Proceedings of the
iEMSs Fourth Biennial Meeting

Main Editors:
Miquel Sànchez-Marrè. Javier Béjar, Joaquim Comas,
Andrea E. Rizzoli, Giorgio Guariso



International Congress on Environmental Modelling and Software iEMSs 2008

Integrating Sciences and Information Technology for Environmental Assessment and Decision Making

Volume 1

Universitat Politècnica de Catalunya, Barcelona, Catalonia

7-10 July 2008

Proceedings of the
iEMSs Fourth Biennial Meeting:

International Congress on Environmental Modelling and Software

iEMSs 2008

International Environmental Modelling and Software Society (iEMSs)
Universitat Politècnica de Catalunya
Barcelona, Catalonia
July 7-10 2008

SPONSORS:

ISBN: 978-84-7653-074-0



INTERNATIONAL
Environmental
Modelling
and Software
SOCIETY

International Congress on Environmental Modelling and Software
iEMSs 2008, July 7-10 2008
Universitat Politècnica de Catalunya, Barcelona, Catalonia



Impressum

Copyright: the copyright of all papers is an exclusive right of the authors. No work can be reproduced without written permission of the authors.

Each session paper in these proceedings was refereed by a member of the Editorial Board and two anonymous referees. Each workshop abstract was refereed by a member of the Editorial Board.

ISBN: 978-84-7653-074-0

Published by the International Environmental Modelling and Software Society (iEMSs)
President: Giorgio Guariso. Address: c/- IDSIA, Galleria 2, 6928 Manno, Switzerland.
E-mail: president@iemss@org. Website: http://www.iemss.org

Printed in Barcelona, Catalonia

Main Editors

Miquel Sànchez-Marrè, Javier Béjar, Joaquim Comas, Andrea E. Rizzoli, Giorgio Guariso

Editorial Board

Alexey Voinov, Andrea Emilio Rizzoli, Anthony J. Jakeman, Arno Krause, Ayalew Kassahun, Barry Croke, Bernd Page, Bert Bredeweg, Brian S. McIntosh, Carlo Giupponi, Carlo Gualtieri, Carmel A. Pollino, Chris Duffy, Claudia Turón, Dagmar Haase, Daryl Hepting, David Arctur, David Post, Dragutin T.Mihailovic, Eduardo Ayesa, Farmani Raziyeh, Geoffrey Holmes, Georgii A. Alexandrov, Gerald Schimak, Giorgio Guariso, Giulia Fiorese, Giulio Iovine, Gürkan Sin, Hirdy Miyamoto, Ian Littlewood, Ignasi Rodríguez-Roda, Ilya Zaslavsky, Ioannis Athanasiadis, Ion Irizar, Ivan Holoubek, Jaroslav Myšiak, Jean Philippe Steyer, Jennifer Koch, Jim Freer, Jiří Hřebíček, Joaquim Comas, Joaquín Izquierdo, Juan M. Lema, Karina Gibert, Keith Matthews, Kostas Karatzas, Krist V. Gernaey, Kristina Voigt, Les Oxley, Lorenz Hilty, Manel Poch, Manuel Pastor, Mark van Loodstrecht, Martin van Ittersum, Martin Volk, Michael Sheridan, Michela Robba, Michiel Blind, Miquel Sànchez-Marrè, Nigel Quinn , Onno Roosenschoon, Paulo Salles, Peter Gijsbers, Peter Struss, Piotr Magnuszewski, Raffaele Giordani, Ralf Isenmann, Ralf Seppelt, Riccardo Minciardi, Rick Sojda, Robert Argent, Roberto Sacile, Roger Moore, Rüdiger Schaldach, Sigrun Kabisch, Steven Frysinger, Thomas Usländer, Thorsten Wagener, Ulises Cortés, Ulrike Bende-Michl, Virgínia Brilhante, Werner Geiger, Werner Pillmann.

Committees

Convenor

Miquel Sànchez-Marrè FiEMSs, Knowledge Engineering & Machine Learning Group, Technical University of Catalonia, Catalonia

Co-organiser

Joaquim Comas, Laboratory of Chemical & Environmental Engineering (LEQUIA), University of Girona, Catalonia

Scientific Committee

Robert Argent, Melbourne University, Australia

James Ascough III, ARS Tucson, USA

René Bañares-Alcántara, Univ. of Oxford, UK

Bruce Beck, University of Georgia, Georgia, USA

Javier Béjar, Tech. University of Catalonia, Catalonia

Virgínia Brilhante, Federal University of Amazonas, Brazil

Luigi Ceccaroni, Tech. University of Catalonia, Catalonia

Joaquim Comas, University of Girona, Catalonia

Ulises Cortés, Tech. University of Catalonia, Catalonia

Neil Crout, University of Nottingham, United Kingdom

Brian Fath, Towson University, USA

Karina Gibert, Tech. University of Catalonia, Catalonia

Carlo Giupponi, FEEM and University of Milan, Italy

Giorgio Guariso FiEMSs, Politecnico di Milano, Italy

Holly Hartmann, CLIMAS/SAHRA, USA

Anthony J. Jakeman FiEMSs, The Australian Nat. Univ., Australia

Ari Jolma, Helsinki University of Technology, Finland

Kostas Karatzas, Aristotle University, Greece

Guy Larocque, C. de Foresteries de Laurentides, Canada

George Leavesley, USGS, Fort Collins, CO, USA

Carlo Lepori, IDSIA, Switzerland

Rebecca Letcher, The Australian Nat. Univ., Canberra, Australia

Ian Littlewood, CEH Wallingford, United Kingdom

Michael McAleer FiEMSs, Univ. of Western Australia, Perth, Australia

Kimio Morimune, Kyoto University, Japan

Jaroslav Myšiak, FEEM, Venice, Italy

Lex Oxley FiEMSs, Univ. of Canterbury, New Zealand

Claudia Pahl-Wostl, University of Osnabrück, Germany

Manel Poch, LEQUIA, University of Girona, Catalonia

Andrea Rizzoli FiEMSs, IDSIA, Lugano, Switzerland

Ignasi Rodríguez-Roda, University of

Miquel Sànchez-Marrè FiEMSs, Tech. Univ. of

Girona, Catalonia

Ralf Seppelt, UFZ Leipzig, Germany

Jean Philippe Steyer, LBE, INRA, France

David Swayne, University of Guelph, Canada

Martin van Ittersum, U. of Wageningen, The Netherlands

Javier Vázquez-Salceda, Tech. U. of Catalonia, Catalonia

Alexey Voinov FiEMSs, University of Vermont, USA

Catalonia, Catalonia

Richard S. Sojda, NRMSC-USG/Montana State University, USA

Achim Sydow FiEMSs, Fraunhofer FIRST, Berlin, Germany

Clàudia Turon, C. D. de la Conca del riu Besòs, Catalonia

Ilias Vardavas, FORTH and University of Crete, Greece

Kristina Voigt, GSF Neuherberg, Germany

Ruth Waldick, Environment Canada

Organising Committee

Georgii A. Alexandrov, CGER/NIES, Japan

David Arctur, OCGii, USA

Ioannis Athanasiadis, IDSIA, Lugano, Switzerland

Ulrike Bende-Michl, CSIRO Land and Water, Australia

Michiel Blind, RWS-RIZA, The Netherlands

Virgínia Brilhante, Federal University of Amazonas, Brazil

Joaquim Comas, University of Girona, Catalonia

Barry Croke, The Australian National University, Australia

Chris Duffy, Penstate University, USA

Krist V. Gernaey, Technical University of Denmark, Denmark

Peter Gijsbers, WL Delft Hydraulics, The Netherlands

Carlo Giupponi, FEEM and University of Milan, Italy

Giorgio Guariso FiEMSs, Politecnico di Milano, Italy

B. Harris, University of Sheffield, UK

Ivan Holoubek,Masaryk University, Czech Republic

Jiří Hřebíček, Masaryk University, Czech Republic

Ion Irizar, CEIT, San Sebastián, Spain

C. W. Anderson, University of Sheffield, UK

Robert Argent, Melbourne University, Australia

Eduardo Ayesa, CEIT, San Sebastián, Spain

Claudia Binder, University of Zürich, Switzerland

Bert Bredeweg, University of Amsterdam, The Netherlands

Brett Bryan, CSIRO Land and Water, Australia

Ulises Cortés, Tech. University of Catalonia, Catalonia

Neville Crossman, CSIRO Land and Water, Australia

Jim Freer, Lancaster University, UK

Karina Gibert, Tech. University of Catalonia, Catalonia

Raffaele Giordani, CNR-IRSA, Bari, Italy

Carlo Gualtieri, University of Napoli, Italy

Dagmar Haase, UFZ, Leipzig, Germany

Geoffrey Holmes, University of Waikato, New Zealand

A. Holzkaemper, University of Sheffield, UK

Giulio Iovine, CNR-IRPI, Italy

Joaquín Izquierdo, Technical University of Valencia, Spain

Anthony J. Jakeman FiEMSs, The Australian Nat. Univ., Australia

Sigrun Kabisch, UFZ, Leipzig, Germany

Kostas Karatzas, Aristotle University, Greece

Ayalew Kassahun, Wageningen University, The Netherlands

Christof Knöri, EMPA/University of Zürich, Switzerland

Jennifer Koch, University of Kassel, Germany

Vikas Kumar, University of Sheffield, UK

Juan M. Lema, University of Santiago de Compostela, Spain

D. N. Lerner, University of Sheffield, UK

Ian Littlewood, CEH Wallingford, United Kingdom

Piotr Magnuszewski, Wroclaw University of Technology, Poland

Keith Matthews, Macaulay Institute, UK

Brian S. McIntosh, Cranfield University, UK

Dragutin T.Mihailovic, University Novi Sad, Serbia

Riccardo Minciardi, University of Genova/CIMA, Italy

Roger Moore, CEH Wallingford, UK

Jaroslav Myšiak, FEEM, Venice, Italy

Lex Oxley FiEMSs, Univ. of Canterbury, New Zealand

Manel Poch, LEQUIA, University of Girona, Catalonia

David Post, CSIRO Land and Water, Australia

Nigel Quinn, Berkeley National Laboratory, USA

Farmani Raziyeh, IRI, University of Exeter, UK

Michela Robba, University of Genova/CIMA, Italy

Ignasi Rodríguez-Roda, University of Girona, Catalonia

Roberto Sacile, University of Genova/CIMA, Italy

Paulo Salles, University of Brasilia, Brazil

Miquel Sànchez-Marrè FiEMSs, Tech. Univ. of Catalonia, Catalonia

Gerald Schimak, ARC Seibersdorf research, Austria

Huub Scholten, Wageningen University, The Netherlands

Ralf Seppelt, UFZ Leipzig, Germany

Rüdiger Schaldach, University of Kassel, Germany

Gürkan Sin, Technical University of Denmark, Denmark

Richard S. Sojda, NRMSC-USG/Montana State University, USA

Jean Philippe Steyer, LBE, INRA, France

Peter Struss, Technical University of Munich, Germany

B. Surridge, University of Sheffield, UK

Martin van Ittersum, U. of Wageningen, The Netherlands

Mark van Loodstrecht, Tech. Univ. of Delft, The Netherlands

Alexey Voinov FiEMSs, University of Vermont, USA

Martin Volk, UFZ, Leipzig, Germany

Thorsten Wagener, Pennsylvania State University, USA

Patrick Wäger, EMPA, Switzerland

Ilya Zaslavsky, University of California, San Diego, USA

Local Organising Committee

Sergio Álvarez, KEMLG, Technical University of Catalonia, Catalonia
Montse Aulinas, LEQUIA, University of Girona, Catalonia
Cristian Barrué, KEMLG, Technical University of Catalonia, Catalonia
Javier Béjar, KEMLG, Technical University of Catalonia, Catalonia
Guiem Bosch, KEMLG, Technical University of Catalonia, Catalonia
David Cabanillas, KEMLG, Technical University of Catalonia, Catalonia
Luigi Ceccaroni, KEMLG, Technical University of Catalonia, Catalonia
Angela Chieppa, KEMLG, Technical University of Catalonia, Catalonia
Joaquim Comas, LEQUIA, University of Girona, Catalonia
Roberto Confalonieri, KEMLG, Technical University of Catalonia, Catalonia
Ulises Cortés, KEMLG, Technical University of Catalonia, Catalonia
Jordi Dalmau, LEQUIA, University of Girona, Catalonia
Patricia García, KEMLG, Technical University of Catalonia, Catalonia
Karina Gibert, KEMLG, Technical University of Catalonia, Catalonia
Juan Carlos Nieves, KEMLG, Technical University of Catalonia, Catalonia
Xavier Flores, LEQUIA, University of Girona, Catalonia
Carlos Mérida, KEMLG, Technical University of Catalonia, Catalonia
Luis Oliva, KEMLG, Technical University of Catalonia, Catalonia
Fernando Orduña, KEMLG, Technical University of Catalonia, Catalonia
Sofia Panagiotidi, KEMLG, Technical University of Catalonia, Catalonia
Manel Poch, LEQUIA, University of Girona, Catalonia
Pau Prat, LEQUIA, University of Girona, Catalonia
Thania Rendón, KEMLG, Technical University of Catalonia, Catalonia
Ferran Ripoll, LEQUIA, University of Girona, Catalonia
Ignasi Rodríguez-Roda, LEQUIA, University of Girona, Catalonia
Gustavo Rodríguez, KEMLG, Technical University of Catalonia, Catalonia
Miquel Sànchez-Marrè, KEMLG, Technical University of Catalonia, Catalonia
Beatriz Sevilla, KEMLG, Technical University of Catalonia, Catalonia
Joao Sousa, KEMLG, Technical University of Catalonia, Catalonia
Arturo Tejeda, KEMLG, Technical University of Catalonia, Catalonia
Pancho Tolchinsky, KEMLG, Technical University of Catalonia, Catalonia
Clàudia Turon, Consorci per a la Defensa de la Conca del riu Besòs, Catalonia
Javier Vázquez-Salceda, KEMLG, Technical University of Catalonia, Catalonia

Organising Groups

- iEMSs, the International Environmental Modelling and Software Society
- KEMLG, Technical University of Catalonia, Catalonia
- LEQUIA, University of Girona, Catalonia

Collaborating Groups

- ISEM, the International Society for Ecological Modeling
- BESAI, Binding Environmental Sciences and Artificial Intelligence
- MODSS, International Conference on Multi-objective Decision Support Systems for Land, Water and Environmental Management
- ISESS, International Symposium on Environmental Software Systems
- ERCIM, the European Research Consortium for Informatics and Mathematics
- TIAS, The Integrated Assessment Society
- IMACS International Association for Mathematics and Computers in Simulation

Acknowledgements

The main editors and all the members of the editorial board would like to acknowledge the work of more than one hundred of ***reviewers***, whose names cannot be listed here for lack of space. Due to their wonderful work, it has been possible to provide a fair and thorough review to all the papers and abstracts in the proceedings.

The convenors would like to thank the great efforts of all the ***Local Organising Committee members***. Without them, the iEMSs 2008 would not have been possible. Thank you all for your collaboration.

Contents

I. Invited Talks

IPT-1: Modelling Ecological Consequences of Climate Changes
M. Gatto , G. Fiorese and G. De Leo 1

IPT-2: Integrated Assessment & Environmental Modelling
Tony Jakeman, Carmel Pollino and Serena Chen 4

IPT-3: The Use of Environmental Modelling in Innovation and Policy Assessment: on Foresight, Insight and Hindsight
M.K. van Ittersum, B. Sterk and C. Leeuwis 5

IPT-4: Environmental Protection, Sustainable Development & Risk Management with Emphasis on Environmental Chemicals
Kristina Voigt 7

IPT-5: Research Needs in Water Modelling
Gustaf Olsson 12

IPT-6: Artificial Intelligence based Modelling for Environmental Applications and Decision Support
Peter Struss 16

II. Session Papers

Session 1: Data and sensor networks and environmental modelling
(*organised by Robert Argent, Gerald Schimak and Nigel Quinn*) **20**

A National Water Resources Data and Prediction System for Australia
R.M. Argent, R.A. Vertessy, L.J. Minty, B.J. Stewart 20

Delft FEWS: A proven infrastructure to bring data, sensors and models together
P.J.A. Gijsbers, M.G.F. Werner, J. Schellekens 28

Design of a sensor network with adaptive sampling
S. Pons, L. Ceccaroni, J. Piera 37

Sensor networks and decision support tools for adaptive real-time water quality management
N.W.T. Quinn 44

New Sensor System for Environmental Monitoring: the Potentiometric Electronic Tongue
M. Gutiérrez, J.M. Gutiérrez, S. Alegret, L. Leija, P.R. Hernández, R. Muñoz, M. del Valle 54

Particle size distributions in an office environment: An experimental approach and a source apportionment analysis through positive matrix factorization
Saraga D., Sfetsos A., Andronopoulos S., Bartzis J., Maggos T., Vasilakos C. 61

Development of an Information System for Processing of Environmental Monitoring Data
J. Ráček, T. Ludík, V. Kubík, J. Ministr, J. Sedláčková, J. Fiala 69

Shared Data Model to Support Environment Sensor Network Data in Hydrologic Models
M. Kumar, C.J. Duffy 75

Primary Environmental Data Quality Model: Proposal of a Prototype of Model Concept
M. Hejč, J. Hřebíček 83

Session 2: Environmental Fluid Mechanics
(*organised by Carlo Gualtieri and Dragutin T.Mihailovic*) **91**

Implicit Large Eddy Simulation - a Promising Method for Turbulence Modelling in High Resolution Models
V. Fuka and J. Brechler 91

Iterative Inverse Modelling Method for Locating a Source of Air Pollution and its Robustness
B. Rajkovic, M. Vujadinovic and Z. Grsic 99

Evolvable Biological Interfaces: Outline of the New Computing System
I.Balaz and D.T.Mihailovic 104

The Impact of Geometrical Effects on Calculated Value of Albedo on the Heterogeneous Environmental Interfaces
D.V.Kapor A.M.Cirisan and D.T.Mihailovic 114

Integration of Environmental Models Based on Fluid Mechanics in Spatial Information Systems
L.Matejicek, Z. Janour and L. Benes 122

Locating a Source of Air Pollution Using Inverse Modelling and Precomputed Scenarios
M. Vujadinovic, B. Rajkovic and Z. Grsic 129

Two Interacting Environmental Interfaces: Folded Bifurcation in Coupled Asymmetric Logistic Maps
D.T.Mihailovic 134

Validation of the hydrodynamic model of a 3D ecological
C.I.Silva and J.F.Lopes 141

Numerical simulation of flow patterns and mass exchange processes in dead zones
C.Gualtieri 150

Spatial Development of Viscous Flow Induced by Wave Propagation over Vortex Ripples
A. Dimas 162

A simplified experimental method to evaluate equivalent roughness of vegetated river beds
S. De Felice, P. Gualtieri and G. Pulci Doria 170

3D ecological modelling of the Aveiro coastal
A. Cardoso and J. F. Lopes 181

Modelling external intrusion into water distribution systems
P.A. Lopez, J.J. Mora, V. Fuertes and F.J. Martínez 191

Session 3: ICTs, Economic Development and the Knowledge Economy
(*organised by Les Oxley*) **200**

Has the ICT Revolution Run its Course?
K. I. Carlaw, R.G. Lipsey and Ryan Webb 200

National Accounting Frameworks for Measuring Information and Knowledge Economies
David Thorns and Les Oxley 213

ICT Diffusion and Productivity in OECD Economies and New Zealand's Industrial Sectors
K. I. Carlaw 221

Knowledge Workers and the Knowledge Economy
David Thorns and Hong Wang 232

Information and Communication Technologies and the Income Distribution: A General Equilibrium Simulation
T. Fukiharu 240

Simulation study on spatial and temporal changes of factories in an industry district
Hiroyasu Nagata, Kei-ichi Tainaka and Jin Yoshimura 248

Session 4: Intelligent Environmental Decision Support Systems **256**
(*organised by Miquel Sànchez-Marrè, Joaquim Comas, Ignasi Rodríguez-Roda, Karina Gibert, Rick Sojda, Peter Struss, Jean Philippe Steyer and Virgínia Brilhante*)

An approach to formal modeling of environmental knowledge via concept maps and ontologies
A. Simón-Cuevas, L. Ceccaroni and A. Rosete-Suárez 256

Managing Insular Tropical Environment through Data and Knowledge Bases by using Web Services: a case study on Corals and Herbarium of La Réunion Island
N. Conruyt, D. Sebastien, D. Payet, Y. Geynet, D. Caron, and D. Grosser 264

An Intelligent Decision Support System for Environmental Impact Assessment of Industrial Projects
S. Aly and I. Vrana 272

Intelligent Environmental Decision Support System for integrated operation of Membrane Bioreactors
I. Rodríguez-Roda, J. Comas, H. Monclús, J. Sipma and M Poch 280

A new definition of Minimum Sustainable Flow based on water quality modelling and fuzzy processing
S. Marsili-Libelli, A. Nocita, E. Giusti and M. Saccà 288

Improving process performance in wastewater treatment plants using an intelligent environmental decision support system
L. Corominas, J. Denamur, X. Flores-Alsina and J. Comas 295

Genetic algorithm based technique for defining threshold regression models
J.C.M. Pires, F.G. Martins, M.C.M. Alvim-Ferraz and M.C. Pereira 303

Optimum Water Allocation at Basin Scale Using Meta-Modeling
M. Shourian and S. Jamshid Mousavi 311

Machine Learning Algorithms for GeoSpatial Data. Applications and Software Tools
M. Kanevski, A. Pozdnoukhov and V. Timonin 320

Semi-Supervised Support Vector Machine for Natural Hazards Forecasting. Case Study: Snow Avalanches
A. Pozdnoukhov, R. S. Purves and M. Kanevski 328

Using Artificial Neural Networks to Predict Local Disease Risk Indicators with Multi-Scale Weather, Land and Crop Data
A. Pechenick and D.M. Rizzo 336

Incorporating Likelihood information into Multi-objective Calibration of Conceptual Rainfall-Runoff Models
A. Azemi, A. H. C. Chan and X. Yao 346

Environmental data inference through fuzzy clustering
S. Marsili-Libelli and E. Giusti 354

Bayesian Challenges in Integrated Catchment Modelling
V. Kumar, A. Holzkaemper, B. Surridge, P.I. Rockett, M. Niranjan and D. N. Lerner 362

Developing Bayesian network models within a Risk Assessment framework
C. A. Pollino and B. T. Hart 372

Spatially Distributed Identification of Debris Flow Source Areas by Credal Networks
A. Salvetti, A. Antonucci, and M. Zaffalon 380

Learning conditional probabilities in event bushes with temporal labels
S. I. Nikolenko, C. A. Pshenichny , R. Carniel 388

Towards A Framework for the Development of Intelligent Environmental Decision Support Systems
M. Sànchez-Marrè, J. Comas, I. Rodríguez-Roda, M. Poch and U. Cortés 398

Session 5: Monitoring strategies and spatial distribution of processes in models for natural resources management on different scales vision and reality **407**
(*organised by Martin Volk and Ulrike Bende Michl*)

Soil Moisture Monitoring at Different Scales for Rainfall-Runoff Modelling
L. Brocca, F. Melone and T. Moramarco 407

Radar-based surface soil moisture retrieval over agricultural used sites – A multi-sensor approach
M. Pause, M. Volk and K. Schulz 415

Overland flow and soil erosion in an undisturbed Brazilian Northeastern Semiarid Experimental Plot
L. F. F. Moreira, F. de O. Silva, A. M. Righetto and V. M. de A. Medeiros 422

Entropy Theory Application for Flow Monitoring in Natural Channels
T. Moramarco, A. Ammari, A. Burnelli, D. Mirauda and V. Pascale 430

On the Lateral Inflows Assessment Within a Real-Time Stage Monitoring Addressed to Flood Forecasting
S. Barbetta, L. Brocca, F. Melone and T. Moramarco 438

Field Measurement Network in the Mesoscale Catchment of the Gera River
M. Fink, M. Wetzel, D. Varga, S. Kralisch and W.-A. Flügel 446

Insights into catchment behaviour for water harvesting assessment in the East India Plateau
B. Croke, S. Norrish, A. Kumar, A. Islam, P.Dey, P. Cornish, S. Kumar, J. Ghosh 454

Understanding riverine wetland-catchment processes using remote sensing data and modelling
A. van Griensven, Y. Xuan, D. Haguma and W. Niyonzima 462

Effects of Rainfall and Soil/Land Use Spatial Distribution on Hydrological Response at Different Scales
N. Berni, A. Viterbo, C. Pandolfo, M. Stelluti, S. Barbetta and L. Brocca 470

The Significance of Spatial Variability of Rainfall on Runoff
I.G. Pechlivanidis, N. McIntyre and H.S. Wheater 478

Integration of In-situ and remote sensing Data for Water Risk Management
Bernhard H.C. Sputh, Oliver Faust, Lasse H. Pettersson, Torill Hamre, Damiano Vitulli, Alastair R. Allen, and Tim Spracklen 486

A rainfall-runoff database to support flood risk assessment
S. Liersch and M. Volk 494

Bottom-up methodologies for the modeling and upscaling of farm Nitrogen losses in European landscapes
T. Dalgaard and C. Kjeldsen 503

Simulation of the flow in the Koiliaris River basin (Greece) using a combination of GIS, the HSPF model and a Karstic– Snow melt model
N.N. Kourgialas, G.P. Karatzas and N.N. Nikolaidis 512

Application of the HSPF Model for Flood Simulation with Analysis of the Results in Terms of Monitoring Uncertainties /Case Study of the Lesnovska River, Bulgaria/
P.Ninov, I.Ribarova, P.Kalinkov and G. Dimova 520

Role of Models for Building an Efficient Monitoring Design
M.G. Erechtchoukova and P.A. Khaiter 528

Multilinear Diffusion Analogy Model for Real-Time Streamflow Routing
M. Perumal, T. Moramarco, B. Sahoo and S. Barbetta 536

Influence of the uncertainties of monitoring data on model calibration and evaluation
A. Ullrich, M. Volk and G. Schmidt 544

Effects of "ENSO-events" and rainforest conversion on river discharge in Central Sulawesi (Indonesia) – problems and solutions with coarse spatial parameter distribution for water balance simulation
G. Gerold and C. Leemhuis 553

Uncertainty in Great Barrier Reef Catchment soil nutrient data - implications for land use management
B. S. Sherman and A. M. Read 566

Data Uncertainty Estimation Tool for Hydrology and Water Quality (DUET-H/WQ): Estimating Measurement Uncertainty for Monitoring and Modelling Applications
R.D. Harmel, D.R. Smith, K.W. King, R.M. Slade and P. Smith 574

Catchment-Scale Water Quality Modelling and Integration of Collateral Information
L.T.H. Newham, J.J. Drewry and B. Fu 584

Session 6: Integrated Catchment Management: How can modelling tools and techniques help? **592**
(*organised by D. N. Lerner, C. W. Anderson, B. Harris, A. Holzkaemper, B. Surridge and V. Kumar*)

CLUES Spatial DSS: From Farm-Scale Leaching Models to Regional Decision Support
S. Elliott, G. McBride, U. Shankar, A. Semandi-Davies, J. Quinn, D. Wheeler, L. Wedderburn, B. Small, A. Hewitt, R. Gibb, R. Parfitt, B. Clothier, S. Green, S. Harris and G. Rys 592

Linking Catchment Environmental Planning to On-Ground Investment: the SCaRPA DSS
N.F. Herron, T.W. Barrett and S.M. Cuddy 560

A consistent framework for knowledge integration to support Integrated Catchment Management
A. Holzkämper, B. Surridge, A. Paetzold, V. Kumar, D. N. Lerner, L. Maltby, J. Wainwright, C. W. Anderson and R. Harris 568

MoGIRE: A Model for Integrated Water Management
A. Reynaud and D. Leenhardt 576

Step by step calibration of an Integrated Model System for irrigation management
T. Arnold, T. Berger and H. Uribe 584

Development of an integrated framework for assessing the impact of urban planning on water quality
G. Fu, S.-T. Khu, F. di Pierro, A. Doglioni and F. Primativo 592

Representing river system behaviour in UML to enhance development of DSS
J. A. E. B. Janssen and Ioannis M. Dokas 598

Model development for integrated water management in the influence area of phosphate mines
K. Bru, A-G. Guézennec, S. Lanini, N. Graveline, K. Soulis, F. Karaouli and A. Guesmi 606

Aspects of the application of decision and information support tools in an integrated water resources management practice
R. Hranova 614

Integrated modelling of water and nitrate fluxes in large catchment areas – assessing the needs and efficiencies of agroenvironmental measures for selected groundwater bodies in Germany
R. Kunkel, M. Eisele, B. Tetzlaff and F. Wendland 622

The Role of Groundwater Resources Models in Integrated Catchment Management
Farrell. R and Whiteman. M 631

Model based sediment quality management on river basin
B. Westrich 639

Large-scale modelling of P-inputs from point and diffuse sources in German river basins
B. Tetzlaff and Frank Wendland 647

A System Dynamics Model for the Assessment of Risks and Risk Mitigation Options at Catchment Scale

M. Finkel and U. S. McKnight 655

Multiple-criteria decision making for integrated watershed management in the Ramian watershed, Golestan, Iran
A. Sadoddin, V. Sheikh, R. Mostafazade and M.G. Halili 662

A Toolbox for the Identification of Parsimonious Semi-Distributed Rainfall-Runoff Models: Application to the Upper Lee Catchment
B. Orellana, I.G. Pechlivanidis, N. McIntyre, H.S. Wheater and T. Wagener 670

Development of matrices to build scenarios of cropping systems distribution for integrated catchment assessment. A proposition for an irrigated area in south-western France
L. Clavel and D. Leenhardt 678

Deposition and Resuspension of Sediments in Near Bank Water Zones of the River Elbe
S. Prohaska and B. Westrich 686

Challenges of alpine catchment management under changing climatic and anthropogenic pressures
C. de Jong, P. Masure and T. Barth 694

Ecosystem Stability in Integrated Catchment Management
P. A. Khaiter and M. G. Erechtchoukova 703

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Modelling the ecological consequences of global climate change

M. Gatto[a] , G. Fiorese[a], G. De Leo[b]
[a]*Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano 20133, Italy (gatto/fiorese@elet.polimi.it)*
[b] *Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Viale Usberti 11/A, Parma 43100, Italy (giulio.deleo@unipr.it)*

***Keywords**:* Biodiversity; Ecosystem functioning; Envelope models; Population dynamics; Alpine ibex

The Earth's climate is changing at an unprecedented rate and, moreover, changes are not driven mainly by natural variations, rather by anthropogenic causes. The recently released Intergovernmental Panel of Climate Change (IPCC) Summary on the Physical Science Basis of global climate change [IPCC, 2007a] stated: "the global increases in carbon dioxide concentration are due primarily to fossil fuel use and land-use change, while those of methane and nitrous oxide are primarily due to agriculture". Carbon dioxide (CO_2) atmospheric concentration has increased from a pre-industrial value of about 280 ppmv to 384 ppmv in 2008, which is the highest concentration in the past 650,000 years [Siegenthaler et al., 2006]. In the last decade an exceptional increase of 1.9 ppmv per year has been experienced [IPCC, 2007a].

The 2007 IPCC report showed that the average surface temperature of the Earth increased more than projected in the 2001 Third Assessment Report [TAR; IPCC, 2001]: the 1906-2005 linear trend indicates a temperature increase of 0.74±0.18°C, which is larger than the corresponding trend of 0.6±0.2°C for 1901-2000 [IPCC, 2001]. The 8 warmest years recorded since 1680, when instrumental measures of temperature became available, occurred in the last 14 years and 2005 has been the warmest year ever. Europe is experiencing even higher rates of temperature increase: 0.95°C versus 0.74°C [EEA, 2004].

The detection of global and regional trends for precipitation is more difficult because precipitation regimes are inherently stochastic and strongly depend on local climatic and morphological characteristics. Nonetheless, improvement in data analysis and modelling have been made and, according to the IPCC report [IPCC, 2007a], significant precipitation increase has been observed in eastern parts of North and South America, in northern Europe and in northern and central Asia. Drying has been observed in the Sahel, the Mediterranean, southern Africa and parts of southern Asia.

1. THE IMPACTS OF GLOBAL CLIMATE CHANGE ON ECOSYSTEMS

There are numerous impacts of Global Climate Change (GCC) on ecosystems: climate is in fact one of the factors that determines ecosystems' composition, productivity and structure. Several plants can successfully grow and reproduce in a specific range of temperatures and precipitation regimes. Regional climate is one of the factors that influences species geographical distributions through physiological tolerance thresholds, e.g., of temperature and rainfall, and through food or nutrient availability. Biota with narrow physiological and phenological ranges will be the most vulnerable to climate change.

Climate change can affect ecosystems, populations and individuals directly (e.g., via rising average temperature) or indirectly (e.g., via changes in the availability of resources). Impacts imply changes in the structure and functioning of ecosystems, in the physiology and phenology of plant and animal species, in their range distribution.

A major negative impact is given by the increased frequency of extreme events (e.g., heat waves, droughts, storms), that directly affects ecosystems; at the same time, small, continuing and progressive variations of temperature and of other climate drivers can affect the biological cycle of many species. Mild winters can affect the competitive interactions existing within the species, altering the structure of the biological community and, in some case, enabling the spread of pathogens that would otherwise die during winter [Harvell et al., 2002].

2. MODELLING THE IMPACTS

One of the most important questions for ecologists is whether GCC will imply a a small or large decrease of biodiversity (at both species and ecosystem level). Biodiversity conservation is usually based on the idea that species extinction occurs on very long time scales; however, the speed of climate change makes this assumption untenable. We must know how species and ecosystems will react on a short time scale [Araújo and Rahbek, 2006].

To predict the expected impacts of GCC on biodiversity and the ecosystem functioning it is necessary to know how climatic variables will evolve in the future and how ecosystems will respond to climatic variations. Present day's AOGCMs (Atmospheric-Oceanic Global Circulation Models) outline scenarios on the future evolution of temperature, precipitation and other climatic variables on geographic grids of approximately 200 x 200 km. There is agreement among the different models as concerns temperature, while there is some disagreement on precipitation trends. Also, AOGCMs do not incorporate soil and coast morphology, the presence of water bodies and differences in vegetation cover, which affect the local climate. This is of course an important limitation for the study of the ecological consequences of GCC [Metzger et al., 2004]. However, efforts are undertaken in several research centers to downscale the AOGCM predictions to a finer scale so that they can be used to assess the vulnerability of the different areas of our planet [IPCC, 2007b].

There are several categories of models that link climatic variation to ecosystem response. First of all we can contrast static vs. dynamic models. Static models use empirical relationships between abiotic factors (including climate) and biotic variables. Biosphere is considered at equilibrium. The relationships are usually obtained via statistical methods (such as logistic regression, generalized linear models, etc.) applied to the past data. Dynamic models simulate one or more ecological processes (such as population reproduction or species migration) that typically vary in time. These models produce transients, not only equilibria. A second contrast is between single species and ecosystem models. The first simulate single populations isolated from other components of the ecosystem to which they belong. The latter simulate changes in ecosystem/biome function and composition, whether at the level of species or of functional groups. A final contrast is between models at the global or local scale. Ecological models that directly use the outputs of AOGCMs on large grids simply predict the presence or absence of species and are frequently based on so-called climate envelopes, i.e., relationships between mesoscale climate variables and species presence. Local models, instead, must be based on predictions downscaled from AOGCMs and can detail the abundance and demography of animal or plant species, the physiological processes, the ecological interactions and local species diversity.

The great majority of the existing models are static, single species and global scale. Actually all the predictions on the future loss of biodiversity [Thomas et al., 2004, Thuiller et al., 2005] are based on models of this type. We will present examples of these predictions and discuss their shortcomings. To increase the realism and utility of

predictions, next generation models should necessarily take steps towards being dynamic and local scale with at least a simple description of ecosystem functioning. To exemplify, we will present preliminary results for some Alpine species with particular regard to the distribution and demography of the Alpine ibex (*Capra ibex ibex*) in the two national parks of Adamello and Gran Paradiso, Northern Italy.

ACKNOWLEDGMENTS

The research has been supported by Fondazione Lombardia per l'Ambiente and Regione Lombardia (Progetto Kyoto Lombardia), and by Ministero dell'Università e della Ricerca (Progetto Interlink II04CE49G8).

REFERENCES

Araújo M.B., and C. Rahbek, How does climate change affect biodiversity? *Science 313*, 1396-1397, 2006.

EEA, *Impacts of Europe's changing climate. An indicator-based assessment*. N2/2004, Copenhagen, 2004.

Harvell C.D., C.E. Mitchell, J.R. Ward, S. Altizer, A.P. Dobson, R.S. Ostfeld, and M.D. Samuel, Climate warming and disease risks for terrestrial and marine biota, *Science 296*, 2158-2162, 2002.

IPCC, *Climate change 2001: The scientific basis*, Cambridge University Press, Cambridge, UK, 2001.

IPCC, *Climate Change 2007: The physical science basis, Summary for Policymakers*, Cambridge University Press, Cambridge, UK, 2007a.

IPCC, *Climate Change 2007: Impacts, adaptation and vulnerability, Summary for Policymakers*, Cambridge University Press, Cambridge, UK, 2007b.

Metzger M.J., R. Leemans, D. Schröter, W. Cramer and the ATEAM consortium, *The ATEAM vulnerability mapping tool*, Quantitative Approaches in Systems Analysis No. 27, CD-ROM publication, Office C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC), Wageningen, The Netherlands, 2004.

Siegenthaler U., T.F. Stocker, E. Monnin, D. Lüthi, J. Schwander, B. Stauffer, D. Raynaud, J.M. Barnola, H. Fischer, V. Masson-Delmotte, and J. Jouzel, Stable carbon cycle–climate relationship during the late Pleistocene, *Science 310*, 1313-1317, 2006.

Thomas C.D., A. Cameron, R.E. Green, M. Bakkenes, L.J. Beaumont, Y.C. Collingham, B.F.N. Erasmus, M. Ferreira de Siqueira, A. Grainger, L. Hannah et al., Extinction risk from climate change, *Nature 427*, 145-148, 2004.

Thuiller W., S. Lavorel, M.B. Araújo, M.T. Sykes, and I.C. Prentice, Climate change threats to plant diversity in Europe, *Proceedings of the National Academy of Sciences 102*, 8245–8250, 2005.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

INTEGRATED ASSESSMENT AND ENVIRONMENTAL MODELLING

Tony Jakeman, Carmel Pollino and Serena Chen
Integrated Catchment Assessment and Management and
Fenner School of Environment and Society
The Australian National University, Canberra

It is clear that the modelling of environmental systems problems poses considerable difficulties. These systems are complex, distributed in space, dynamical (with time-spread responses to change), and heterogeneous. Often the problems are wicked ones whose formulation is not totally agreed upon by stakeholders and who may have different and changing preferences. Their models must cover a range of sciences (e.g. hydrology, ecology, agriculture, forestry, economics, psychology, even demography, sociology and politics) and a range of categories of people affected. The evolving discipline of Integrated Assessment (IA), which will be discussed briefly to provide context, aims to deal with such systems, including the human component. Integrated Scenario Modelling (ISM) is a core activity of many IA exercises. It involves a model as an approximation of the system under study. The model allows the simulation of how input drivers (scenarios) such as climate and human activities yield outputs (indicators) representing the states of the system. In IA, the system is extended so that policy and management link to the controllable human activity inputs. The talk will characterise the different types of integrated modelling frameworks that are available and illustrate their appliability for particular uses. It will take as an example the issue of how managers in river basins can prioritise their investments to meet resource condition and socioeconomic outcomes. It will also summarise the lessons that we have learnt.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

The Use of Environmental Modelling in Innovation and Policy Assessment: on Foresight, Insight and Hindsight

M.K. van Ittersum[a], B. Sterk[a,b,c], C. Leeuwis[c]

[a] *Plant Production Systems group, Wageningen University, P.O. Box 430, 6700 AK Wageningen, The Netherlands (martin.vanittersum@wur.nl)*
[b] *Land Use Planning group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands (barbara.sterk@wur.nl)*
[c] *Communication and Innovation Studies group, Wageningen University, Hollandseweg 1, 6706 KN Wageningen, The Netherlands (cees.leeuwis@wur.nl)*

Keywords: agricultural systems, integrated assessment, learning, policy-oriented research

Almost any scientific paper on environmental modelling, decision support systems or integrated assessment tools claims and argues that the management of land and natural resources can potentially benefit from such research efforts. Whether this potential is actually fulfilled is a question that only recently has been receiving some explicit attention (see McIntosh et al, in press for an overview from iEMSs 2006), and the relatively few documented analyses [e.g., McCown et al., 2002; McIntosh et al., 2005] show that the potential in terms of measurable impact is often not achieved. Also, impact is not easy to measure as the processes at which these research tools target are frequently highly non-linear and dispersed in time and space. This paper reports and reflects on the contributions of *land use* models, as one type of environmental models, to learning for societal problem solving, i.e., learning of farm managers and/or land use planners at local, regional, national or international level, to solve land use related problems. This is done through seeking a suitable opportunity to introduce a research model [e.g., Sterk et al., 2006], a comparative analysis of several cases where research models had a demonstrated impact [Sterk, 2007] and experiences from recent projects [e.g., Van Ittersum et al, 2008]

The research models that form the basis of this paper all attempt in some way to address the future, but with different purposes and methods. Foresight studies can be classified, for instance, as either projective, predictive, explorative or speculative [Van Ittersum et al., 1998] and their roles may be heuristic, improving understanding, symbolic, putting an issue on the political agenda, and relational, creating a community [Shackley & Wynne, 1995; Van Daalen et al, 2002; McIntosh et al., 2005]. These may be, implicitly or explicitly, targeted at different phases of innovation or policy cycles [e.g. Van Daalen et al, 2002]. Do computer models play such roles and if so, what kind of arrangements, conditions, model qualities, or other factors harness land use modelling to perform specific roles in contexts where different stakeholders play a role?

In our research, in successful cases, learning through modelling took the form of a new perspective on a land use system, frequently in combination with a better understanding of the position of other stakeholders, resulting in adapted problem definitions, a changed solution space and/or the formation of new coalitions to tackle a particularly land use related problem. Models were found to contribute not only to improving understanding (heuristic role) but also to agenda-setting (symbolic role) and the creation of communities (relational role). Literature suggests critical success factors for research models, such as the need for proper timing of availability, ease of graphical user interfaces, transparency and representation of uncertainties. In addition to, and at times perhaps instead of, such rather

technical, static and distinct factors, the study suggests that societal context, actors aspirations, experienced interdependency, network building and model contextualisation were explanatory variables for impact of computer models. Thus, we (researchers) need to anticipate the relatively fluid and fuzzy features of social contexts and problem solving processes to harness land use modelling for societal learning.

What do the findings imply for those who wish to pursue the use of science-based land use models to contribute to societal problem solving? First of all, the analysis demonstrates that the contributions of land use models to societal problem solving can be various and distinct. The contributions are not limited to learning about a land use system but are more diverse and extend to learning about the views, norms and values of other factors, mediation of conflicts between stakeholders and community building when the organization of stakeholders is desirable for coping with a problem. Furthermore, the research suggests that in designing a modelling strategy, equal attention needs to be paid to the requirements for model development, and the embedding of the work in a given/intended societal context. For policy-oriented research specifically, the notion 'boundary arrangement' of science and policy proved helpful in understanding the position of science vis a vis policy and its institutions and hence in better devising a strategy for contextualisation that enhances impact of model-based research [Sterk et al, in press].

REFERENCES

McCown, R.L., Z. Hochman, and P.S. Carberry (Eds.), Learnings from 25 years of agricultural DSS: Case histories of efforts to harness simulation models for farm work, *Agricultural Systems, 74*, 1-220, 2002.

McIntosh, B.S., P. Jeffrey, M. Lemon, and N. Winder, On the design of computer-based models for integrated environmental science, *Environmental Management, 35*, 741-752, 2005.

McIntosh, B.S., C. Giupponi, A. Voinov et al., Bridging the gap: developing tools for environmental policy and management, In: Jakeman, T., Rizzoli, A., Voinov, A. & Chen (eds.). State of the Art and Futures in Environmental Modelling and Software, Elsevier, in press.

Shackley, S., and B. Wynne, Global climate change: The mutual construction of an emergent science-policy domain, *Science and Public Policy, 22*, 218-230, 1995.

Sterk, B., A window of opportunities. The contributions of land use modelling to societal learning, PhD Thesis Wageningen University, 136 pp, 2007.

Sterk, B., M.K. van Ittersum, C. Leeuwis, W.A.H. Rossing, H. van Keulen, and G.W.J. van de Ven, Finding niches for whole-farm design models – contradictio in terminis? *Agricultural Systems, 87*, 211-228, 2006.

Sterk, B., P. Carberry, C. Leeuwis, M.K. van Ittersum, M. Howden, H. van Keulen, and W.A.H. Rossing, The interface between land use system research and policy: multiple arrangements and leverages, *Land Use Policy*, in press.

Van Daalen, C.E., L. Dresen, and M.A. Janssen, M.A., The roles of computers models in the environmental policy life cycle, *Environmental Science and Policy, 5*, 221-231, 2002.

Van Ittersum, M.K., R. Rabbinge, and H.C. van Latesteijn, Exploratory land use studies and their role in strategic policy making, *Agricultural Systems, 58*, 309-330, 1998.

Van Ittersum, M., F. Ewert, T. Heckelei et al., Integrated assessment of agricultural systems – A component-based framework for the European Union, *Agricultural Systems, 96*, 150-165, 2008.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Environmental Protection, Sustainable Development & Risk Management with Emphasis on Environmental Chemicals

Kristina Voigt
Helmholtz Zentrum, Muenchen, German Research Center for Environmental Health, Ingolstaedter Landstr. 1, 85764 Neuherberg, e-mail: kvoigt@helmholtz-muenchen.de

Keywords: Environmental risk, *environmental risk* assessment, *environmental risk management, sustainability*, *sustainable chemistry, green chemistry, environmental models, environmental decision support systems, Hasse Diagram Technique*

The relevance and acceptance of **environmental protection** in society has been changing over the past decades. According to my own experience in Germany the topic was of high importance at the end of the seventieth till mid 1980ieth. Then the interest turned to other important societal subjects such as risk of unemployment, retirement pay, reunification, warlike operations etc. These days a lot of environmental research is performed in the direction of human health and the environment. Currently another focus lies on the issues regarding climate change. Furthermore the subjects energy and the environment receive more attention.

Only recently the terms environmental protection often in connection with **sustainability** and **sustainable development** receive an increasing interest. **Sustainability** is a characteristic of a process or state that can be maintained at a certain level indefinitely. The term, in its environmental usage, refers to the potential longevity of vital human ecological support systems, such as the planet's climatic system, systems of agriculture, industry, forestry, and fisheries, systems on which they depend in balance with the impacts of our unsustainable or sustainable design (WIKIPEDIA, 2008). **Sustainable development** is a pattern of resource use that aims to meet human needs while preserving the natural environment so that these needs can be met not only in the present, but in the indefinite future. The term was used by the Brundtland Commission which coined what has become the most often-quoted definition of sustainable development as development that "meets the needs of the present without compromising the ability of future generations to meet their own needs"(United Nations, 1987). **Sustainable development** does not focus solely on environmental issues. The United Nations 2005 World Summit Outcome Document, refers to the "interdependent and mutually reinforcing pillars" of sustainable development as economic development, social development, and environmental protection (United Nations, 2005). Concerning the chemicals' policy a stable balance between economic interests and the challenging protection of the environment and human health must be given. Only with respect to this balance we will be able to further profit from the usefulness of chemistry and chemicals, e.g. pharmaceuticals, pesticides, polymers, fertilizers, building materials, flame retardants, etc. In this context we often talk about **sustainable chemistry** or **green chemistry** (Richter, 2008). The discussion, can chemists contribute to sustainable development, and in which way is rather young. The chemical industry has achieved significant improvements with regard to the reduction of direct emissions and waste. This is a result from the high technical level which has been developed in the meantime, under consideration of the saving of energy and raw material. The most hazardous chemicals like PCBs (Polychlorinated biphenyls) and other POPs (Persistent Organic Pollutants) are banned.

Sustainability targets, however, continue to play only a subordinate role in the development of new chemicals and chemical processes. New chemicals should fulfil their function in an optimised manner; but the proportion of new chemicals which are classified as dangerous is not lower than that of existing chemicals. When developing new processes or products, criteria like low resource demand, low waste, low toxicity are mostly of minor importance. Sustainability is still waiting to become a main goal worthwhile being conquered (Steinhäuser, 2004). Today, the controversial discussion on chemicals has generated a more differentiated thinking necessary to minimize the **risks** posed by chemicals and chemical processes.

The risk decision process is traditionally divided into two stages, **risk assessment** and **risk management**. **Risk assessment** is the major bridge linking science to policy.

The European risk assessment principles for new and existing chemicals are laid down in Commission Directive 93/67/EEC and 1488/94 (European Commission, 2003), respectively. Increasing concern that these European Commission regulations do not provide sufficient protection and that less than hundred high priority substances underwent a risk assessment in the past 10 years led to a review of the current policy on chemicals. A new system called REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) has recently been adopted (European Commission, 2006). The aim of REACH is to improve the protection of human beings (comprising of workers, consumers, and humans indirectly exposed via the environment) as well as ecosystems in the aquatic (water and sediment) and terrestrial compartments (including top predators) from adverse effects of chemicals while maintaining the competitiveness and enhancing the innovative capability of the EU chemicals industry. Within the context and scope of REACH, there is a need to be able to efficiently perform risk assessments on thirty thousands chemicals manufactured in or imported into Europe.

An e**nvironmental chemical risk/safety assessment** usually proceeds in the following sequence: **hazard assessment**, **exposure assessment** and **risk characterisation**. In the **hazard assessment**, reliable and relevant longterm (chronic) ecotoxicity data for organisms belonging to different trophic levels are gathered. For a limited effects database, the predicted no effect concentration (PNEC) is calculated by applying an assessment factor (AF), reflecting sources of uncertainty, to the lowest ecotoxicity value observed. For a sufficiently large effects database, a species sensitivity distribution (SSD) can be used to derive the PNEC value. The 5th percentile is used as the PNEC estimate, after application of an AF between one and five to cover remaining uncertainties (European Commission, 2003). In the **exposure assessment**, a distinction is made between different spatial scales (European Commission, 2003). The local scale considers the vicinity of a point source and the local predicted environmental concentration (PEClocal) is calculated. The regional scale assesses the exposure levels due to diffuse/widespread releases in a larger region (PECregional). The PECregional acts as the background concentration for the local assessment. The technical principles, described in the EU TGD – Technical Guidance Document (European Commission, 2003), are implemented in the computer program EUSES (European Commission, 1998). EUSES (European Union System for the Evaluation of Substances) software first calculates releases of chemicals based on the volume produced or imported, the use pattern, and the physical-chemical properties of the chemical concerned. These release estimates are subsequently translated into PECs for each environmental compartment (air, water, sediment, soil) based on the transport and fate of the substance. The **risk characterisation** comprises of a quantitative comparison of the PEC, for most substances under REACH estimated through modelling, with the PNEC. The risk characterisation ratio (RCR), or PEC/PNEC ratio, larger or equal to one signifies that there is a potential risk of adverse effects occurring. A RCR smaller than one signifies no need for further information and/or testing and/or implementing risk management measures (RMMs) (Verdonck, 2008). Improvements and enhancements concerning the EUSES risk assessment model / software are suggested by several authors.

Jager et al. are of the opinion that probabilistic risk assessment is preferable to deterministic quotients to address the complex problems of chemical risk assessment and risk management. Nevertheless, it takes time to change the conventional and well-accepted approaches, and risk managers as well as risk assessors need to familiarise themselves with probabilistic methods and distributions of risk. Perhaps it is possible to use a deterministic and a probabilistic approach side–by-side in a decision support system. In this way, risk assessors and risk managers can compare the outcomes and get accustomed to dealing with probability distributions (Jager, 2001). A more recently published paper by Verdonck et al.

demonstrates the usefulness of an efficient risk refinement tool based on sensitivity and uncertainty analysis on the risk assessment model EUSES (Verdonck, 2008).
In **risk assessment**, scientific data on toxicological and ecotoxicological effects are used to determine possible adverse effects and the exposure levels at which these effects may be expected. These effect levels are then compared to predicted (or monitored) exposures in order to characterize the risk. The risk characterization is used as a part of the basis for risk management decisions on appropriate measures to handle the risk. Such decisions range from taking no actions at all, via limited measures to reduce the highest exposures, to extensive regulations aiming at completely eliminating the risk, for instance by prohibiting activities leading to exposure. In the **risk management** decision, factors other than the scientific assessment of the risk are taken into account, such as social and economical impacts, technical feasibility, and general social practicability. According to the European Commission Technical Guidance Document for risk assessment (European Commission, 2003) "the risk assessment process relies heavily on expert judgment". Issues to be determined by expert judgment include central issues in many risk assessments, for instance the relevance of test data obtained with non-standardized methods, the interpretation of conflicting data, data quality evaluation, assessment of carcinogenicity and mutagenicity, and whether an effect is causally connected to exposure or not (European Commission, 2003). The combined effect of scientific uncertainty and a high degree of flexibility and reliance on individual experts make it in practice impossible to achieve a risk assessment process that is fully consistent and systematic in all aspects. It is therefore essential to scrutinize and evaluate the risk assessment process, in order to learn more about (1) how scientific information is reflected in risk assessment, (2) how and to what degree risk assessment influences risk management, (3) to what degree the risk decision process as a whole satisfies general quality criteria such as efficiency, consistency, and transparency. Hansson and Ruden made an approach to evaluate the risk decision processes (Hansson, 2006).
The complex topic of environmental chemical risk assessment has hence to be seen in connection with environmental models and environmental decision support systems (EDDSs). A DSS (decision support system) has been defined in many different ways, but it can be regarded in general as an interactive, flexible, and adaptable computer based information system especially developed for supporting the recognition and solution of a complex, poorly structured or unstructured, strategic management problem for improved decision-making. It uses data and models, provides an easy, user-friendly interface, and can incorporate the decision-makers own insights. In addition, a DSS is built by an interactive process (often by end-users), supports one or more phases of decision-making, and may include a knowledge component. An environmental decision support system (EDSS) often consists of various coupled environmental models, databases and assessment tools, which are integrated under a graphical user interface (GUI), often realized by using spatial data management functionalities provided by geographical information systems (GIS). An overview of the current activities, methods and tools in this area is outlined by Matthies et al. (2007).

In my further presentation I will give an example applying a software tool for environmental decision support based on the theory of partially ordered sets. The background and application of the software tool named Hasse Diagram Technique were introduced by the author recently (Voigt, 2006). This environmental decision support tool is now applied on the ranking of environmental pollutants found in soil samples in the Alps. The increasing worldwide contamination of soil and freshwater systems with thousands of industrial and natural chemical compounds is one of the key environmental problems facing humanity (Schwarzenbach, 2006). Although most of these compounds are present at low concentrations, many of them raise considerable toxicological concerns, particularly when present as components of complex mixtures. I frame the concerns primarily from an environmental-protection perspective with a focus on soil ecosystem in the Alps, but without neglecting the human health issues. Protecting natural waters and soils against chemical pollution safeguards soil and aquatic life and thus, directly or indirectly, human health. Hence, any measures taken to prevent the chemical pollution of surface and groundwater resources as well as soils will not only improve ecosystem health, but will also benefit both the production of clean water and safe food for human consumption

(Schwarzenbach, 2006). As an example we use a data-matrix which was generated in an international project named MONARPOP (Monitoring Network in the Alpine Region for Persistent and other Organic Pollutants) in which selected chemicals in environmental media in the mountain area of the Alps were analyzed in the years 2004 and 2005. 17 pesticides were chosen and analyzed in soil samples in Germany, Austria, Switzerland, Italy, and Slovenia. The samples were taken at 1400 m level and at other different heights (MONARPOP, 2008). The data analysis showed that chemicals like Dieldrine, Hexachlorobenzene, p,p'-DDT and p,p'-DDE showed the maximal contamination of the chosen soil samples. A further and more sophisticated method, also available in the newly developed Hasse software named PYHASSE the similarity analysis revealed that the height where the soil samples were taken has some kind of influence on the ranking of the pollutants (Voigt, 2008).

To round up the very complex topic of environmental protection, sustainable development, and risk management focussing environmental chemicals some new activities will be mentioned, like the recently founded Working Group 'Risk Management' of the Technical Committee 'Environmental Informatics' of German Society of Computer Science, GI) chaired by Alberto Susini (Working Group Risk Management, 2008). The use of non-testing methods in the risk assessment of chemicals is summed up in a special issue of the scientific journal QSAR & Combinatorial Science, edited by Andrew Worth (2008). The journal Environmental Science and Pollution Research dedicated a subject area to the topic "Risk Assessment and Management, Health" (Young, 2008).

LIST of ABBREVIATIONS

AF: Assessment Factor

DSS: Decision Support System

EDSS: Environmental Decision Support System

EUSES: European Union System for the Evaluation of Substances

GIS: Geographical Information Systems

GUI: Graphical User Interface

MONARPOP: Monitoring Network in the Alpine Region for Persistent and other Organic Pollutants

PCBs: Polychlorinated biphenyls

PEC: Predicted Environmental Concentration

PNEC: Predicted no Effect Concentration

POPs: Persistent Organic Pollutants

RCR: Risk Characterisation Ratio

REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals

RMM: Risk Management Measure

SSD: Species Sensitivity Distribution

TGD: Technical Guidance Document

REFERENCES

European Commission, EUSES 2.0, the European Union System for the Evaluation of Substances. National Institute of Public Health and Environment (RIVM), Bilthoven, The Netherlands. Available through the European Chemicals Bureau, Ispra, Italy, 1998.

European Commission, Technical guidance document in support of the commission directive 93/67/EEC in risk assessment for new notified substances and the Commission regulation (EC) 1488/94 on risk assessment for existing substances. www.ecb.it, 2003.

European Commission, C Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/ 45/EC and repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/ 105/EC and 2000/21/EC. Official Journal of the European Union, L 396/1, 2006.

Hansson, S. O., Ruden , C. Evaluating the risk decision process, *Toxicology*, 218: 100-111, 2006.

Jager, T, Vermeire, T.G., Rikken, M.G.J., van der Poel, P., Opportunities for a probabilistic risk assessment of chemicals in the European Union, *Chemosphere*, 43, 257-64, 2001.

Matthies, M., Giupponi, C., Ostendorf, B., Environmental decision support systems: Current issues, methods and tools Preface, *Environmental Modelling & Software*, 22, 123-127, 2007

MONARPOP (Monitoring Network in the Alpine Region for Persistent and other Organic Pollutants), http://www.monarpop.at/, 2008.

Richter, S. Nachhaltigkeit und Innovation im Chemikaliensektor: Welche Rollen nehmen Staat, Wissenschaft, Industrie und Verbände ein? *UWSF - Z Umweltchem Ökotox*. 20 (2), 86-89, 2008.

Schwarzenbach, Escher, B. I.; Fenner, K.; Hofstetter, T., B.; Johnson, A. C.; von Gunten, U.; Wehrli, B. The challenge of micropollutants in aquatic systems. Review. *Science*, 313: 1072-1077, 2006.

Steinhäuser KG, Greiner P, Richter S, Penning J, Angrick M., Sustainable Chemistry: Signal for Innovation or only Slogan? *ESPR – Environ Sci & Pollut Res.* 11 (5), 281-283, 2004.

United Nations. Report of the World Commission on Environment and Development, General Assembly Resolution 42/187, 11 December 1987.

United Nations General Assembly, World Summit Outcome Document, http://www.who.int/hiv/universalaccess2010/worldsummit.pdf, 15 September 2005.

Verdonck, F.A.M., Van Sprang, P.A., Vanrolleghem, P.A. An intelligent data collection tool for chemical safety/risk assessment, *Chemosphere*, 70, 1818-26, 2008.

Voigt, K., Brüggemann, R., Pudenz, S., A multi-criteria evaluation of environmental databases using the Hasse Diagram Technique (ProRank) software, *Environmental Modelling & Software*, 21 (11), 1587-1597, 2006.

Voigt K., Brüggemann, R., Schramm, K.-W., Kirchner, M., PYHASSE a New Software Tool for Partially Ordered Sets Example: Ranking of Contaminated Regions in the Alps, EnviroInfo 2008, Lüneburg, September 10. – 12., Proceedings, Shaker-Verlag, 2008, accepted.

WIKIPEDIA, Sustainability, http://en.wikipedia.org/wiki/Sustainability (06.05.08), 2008.

Working Group Risk Management, http://www.iai.fzk.de/Fachgruppe/GI/akrisk.html, 2008.

Worth, A., Computational Assessment of Toxicity and Environmental Fate, Special Issue, *QSAR & Combinatorial Science*, 27 (1), 2008.

Young, A., *Environmental Science and Pollution Research International*, Springer Verlag, 2008.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Research Needs in Water Modelling

Gustaf Olsson
Professor in Industrial Automation
Lund University, Sweden
Gustaf.olsson@iea.lth.se

Today's process engineering systems are of ever-increasing complexity. This is also true for water and wastewater systems. The processes need to be better understood in order to achieve an improved final product quality while ensuring a safe, economic and efficient plant operation. Advanced water and wastewater engineering systems are, inherently multidisciplinary. The processes are composed of interacting subsystems of parts from different engineering disciplines, requiring an integration of chemical, bio, mechanical, electrical, control, and process engineering. This also means that design of proper control laws for large systems almost always requires a well developed process model.

Many models are developed to summarize and encapsulate our current understanding of the processes. Well-known examples of this are the Activated Sludge Models (ASM) no 1, 2 and 3 and the Anaerobic Digestion Model (ADM), published by task groups in IWA. These models describe the kinetics and reactions taking place in the biological reactors in wastewater and sludge treatment systems. However, they do not describe the interaction between the plant unit processes, the hydraulics, the solids/liquid separation and many other phenomena in a complex plant.

Many models, including the ones mentioned, are used for design. Increasingly such models describe the dynamical response to load changes. It is obvious that these models cannot be verified until afterwards, since the plant does not yet exist. A reliable system operation also requires that faults and failures are detected automatically at an early stage. This requires other type of models, to support the detection process. Such models are part of the operator guidance and support system and should give a basis for manual decisions, where operating changes are performed by a human operator.

In most cases models are represented numerically as differential or difference equations. Increasingly we also see models based on rules – sometimes numerical and sometimes logical or linguistic - and they often aim at mimicing the operator or human behaviour.

Modelling other phenomena

Modelling water and wastewater processes and systems involve many kinds of models on top of the biological reaction models, mentioned above. The hydrodynamics is often assumed to be quite simple and mostly the biological reactors in wastewater treatment systems are represented as either a series of completely mixed reactors or as a plug flow tank. Incomplete mixing may often be a reason for significant modelling errors.

In solids/liquid separation there are other sources of incomplete knowledge and of modelling difficulties. One challenge has to do with the modelling of the relative settling velocity of the solids in the liquid. This is related to the floc forming process in the biological reactor. The other challenge is related to the hydrodynamics. Very often it is assumed that the solids/liquid separation is represented by an ideal separator or as a one-dimensional system, where hydrodynamics is not considered. In reality, however, the liquid forms significant flow patterns depending on the influent flow rate and on the boundary conditions of the tank.

Sludge bulking is a well-known phenomenon that causes large operational problems. Still we have quite an incomplete knowledge about the mechanisms. One is the formation of filamentous organisms. We know that this is related to the availability of oxygen and food. Still we lack reliable models of the floc formations in combination with the behaviour of the solids/liquid separation with both filamentous and floc forming organisms present.

Wastewater treatment presents interesting combinations of microscopic and macroscopic challenges. We can only manipulate the system with macroscopic methods, such as flow rates, air flow rates, recycles, dosing carbon sources and chemicals, etc. Still the consequences can often be seen at the microscopic level, where the organism behaviour is determined by our plant manipulations. As an example we can favour simultaneous nitrification-denitrification depending on the availability of dissolved oxygen and organic carbon.

A major challenge for the system wide control is to predict the flow rate and concentrations at important points in the sewer system some time ahead. Dynamic flow predictions should at the same time be used to minimise flooding problems of streets, basements, etc. This will also include the prediction of the influent load (including both flow rate and concentrations) to the wastewater treatment plant. Such a tool is still not generally available. However, some different tools are available and have been applied.

In a water distribution system the pressure in the water pipes is seldom constant. Instead, it is changing every time the consumption changes, or when a pump starts or stops or a valve opens or closes. Another important reason for pressure changes is small or large leakages. If there is a sudden burst due to a pipe break there will be a pressure wave propagating along the pipe. High frequency measurements of the pressure in combination with low frequency pressure and flow measurements can give significant information not only for the detection of a leakage but for the location of it, both in single pipes and in networks of pipes.

Models for detection, estimation and control

The tasks of an integrated supervision platform are many fold: plant-wide monitoring, database and knowledge base maintenance, fault detection, fault diagnosis and control. Failures or malfunctions of process components or instrumentation increase the operating costs of the plant. Gross failures, such as accidents, massive hydraulic overload or toxic influent in a wastewater treatment plant or large bursts and leakages or toxic spills in a water distribution system, are definitely more serious.

Modern process engineering requires prompt detection and classification of process anomalities, which would minimize the overall repair time by assisting operators and engineers in the diagnosis of system degradation. A specialized field called condition monitoring and diagnosis is devoted to determine the life span of operating devices. However, some devices tend to fail abruptly without showing any symptoms of warnings. Moreover, complex control laws often compensate for and conceal the faults. That is why real time process supervision assumes a central role in timely fault detection and isolation of complex systems.

Generally speaking we need models as a basis for controller design. However, for many processes in the water and wastewater industry a control engineer have used the knowledge of the process and its underlying physics to design very simple controllers, such as PID controllers. Analysis methods can be combined with an empirical approach of the problem to design controllers with an acceptable level of performance. During the last decades more powerful control techniques have been developed and used in various industrial sectors in order to increase their productivity. A typical example is that of model-based predictive control (MPC) in the chemical industry, where improving the pureness of some products by a few percent can yield very important profits. An interesting aspect of MPC is that it was born more than 30 years ago and developed in the industrial world, based on a very pragmatic approach of what optimal industrial control could be. It only began to arouse the interest of the scientific community much later.

Nearly all modern optimal control design techniques rely on the use of a model of the system to be controlled. As the performance specifications became more stringent with the advent of new technologies, the need for precise models from which complex controllers could be designed became a major issue, resulting in the theory of system identification. The theorists of system identification quickly oriented their research towards the computation of the 'best' estimate of a system, from which an optimal controller could be designed, using the so called certainty equivalence principle: the controller was designed as if the model was an exact representation of the system.

However, a model is never perfect, with the result that the controller designed to achieve some performance with this model may fail to meet the minimum requirements when applied to the true system. Robust control is an answer to this problem. It uses bounds on the modelling error to design controllers with guaranteed stability and/or performance. During the last twenty years many efforts have been accomplished in order to bridge the gap between robust control and system identification and a new sub-discipline has appeared, called identification for control. This became a very active research area during the 1990s. It resulted in iterative schemes where steps of closed loop identification – using the latest controller designed – alternated with steps of control design, using the last developed model.

Verifying models

Verifying a model to the physical reality is far from trivial. First of all one has to define the criteria for the basis of the decision. Traditionally this has been determined based on the residuals, i.e. the different between the model output and the true plant output. Some statistics can be evaluated based on these residuals or the prediction errors.

The activated sludge models are represented by both organism and various substrate concentrations; each one of them is a state in the model. The organism concentrations can not be measured on-line and still it is very desirable to know them. In principle this can be accomplished by state estimation, where on-line measurements are combined with a process model to find out the current values of the unknown states. A lot of attempts have been made and many estimation methods are available. Most of them are variants of the famous Kalman filter. A major problem that is often overlooked is the model accuracy. A model error in the filter will cause estimation errors.

Another challenge is to identify model parameters, given on-line (and noisy) measurements. Typically the activated sludge models have many parameters and not all of them can be estimated from plant measurements. Given a model having "perfect" measurements it is possible to find out the identifiability of some of these parameters. However, with more noisy measurements the identifiability is often in doubt.

To track the current process operational state via the instrumentation is called monitoring. In a sophisticated treatment plant there is a huge data flow from the process. More instrumentation and new instrumentation development will further provide more data. Unlike humans, computers are infinitely attentive and can detect abnormal patterns in plant data. Information technology can be used more to encapsulate process knowledge, i.e. knowledge about how the process works and how to best operate it.

Model complexity vs. accuracy

When characterizing models we talk about both complicated and complex models. In many processes there are complicated ingredients, like nonlinearities, time varying parameters and time delays. When modelling large, integrated water systems the complexity is of another kind. They include several phenomena of biological, physical and chemical nature. They include a large number of states. The systems are often stiff, i.e. the ratio between the fastest and the slowest response times are several orders of magnitude. Measurements are often scarce, which means that the identifiability and model verification is difficult. Furthermore, the systems are often widespread geographically.

A lot of research is being spent on integrated modelling of systems. Already during the 1970s and 1980s large and comprehensive water models were developed. Relatively few measurements were gathered to fit the models to data and often the model was fit to the data. There are many examples to show that such models perform only marginally better, or sometimes worse, than simple models. The lesson to learn is that the accuracy of a model does not always increase if the number of states increases. If the purpose of the model is to find an accurate relationship between manipulated variables and measurements, then a small dimension model may be satisfactory. The parameter identifiability problem is always present in large systems. This means that several sets of parameters may satisfy the data set. Therefore we may never be confident that the model will predict correctly, once we are outside the operating range of the data set. However, often we do not need to know the exact answers in terms of amplitudes and response times. We wish to understand the various interactions between phenomena and illustrate qualitative behaviour. The human has often difficulties to understand couplings, especially if there are a lot of internal feedback loops in the system.

Large systems are not only a collection of equations for different mass balances. The crucial question is often to formulate the purpose of the model, the relevant time scale, and identifying manipulated variables. In particular, for a plant wide model we need to find a systematic and structured way to coordinate all control actions.

In the full presentation more examples and references will be given about research results and needs.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Artificial-Intelligence-based Modeling for Environmental Applications and Decision Support

Peter Struss

Tech. Univ. Munich, Comp. Sci. Dept., Garching, Germany (struss@in.tum.de)

Keywords: Modeling; Artificial Intelligence; Environmental Decision Support.

1. THE PROBLEM

What we call "environment" or environmental systems is the natural world (or sections thereof) including humans and the artifacts they create. Environmental systems are complex systems, comprising a large number of interacting objects, substances, and processes which are subject to different specialized areas of science, such as biology, chemistry, hydrology, meteorology, and, when it comes to human activities, sociology, economics, even psychology. If we want to intervene in order to achieve certain goals – avoidance of environmental damage, sustainable development of resources, preservation of species – we rely on knowledge about the respective systems and the various relevant phenomena they comprise, both scientific knowledge and traditional knowledge and experience. We need a **model**, a model in a general sense (not simply a mathematical simulation model): a formal representation of concepts and interdependencies that allow explaining what we observe and predicting future developments and the impact of interventions.

Research in the different fields has contributed lots of relevant results, very often as reports, empirical data, and scientific papers, but also delivered executable models of certain systems at different scales, from small water catchments to global weather patterns. Models are accepted as an important means for validating or refuting scientific hypotheses. The corpus of available knowledge and well-supported hypotheses has increased tremendously over the past years. We should be in a much better position to understand environmental problems, develop possible solutions, take well-founded decisions and perform successful interventions.

However, we face a number of fundamental and hard problems that arise from the – inevitable – fact that all relevant scientific results and also other elements of knowledge and insights

- deal with only a limited number of aspects relevant to an environmental system or problem, and
- have been obtained in a particular context.

In order to exploit these results, we have to combine them to "obtain a more complete picture", to form a holistic model of a system. This usually requires crossing the borders of different scientific fields, which is difficult. It also requires assessing whether and which part of certain results can be transferred to a different context (chemical reactions will remain the same in different places, but whether their necessary ingredients are present or whether there is another compensating reaction may have to be established explicitly for each place). This can be stated as the problem of **integrating knowledge** elements and the problem of **context-dependent knowledge** and its re-use and adaptation.

If the results take the form of models, e.g. numerical simulation models, these general problems show up as various difficulties, such as

- Contradictory results and comparison of models. It is often hard to compare two models and identify the origin of discrepancies between the results in a rigorous and formal way (beyond the, usually informally stated or implicit, assumptions of the originators).
- Difficult re-use of models or model elements. It may not be obvious whether a model that has been validated in some area is appropriate for a similar system in another area. Sometimes, it may be known that some changes in the context violate certain assumptions underlying the model, which, hence, has to be modified. But it may be hard to tell which parts of the model are affected and need adaptation.
- Difficult integration of models. Even if we understand that certain insights underlying existing models need to be combined for obtaining a more appropriate model, integrating these models is usually a non-trivial task: each individual model has its boundaries which now have to and can be overcome, but it is not obvious where and how this can be done and how the underlying concepts can be mapped onto each other.

It is important to note that these difficulties are not simply technical issues. They are very deeply rooted in the very nature of the complexity of the systems to be modeled and in the (reductionist) way we can derive knowledge about them. “Divide and conquer” is a good research strategy which delivers insights each under a small set of aspects. However, in order to establish the basis for a broader understanding and well-founded interventions, we need to put the pieces together. We cannot divide the problem world into “improving agriculture”, “maintaining water quality”, “preservation of bio-diversity”, … and conquer them separately.

2. THE ARTIFICIAL INTELLIGENCE ANSWER

2. 1 What Is Needed?

The challenge lies in integrating knowledge from different sources and about different subjects stated in terms of models and adapting models to different contexts. Models can only be integrated if they share some concepts. Otherwise, they will stay unrelated. The problem may resemble the problem of combining data from different data bases, but it is much harder than this, because one has to identify “mappings” between different concepts of describing the behavior of certain environmental, social etc. (sub) systems, which is more complex than identifying corresponding variables in different models.

A fundamental consequence is that such models have to include a **conceptual model**, rather than some mathematical constructions (e.g. ordinary or partial differential equations) relating a number of variables. They need an explicit representation of the knowledge underlying the model, knowledge about the real-world objects whose properties are captured by the variables and their interrelations.

Integration and adaptation of models to a new scenario and/or context involves identifying and including those parts of a model that capture the aspects relevant to the new situation and dropping the irrelevant and inappropriate ones. To enable this, we need to develop **compositional models**, i.e. models that are composed of independent, elementary, combinable model fragments that can be re-used in different contexts, rather than holistic models whose structure does not reflect and preserve the structure of the modeled systems in terms of various interacting phenomena. We need to develop libraries of such elementary models that are building blocks for models of various complex systems and situations that comprise varying sets of objects and processes.

In order to be combinable with others und in different contexts, such model units from a library have to describe basic phenomena independently of others and a specific context.

They have to be **context-independent models**. If they are not, we could not be sure that such a model unit would be appropriate in a new context.

Since building and using compositional, context-independent, conceptual models is a complex task in itself, we need powerful tools supporting the development, validation, and maintenance of models and model libraries, and we need inference systems for reasoning about such models and for solving tasks like interpretation of observations, predicting the evolution of systems, or developing remedial strategies.

2. 1 What Is Offered?

The requirements outlined above are obviously a challenge to knowledge representation and automated problem solving, which belong to the key areas of Artificial Intelligence. Indeed, AI has delivered theories and systems facing this challenge, especially in research on qualitative modeling and model-based systems. The former work has developed formalisms for the kind of reductionist modeling we need, both for artifacts that are established by interconnected components and natural systems comprising different interacting processes. In this approach, **process models** describe configurations of objects or substances and elementary dynamic changes imposed on properties of the participating objects (see Forbus [2008], Heller-Struss [2001]). Examples are a chemical reaction involving certain substances, heat transfer between objects, a prey-predator relation, or the impact of an increase in agricultural productivity on the income of small farms.

Since the conceptual entities (objects and their relations, processes) are explicitly represented and subject to automated reasoning, process models are **conceptual** models in the sense stated above. They are also **compositional**. Each process model explicitly states the preconditions for a certain effect to occur. Different process models can be combined and interact via their impact on quantities of the objects in the system. The effect of one process can trigger another process by establishing its preconditions, and it can destroy process preconditions, even of itself (evaporation may finally reduce the amount of liquid to zero). The modeling formalism cannot guarantee **context-independent** models, because it cannot prevent a modeler from hard-wiring certain assumptions in the model. But it provides a construct, which is quite fundamental and lies beyond other representation of changes, for instance, in ordinary differential equations. An individual process model in a library can only state its own contribution to dynamic changes, but must not assume the non-existence of other, processes that could amplify, compensate for, or overrun this contribution. As a consequence, its impact cannot be described by a differential equation, which deterministically fixes a first-order derivative. The impact of an individual process can only be stated as an **influence** that determines the derivative only in combination with all other processes that have an impact on the same quantity.

The models developed in this field of AI satisfy the requirement for context independence also in a more specific sense, namely in refraining from implementing a particular computational direction and order , which may vary with the task, available data etc.

Models do not have a value unless they are used for some purpose. The field of model-based systems has developed theoretical foundations and systems to perform **problem solving** exploiting models (see Struss [2008]). While most of this work is concerned with artifacts and based on component-oriented modeling, some of the basic ideas and techniques have been transferred to the environmental domain and process-oriented modeling (e.g. Heller-Struss [2002]). Different tasks can be formalized in logic and supported or automated by inference engines that exploit a process library (see Struss [2004]). For instance, **situation assessment** based on a set of observations (and, perhaps, assumptions) can be formalized as the task of composing a model from the library that is (minimal and) consistent with the observations. Since such a model contains the conceptual layer, it also forms the basis for providing **explanations** of the observations, namely in terms of the processes and their preconditions that contribute to certain observed effects. The resulting model can be used for **prediction** of the potential future behavior of systems. **Remedy proposal** in unwanted situations can be formalized as extending a model of this situation by human-controlled processes such that the extended model becomes consistent with a formulated goal or healthy development. All these inferences crucially depend on the

mentioned properties of the models: explicitly represented concepts are the basis for automated reasoning about the model, and compositionality and context-independence allows for the automated composition and revision of models, without the intervention of a modeler. Furthermore, the building of model libraries itself can exploit the formal theories and techniques, e.g. by checking for inconsistencies among model fragments or between models and data, which supports debugging and modifying, or even discovering models.

2. 1 The Vision

What we are proposing is more than yet another application of the modeling technology developed in AI. Rather, it provides the starting point for a development that aims at a major qualitative step in the research on environmental problems and decision support. The grand vision is that **research** in relevant areas does no longer produce results only in terms of reports, scientific papers, and collections of empirical data. Instead, it **produces models**, more specifically model fragments as contributions to a large general library of phenomena that are relevant to a subset of environmental issues. This way, one cannot only read and understand the results obtained by other researchers (or from experience and traditional knowledge); these results are also incorporated in a set of new model fragments that are ready for being integrated in existing models, replacing refuted old models or establishing an alternative hypothesis. The minute they are published this way, the results would immediately be available to other researchers, to enhancing existing models and also to checking their validity in a different context. Comparison of rivaling hypotheses would become much easier, because alternative models could easily be generated by replacing well-identified model fragments. Wouldn't this not only ease model building itself, but speed up scientific progress in the relevant disciplines?

Of course, a lot work remains to be done. The field of model-based systems needs to solve a number of open theoretical and technical problems. Not only simplistic qualitative models need to be composed and exploited, but also numerical or semi-quantitative (e.g. interval) models. The integration with other modeling approaches, e.g. finite elements analysis, has to be established. In general, stronger spatial representations need to be incorporated. Also temporal reasoning has to be included, adding a dimension of complexity, e.g. in stepping from situation assessment to "evoluation assessment" and from remedy proposal to complex "remedy planning".

However, for environmental researchers, there is no use in simply waiting for the model-based technology to become more mature and powerful. It will never, if not exposed to and challenged by the application domain in real contexts. The research relevant to environmental systems has to contribute its own major share to the solution. An ontology has to be developed that comprises the basic concepts as a foundation for being able to state model fragments in a coherent way and for combining them. It has to be open and flexible enough for future modifications and revisions of the existing model fragments. A uniform modeling formalism with the necessary expressive power has to be designed or, perhaps, a set of specialized formalisms with appropriate interfaces between them. It is a huge research project in itself, and it is certainly overwhelming when attempted in a comprehensive way. However, for certain focused areas (water quality as an example) the time and the technology appear to be mature to start such an enterprise.

REFERENCES

Forbus, K.., Qualitative Modeling. van Harmelen, F., Lifschitz, V., and Porter, B. (eds.). *Handbook of Knowledge Representation*, Elsevier, 2008.

Heller, U., Struss, P., Consistency-Based Problem Solving for Environmental Decision Support. *Computer-Aided Civil and Infrastructure Engineering, 17*, 79-92, 2002.

Struss, P., Artificial Intelligence Methods for Environmental Decision Support. *e-Environment: Progress and Challenge*, ISBN: 970-36-0215-0, 1- 14, 2004

Struss, P., Model-based Problem Solving. van Harmelen, F., Lifschitz, V., and Porter, B. (eds.). *Handbook of Knowledge Representation*, Elsevier, ISBN-13: 978-0-444-52211-5, 395-465, 2008.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A National Water Resources Data and Prediction System for Australia

R M Argent[ab], R A Vertessy[b], L. J. Minty[b], B.J. Stewart[b]

[a] *Department of Civil and Environmental Engineering, The University of Melbourne, Australia*

[b] *Water Division, Bureau of Meteorology, Australia (R.Argent@bom.gov.au)*

Abstract: Australian water resources have a turbulent history of management, arising from a combination of jurisdictional and physical system characteristics. The position of state boundaries along rivers and across landscapes, and the imposition of state-based management regimes, have resulted in the water resources of large river basins being assessed and managed in significantly different ways. In addition to this, Australian water resources exhibit some of the world's highest variability in flows, leading to large uncertainties in prediction, and management, of the resource. This situation has been exacerbated in recent years as precipitation has diminished and become more variable, enhancing Australia's reputation as a land of 'droughts and flooding rains'. Under various future predictions of climate it is forecast that Australia's water resources will become more scarce and less reliable. In this situation the keys to improving future management of Australian water resources are to monitor the water resource, understand the interaction of water system components of demand and supply and the characteristics of the system, to predict the future state under a variety of conditions and management regimes, and to use system monitoring to improve such understanding and prediction. To achieve these ends the Government of Australia approved, in 2007, the development of an Australian Water Resources Information System (AWRIS). This system is designed as a national end-to-end system, taking data from long term and real-time monitoring networks, through a national data collection and collation centre, to web-based provision of water data, information, prediction and forecasting tools, and predictions. This paper describes the vision, and architecture, implementation and development issues of the AWRIS (http://www.bom.gav.au/water).

Keywords: AWRIS, Australian Water Resources Information System, Water Information, Data interoperability, Web Services

1. BACKGROUND

Water is a key area of concern for many societies and countries, particularly in the face of climatic change and increasing climatic variability. Australia is one country that is highly susceptible to these influences for, although having an overall low population density, much of land is arid, with a highly urbanised society and a significant dependence upon irrigated agriculture.

In Australia, water resources are largely managed or overseen at a state level, with various levels of management also occurring at local, catchment, region, system and basin scales. As a result, it is difficult to obtain a whole and consistent understanding of the nature of the resource, and even more difficult to make reliable forecasts of future water availability. There are over 200 agencies, authorities, corporations and organisations that manage part of the resource, and which therefore hold some of the pieces of the water resources puzzle. Effective and holistic water resources management to meet competing needs is thus

impeded, as is water resource policy development and decision making on infrastructure investment.

Particular aspects of the Australian water resources system provide an impetus for improved understanding and management. For example, over the period since 1997 much of the country has been subject to a long term and widespread drought and although drought is not uncommon in particular regions or states, the coincidence and longevity this time has sustained water resources debate at a national level for a significant period. Additionally, significant surface and groundwater systems are positioned across multiple jurisdictional boundaries, increasing the difficulties in monitoring and understanding systems from a national perspective. This issue is further complicated by a strong reliance upon irrigated agriculture as a food source, which may at times be competing with other users in other jurisdictions for the diminishing resource.

In this climate and situation the importance of understanding the national water resource increased to the point where a series of coordinated actions were required at a national level. These activities include monitoring the water resource across the full spectrum of water processes, understanding the interaction of the system components in both space and time, investigating the characteristics of the natural and constructed water systems and the broad range of interactions occurring between the two, using such monitoring and investigation to assess the current state of the resource, and, finally, to provide forecasts of future states of the system. These needs have resulted in the establishment of a project to develop an Australian Water Resources Information System (AWRIS).

2. THE AUSTRALIAN WATER RESOURCES INFORMATION SYSTEM

The concept of a national system for coordinating and managing water data and information has existed in Australia, in various guises, since the 1980's. At that time, and following a significant national level drought in 1982/83, the federal government invested significant resources to expand the national surface water monitoring scheme. With the considerable advances in information science and technology over the intervening decades, the concept re-emerged in the 00's in the form of a national network for water observations, akin to surface hydrology monitoring and management systems in countries such as Korea and USA. In 2007, development of this observation network (O'Hagan et al., 2007) informed planning for the AWRIS.

2.1 The Concept of the AWRIS

In its simplest form the AWRIS will be a data system for housing all of Australia's critical water data. It will be supported by systems for effective gathering and ingestion of data from multiple sources, in multiple forms, and the spatial and temporal relationships between these data will be represented in ways that provide effective support for querying and analysis to meet a broad set of data and information needs. The system will be constructed to provide robust and reliable, nationwide information on water availability, water quality and water usage, and will be the authoritative repository for water data and reporting in Australia.

A key role for AWRIS will be transformation of data into information, and the system will therefore deliver not only data and information, but also reports and tools to enhance the decision-making capabilities of users engaged in:

- Policy development;
- Planning;
- Operations;
- Public enquiry, and
- Research.

AWRIS is conceived as having three core components (Figure 1) - a central or 'enabling' framework for harvesting and organising water data, an external face providing public access to water data, tools and information, and an internal face providing analysis and reporting tools for delivery of key regular national water reports, such as Australian Water Resource assessments (AWRs) and National Water Accounts (NWAs).

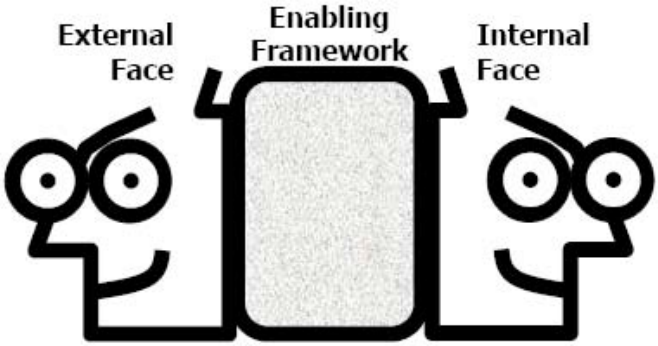


Figure 1. AWRIS Core Components – a central enabling framework delivering services to external and internal clients

As implied by the name, AWRIS will be an information system, providing end-to-end processing and delivery of data (Figure 2). The workflow starts with harvesting and checking data from primary collecting agencies, proceeds through an organising step that imposes a strong geo-spatial relationship formulation upon the data (in the form of a 'geofabric'), on to an analysis and interpretation component, to final internal and external delivery of customised reports. User needs may require all these steps or possibly only a few, such as in the provision of primary data to end users.

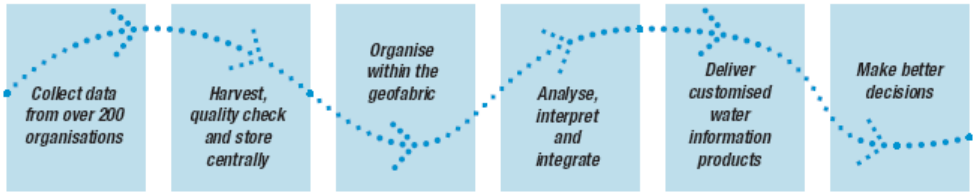


Figure 2. The AWRIS data harvesting, organisation, analysis, reporting and delivery pathway

The operation of AWRIS will occur through, and be enhanced by, data exchange standards and agreements, common data models and terminology, and appropriate systems infrastructure and architecture.

2.2 Legislation

Establishment of a system like the AWRIS can take place in a number of ways, such as through federated agreements and alliances, or through enablement via central legislation. In this case, AWRIS development is initially supported through a two-year (2008/09) grant under the Raising National Water Standards program of the National Water Commission. This 'phase 1' activity is further supported over a 10-year horizon through the national water plan.

The primary associated legislation for AWRIS is the Water Act, 2007, and in particular Part 7, dealing with water information (Government of Australia, 2007; http://www.comlaw.gov.au). Under this the Bureau of Meteorology (the 'Bureau') takes a central role, being charged with:

- collecting, holding, managing, interpreting and disseminating Australia's water information;

- providing regular reports on the status of Australia's water resources and patterns of usage of those resources;
- compiling and maintaining water accounts for Australia, including a set of water accounts to be known as the National Water Account;
- providing regular forecasts on the future availability of Australia's water resources;
- undertaking and commissioning investigations to enhance understanding of Australia's water resources;
- issuing National Water Information Standards;
- giving advice on matters relating to water information, and
- any other matter, relating to water information, specified in the regulations

These activities significantly expand the Bureau's role in water information, from traditional areas of climate monitoring and flood forecasting. AWRIS is being built and will be owned by the Bureau, consistent with this expanded role.

2.3 Providers of Water Information

A key aspect of the Water Act 2007 is that it *requires* 'persons' to provide water information to the Bureau. In this context, 'persons' refers to over 200 agencies, authorities and other bodies collecting and holding relevant water data. An indication of the breadth of this endeavour, and possibly the complexity of the data harvesting and collation task, is found in the list of persons in Regulations under the Act. These include leading (eg water) and other (eg agriculture or environment) national and state water agencies, hydroelectricity organisations, owners and operators of major storages, rural and urban water utilities, catchment management authorities, and collectors of rainfall and/or stream height data (e.g. councils). In a country such as Australia, with three-tiered government, this spread of authority and control is not surprising when dealing with a commodity as important as water.

2.4 Data Categories Being Supplied

The complexity of obtaining data from a range of agencies, with different skills, capabilities and data systems, is compounded when considering the information to be provided. These data, listed in Table 1, are required so that comprehensive analyses can be undertaken, in line with the legislative needs for national water resource assessments and accounts.

Table 1. Types of water information to be provided to the Bureau

Category number	Description of category information
1	Surface water resource information
2	Groundwater resource information
3	Information on major and minor water storages
4	Meteorological information
5	Water use information
6	Information about rights, allocations and trades in relation to water
7	Information about urban water management
8	Information about water restrictions
9	Water quality information
10	Descriptive and reference information about water information in other categories

3. CONCEPTUAL WORKFLOW FOR THE AWRIS

As indicated in Figure 2, a reasonably straightforward process is defined for operation of the AWRIS. These are expanded in Figure 3, along with an indication of feedbacks that will need to operate, and which will naturally complicate the process.

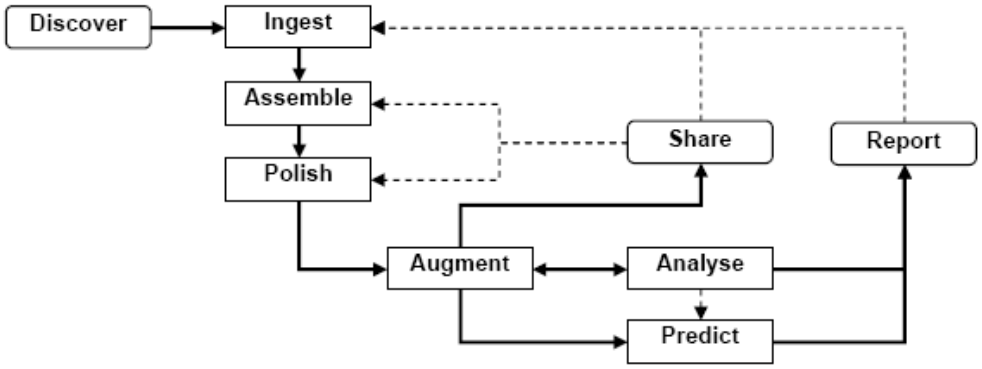


Figure 3. Data processing through the AWRIS, where solid lines show primary flows, and dashed lines represent feedback mechanisms

In Figure 3, data from multiple sources, and in multiple formats, is ingested into the central data system, having possibly been 'discovered' via web services methods (e.g. Horak et al., 2008). These data are then assembled and polished into a consistent geo-spatial and temporal framework through use of appropriate data models and quality control checks. Data augmentation can then be undertaken by inferring additional water balance components. These data are then made available to users, and are also analysed and used in predictions to provide a variety of internal and external water resources reports. In terms of feedbacks, the needs and outputs of analysis will likely inform both the augmentation and prediction processes. For example, a water balance analysis may require prediction at places in the system, or possibly provide model output that can further augment the data pool. The gaps, errors and glitches identified in data and information by users through sharing and reporting will provide a valuable response mechanism that will inform not only the assembly and polishing processes, but also feed back through the ingestion step to the original data providers. This latter point raises a key problem identified by jurisdictional representatives of the data providers, and also in architectural discussions, of the 'point of truth' for the data. As ever with data management and publishing by multiple sources with a range of capabilities, there is a requirement for clear procedures and mechanisms by which data updating (eg fixing errors or changing quality codes) is undertaken.

4. ANALYSIS, SHARING, PREDICTION AND REPORTING REQUIREMENTS

Whilst the complexity of the data ingestion and assembly tasks is significant, the current business needs for the data held in AWRIS are reasonably clear. These reflect the first four points noted from the Water Act (Section 2.2), and can be described as:

- Make all data available publicly, with the exception of data deemed 'not to be in the public interest', or data which 'expressly identifies a persons water use' (Government of Australia, 2007).
- Support the undertaking of water balances (e.g. Zhang et al., 2005; Welsh et al., 2006) at various scales and complexities of systems, the dominant components of which may include storages, incoming and outgoing volumes, groundwater stores and level changes, reuse and recycling of water, diversions and extractions.
- Support production of water accounts (e.g. Turner et al., 2007; Kirby et al., 2008) for various systems, and the National Water Account.

- Support production of water resource assessments for various reporting units, such as catchments, basins, states and systems.

It is useful at this point to return to the concepts described previously, whereby the AWRIS includes not only the central framework for data management but also the analysis tools and reporting systems underpinning the internal and external faces. The development of these has significant implications for the system architecture and geo-spatial and temporal data systems.

5. ARCHITECTURE, DATA SYSTEMS AND INFORMATION DELIVERY

Development of the AWRIS by the Bureau began early in 2008, with attention focussed across the spectrum of data ingestion, assembly, sharing and reporting, supported by a strong focus on stakeholder engagement. Given the nature of the newly legislated data delivery requirement for 'persons' holding water data, this latter point is essential.

AWRIS development activity is being driven by a 2-year Phase 1 project plan, with the purposes of delivering 'useful functionality' in the short term, and a 'reliable and relevant system' for the long term. Development is occurring in four key areas – data, architecture, 'geofabric' and products.

5.1 Data Interoperability and Management

The first data harvest under the new legislation is currently timed to occur in September 2008. This includes only a portion of the total data pool, and only a sub-set of the data providers, as the Regulation stipulate that different 'persons' need to deliver data at different times and frequencies, with frequencies range from daily (or possibly finer for data sourced automatically via e.g. web services), to annually for types such as irrigation system allocations.

With such a varied data stream, understanding of the data types and flows is important. The AWRIS team is thus undertaking a data inventory to determine the data types held, although this is difficult due to the sheer range of authorities and responsibilities. It is recognised that 'discovery' will be an ongoing process through the first year and beyond – not only with respect to data held, but also the formats used.

Some of the larger data collection organisations, with larger data holdings, are well set up in terms of data management, using robust commercial data systems. We are currently undertaking a first pilot data harvest with these groups to identify problems and issues that may arise down the track, and to create solutions to serve future needs.

5.2 Solution Architecture

In a system such as AWRIS the architectural needs are many and varied. In the initial phase, focus is on development of a solution architecture that can support the highest priority and most immediate requirements, whilst providing a platform for future development beyond Phase 1. Principles of operation in this area therefore include the adoption of an agile and evolutionary approach built around components and embodying some of the fundamentals of service oriented architectures, such as loose coupling, autonomy, statelessness and reuse. A number of requirements identification activities have been undertaken to inform architectural design in the period leading up to 2008, and early activity is continuing in this area, focussing particularly upon both functional and non-functional requirements of internal users.

5.3 Geofabric

Undertaking water balances and other water resource analyses across different spatial and reporting units (eg catchment, system, basin, state) requires a consistent approach to the

spatial relationships between data. Support for this in AWRIS will be done through a 'geofabric' - a concept that is easy to grasp but hard to define. Through the developmental phases of the AWRIS the geofabric has been described variously as:

- a sophisticated three-dimensional representation of all landscape elements and their interrelationships that impact on the water cycle. It explicitly describes the spatial relationships and functional connectivity between geophysical landscape elements from a water balance perspective
- a baseline suite of common data sets that allow shared understanding of definition, naming and characterisation of landscape elements
- a specialised Geographic Information System that describes spatial relationships between important hydrologic features.
- a single, consistent, national geospatial framework for hydrological features, spatially coincident and topographically enforced to a 1" national digital elevation model

Thus, the geofabric not only encompasses fundamental geospatial data, but also i) the description of features, ii) the hydrological data associated with those features, and iii) the relationships between and amongst the features and their data from a hydrological perspective. As an example, a feature such as a monitoring point may be part of a surface runoff system, a groundwater system, and an engineered water supply system. It will thus have a different hydrological 'meaning' and information role depending upon the system of interest. The geofabric will support these multiple identities and roles for such monitoring points. Initial activities in defining and designing the geofabric have focussed upon conceptual and abstract design (eg geometry v network v feature based approaches), the utility of existing hydrological data models and tools (eg ArcHydro), and development of protocols for testing and comparing alternate models within the geofabric.

5.4 Products

One of the advantages in constructing a national water information system is that many users can readily identify and articulate their needs for such a system. Thus the opportunity exists from very early in the project to meet some of these needs, particularly in the area of data exploration and delivery. Early developments in the area of 'products' are therefore focussing upon provision of services by which users can explore, become familiar with, and obtain data of interest in a variety of raw and derived formats and forms. A data exploration system and 'dashboard' of data reporting tools are being developed to this end. An example of a derived product, in the form of a national dam level reporting system, is also being developed as a both a functional report and a testing ground for other products.

6. FUTURE PROGRAM

This paper has provided a brief introduction to the vision of the AWRIS, some of the issues in architecture, development and implementation, and an overview of current activities. There is a risk in considering the design, development, implementation and operation of such a broad system that the complexity of the data harvesting, processing, analysis and reporting will outweigh the elegance of the concept. To overcome this, we have chosen a multi-faceted approach, built around the principles of 'adopt, adapt, develop', and encompassing a range of strategies. Firstly, we have selected a team with a broad range of experience and contacts in the areas of spatial and temporal water information systems. Secondly, data management and reporting systems of this complexity have been built before and so we aim to understand and adopt technologies and solutions that are already well proven, or which are at the forefront of current operational thinking. For this, we will draw upon international linkages and solution developers and providers to inform our selection of methods and technologies. Thirdly, we have established a research alliance with Australia's leading researchers in water information. These will provide a pool of

talent upon which we can draw in future to adapt the technologies and methods initially adopted, and to extend and develop these based upon longer term research investigation. Finally, we are taking a phased approach to development, accepting that will make mistakes as we go along, and that the things learned from these provide the 'experience' we draw upon in the long term.

ACKNOWLEDGEMENTS

Development of AWRIS is supported by the National Water Commission via the Raising National Water Standards program, and by the Government of Australia under the national water plan.

REFERENCES

Government of Australia, 2007. Water Act 2007, Government of Australia, p. 235.

Horak, J., Orlik, A., Stromsky, J., 2008. Web services for distributed and interoperable hydro-information systems. Hydrology and Earth System Sciences, 12, 635-644.

Kirby, M., van Dijk, A. I. J. M., Mainuddin, M., Peña-Arancibia, J., Guerschman, J.-P., Liu, Y., Marvanek, S., McJannet, D. L., Paydar, Z., McVicar, T. R., van Niel, T. G., Li, L. T., 2008. River water balance accounting to evaluate model adequacy and uncertainty in climate and development scenario assessment. In: Lambert, M., Daniell, T. and Leonard, M., (Eds.), Proceedings of Water Down Under 2008: Adelaide, Engineers Australia, p. 1992-2003.

O'Hagan, R. G., Atkinson, R., Cox, S., Fitch, P., Lemon, D., Walker, G., 2007. A reference model for a water resources observation network. In: Oxley, L. and Kulasiri, D., (Eds.), MODSIM 2007 International Congress on Modelling and Simulation: Christchurch, Modelling and Simulation Society of Australia and New Zealand, p. 1145-1151.

Turner, G. M., Baynes, T., McInnes, B., West, J., Hoffman, M., 2007. Water accounting system for strategic resource management. In: Oxley, L. and Kulasiri, D., (Eds.), MODSIM 2007 International Congress on Modelling and Simulation: Christchurch, Modelling and Simulation Society of Australia and New Zealand, p. 226-232.

Welsh, W. D., Barratt, D. G., Ranatunga, K., Randall, L. A., 2006. Development of a national, land-use based water balance model for Australia. In: Voinov, A., (Ed.), iEMSs 2006. Summit on Environmental Modelling and Software: Burlington, USA, International Environmental Modelling and Software Society, p. 6.

Zhang, L., Hickel, K., Shao, Q., 2005. Water balance modelling over variable time scales. In: Zerger, A. and Argent, R. M., (Eds.), MODSIM 05 International Congress on Modelling and Simulation: Melbourne, Modelling and Simulation Society of Australia and New Zealand, p. 2988-2994.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Delft FEWS: A proven infrastructure to bring data, sensors and models together

P.J.A. Gijsbers[a] **, M.G.F. Werner**[a,b] **and J.Schellekens**[a]

[a] *Deltares-Delft Hydraulics, Rotterdamseweg 185, 2629 HD, Delft, The Netherlands (peter.gijsbers@deltares.nl)*

[b]*Department of Water Science & Engineering, UNESCO-IHE, P.O. Box 3015, 2601 DA Delft, the Netherlands*

Abstract: Delft FEWS is a proven real time software infrastructure for operational water management and forecasting, having been applied in several operational forecasting centres all over the world. The system has contributed to changing the paradigm of flood forecasting systems from a model centric approach to a data centric approach. The system is a time series oriented ETL-infrastructure (Extract, Transfer, Load) with the ultimate aim to provide the forecaster with relevant information in support of the operational process of flood forecasting and warning. Key to its success is the design choice for separation of data communication protocols from the content, as well as its openness to integration of other applications and data. While keeping its focus on its core tasks, the design concept has allowed the system to evolve from a stand alone application, through a client-server concept into a (web) service oriented architecture. Continued research is ongoing to explore new technologies and turn these into operational features available to operational forecasters. This paper will highlight some of our experiences in system design to combine real time data with models.

Keywords: FEWS; ETL; real time; forecasting; modelling; time series; data; sensors

1. INTRODUCTION

Delft FEWS is a proven, state-of-the-art, real time software infrastructure for operational water management and forecasting. The system is a sophisticated collection of modules designed for building an operational water management system customised to the specific requirements of an individual agency. The philosophy of the system is to provide an open shell system for managing the operational management process [Werner et al.2004]. This shell incorporates a comprehensive set of general data handling utilities, while providing an open interface to a wide range of forecasting models. The modular and highly configurable nature of the system allows it to be used effectively both in rudimentary systems and in highly complex systems utilising several simulation models. Delft FEWS can either be deployed in a stand-alone, manually driven environment, as a fully automated distributed client-server application, or as a web-service component within another forecasting system.

Over the last few years, Delft FEWS has been implemented as the prime operational system in several operational flow forecasting systems used both by national authorities in for example England, Wales, Scotland, Germany, Switzerland, Taiwan and the Netherlands, and by regional authorities in amongst others the Netherlands, Austria,

Thailand and Italy. In addition to these, several operational and pilot systems are currently being implemented in the Netherlands, USA, Singapore, Russia, Spain and in the Mekong basin. Its expansion has, however, not been constrained to a geographic expansion as more recently its range of application has evolved from a focus on the use in operational flood forecasting systems to a much wider application range for operational water management forecasting of both high and low flows, reservoir and storm surge barrier operation, groundwater management, algae bloom and water quality forecasting. In addition it is also being applied as a Water Information System, where it forms the heart of an infrastructure that focuses on automated data collection, validation, visualization, and persistent storage.
This paper will discuss the philosophy behind Delft FEWS. It will analyse the design choices that have enabled the expansion and it will address some of the lessons learned in implementation across such a wide range of domains.

2. THE PHILOSOPHY BEHIND Delft FEWS

2.1 Support the operational team

Flood Early Warning Systems, from which the FEWS acronym has been derived, have been established for many years [see e.g. Grijssen et al, 1992]. These have a primary role in providing comprehensive information to the operational forecaster in a concise and efficient manner. It is then the responsibility of the operational forecasting team to take guidance from the information provided, and when relevant issue a warning to relevant authorities and if applicable the general public. From the perspective of this role in support of the operational flood forecasting, warning and response process [Parker & Fordham, 1996], describing the system as a warning system is somewhat misleading, as the warning itself is issued by the forecasting team, having taken guidance from the real time data and the forecasting system. Within this perspective, the main purpose of a forecasting system in support of the warning process, as compared to when a monitoring system only is used, is to increase the lead time of warnings so that authorities can act upon expected conditions [see Werner *et.al,* 2005], Werner & Heynert [2006] discuss this role in more detail. Within this perspective, two key factors should be mentioned that have had significant impact on system design: (i) robustness and (ii) confidence.

2.2 Deliver robustness and confidence

As discussed, the primary role of a forecasting system in the operational setting is to provide data to be used as guidance to the forecaster in making a decision on whether or not to take action. Within this context, the confidence that the forecaster vests on the data provided is essential if that information is to be used meaningfully. Typically the confidence of operational forecasts are addressed in terms of how trustworthy observations and forecasts are, how large uncertainties in the forecasts made is and, in that same light, how skilful these forecasts are in predicting forthcoming events.
These considerations on factors that contribute to forecaster confidence are obviously important, but additional factors such as the robustness of the system itself will also need to be considered. Robustness refers to the fact that the system should offer duty forecasters the latest and best data available, given the conditions in the field. If the system fails, no or very little information may be available, and decisions may need to be taken unsupported by data. Should the system fail regularly, forecasters will naturally lose confidence in its ability to provide relevant information, and seek guidance from other sources. Much attention has been paid to this aspect in the design of Delft FEWS. Robustness has been one of the prime drivers in establishing a highly modular concept. As there is a high degree of independence of these modules, some may be allowed to fail and lead to degradation in the information provided. The modularity ensures this will not cause a complete failure of the system.

2.3 Focus on data handling and provision, not on running models

Many flood forecasting systems used to be, and sometimes still are, based on a model centric approach [Werner & Whitfield, 2007]. Having been established essentially as a wrapper around a specific model or models, the execution of the hydrological and hydraulic models seemed to be the main task of these systems and as such every item was designed for this task. Operational users, however, are not always focused on the models themselves but rather on gaining insight in a variety of data streams, ranging from tidal surge data and in-situ observations through to meteorological and hydrological forecasts. A key task for an operational information system is to bring all information together and create a consistent and reliable information base for decision making. The collection, validation, transformation, storage and visualization of data play an equal role next to the running of models. It should be stressed that efficiency in combining data from several sources in differing ways efficiently is paramount to allow the forecaster gain a rapid understanding on the basis of the data available.

2.4 Focus on essential system tasks, not on domain specific content or formats

Tasks such as data collection, validation, transformation, storage and visualization are common to many domains in and outside the flood forecasting or even the wider water domain. System-infrastructures that conduct these types of tasks are typically called ETL-infrastructures, i.e. extract data from external sources, transform this into a consistent format that can be loaded to other applications or components.

A characteristic of hydrological forecasting systems is the spatial and temporal nature of the data that can be handled. Delft FEWS has no additional domain specific knowledge to conduct all data processing tasks. Its design choice to strictly separate the water domain from the system design and to separate data/communication protocol and content is considered a major contributor to its evolution from its application to the flood forecasting domain towards its current wide range of application.

2.5 Create openness for other system components

Water authorities often have knowledge bases and computer models that have been developed over time. When adopting a new comprehensive forecasting system they will be reluctant to throw these components away, since much knowledge and experience is vested in these models and knowledge bases, which may also represent large financial investments. Incorporating such existing knowledge components in a new system that in essence provides only an operational framework is a considerable asset. Not only does reuse of existing components ensure that both knowledge and financial investments are maintained, their reuse also bolsters acceptance of and confidence in any new system. Openness in system design is therefore considered essential. It is also an asset that reduces supplier dependency for new system components (e.g. models). This need for open systems is discussed in detail by Werner & Heynert [2006], and its value is also reflected in e.g. Funke et al. [2006]. The separation of form and function also makes it easy to identify interfaces that create openness to existing system components or new system components from other suppliers.

2.6 Deliver product stability

Operational forecasting teams typically encompass a large number of qualified staff to be able to provide a continuous service. As a consequence, the stability of the system in terms of how staff interacts with it are of major concern, particularly for those systems that are

used nationally in several operational centres. The investment associated with developing forecasting and warning procedures, documenting these procedures and training relevant staff is extensive for organisations such as the Environment Agency of England and Wales [Whitfield, 2005]. Werner & Heynert [2006] have highlighted how an open system enables improvement of the forecasting and data processing capabilities with minimal impact on how operational staff uses the system on a day to day basis. As a consequence, product developers should ensure that the appearance of the system to its direct users, in particular the user interface, does not evolve too fast, as agencies will be reluctant to repeatedly send large numbers of staff on training to understand such 'improvements'. In the same line of thought, backward compatibility should be ensured and plans carefully made for removing the utilization of functionality that has been superseded by other components.

2.7 Hand back responsibility for system extension

Many agencies are all too aware of dependency on external suppliers. They often are anxious for vendor-locks, i.e. contracts and products where they are bound to one supplier for a long period over time. The ability to separate the modelling task from system configuration enables agencies to outsource the modelling work to other suppliers. Once trained in system configuration, they also can take responsibility for that configuration. In this way, they are in control again, with only minor dependency on the supplier for basic system maintenance and support. Deltares-Delft Hydraulics, the system supplier of Delft FEWS, is keen to stimulate agencies to take responsibility for their Delft FEWS application as innovation in forecasting techniques will then not only be driven by the ideas of a central supplier only, but will be complemented by all users who are actively involved in configuring and working with the system. The aim of Deltares-Delft Hydraulics is then to act as an intermediary in sharing innovation across users.

3. SYSTEM FUNCTIONALITY

In functional terms, data processing within Delft FEWS is guided by the ETL-paradigm; extract data, transform into a consistent format and load into forecasting models. These functionalities are provided by four modules that work on top of a high performance data access layer which combines relational database technology with a highly compressed data structure that can handle large sets of time series data, including spatial time series.

3.1 Extract data

The import module provides the database with a quality controlled data set, obtained from external sources, such as telemetry systems, remote sensing observations (radar and satellite) and (ensemble) weather predictions. The design of this import module has focussed on facilitating automated detection of missing values and validation against prescribed ranges and maximum rates of change.

There are currently no standards in data formats made available through telemetry systems, and as a consequence a range of data formats need to be catered for. In Delft FEWS a design pattern is chosen that requires a (small) code extension for each new format to be considered. This can be done within the Delft FEWS code base but also by registering a Java class that has been made according to the Delft FEWS import plug-in specifications. The latter option does not require the original code developer to be involved. This approach was chosen in preference to the development of a generic import class that could be configured to the specifics of a format. Experience with such generic classes has shown that these quickly become overly complex as there is in principle an unlimited combination of options. Establishing a targeted class specific to a format has

been found much more efficient and resilient. For remote sensing data and weather predictions industry standards such as GRIB and BIL/BSQ are, however, supported.

The default operation method of this import module is to fetch data at regular intervals from a disk device within the LAN. Recently, the module has been extended to acquire data from an external server (FTP or HTTP). However, the preferred approach is often to keep the import mechanism used as simple as possible, with the simple file based mechanism often being the most appropriate. More complex modes of operation, e.g. direct database queries or web service requests are available, but often not considered appropriate as for example low server response may result in time-outs or blocking of other server processes. Essentially experience has shown that tight integration of separate systems using complex data transfer mechanisms may be technically desirable, but may be detrimental to overall robustness due to dependency on multiple systems being fully available simultaneously. Within the sensor world, initiatives are being undertaken to develop standard data exchange formats for sensor data. The Sensor Web Enablement protocol (SWE), developed by the Open Geospatial Consortium [OGC, 2006] is one of these initiatives. This protocol has been evaluated in a research project co-funded by the Dutch program Ruimte voor Geo informatie. After a test implementation to query various Sensor Observation Services for observation data, we concluded that this protocol has various drawbacks when used within a high-performance operational setting as is required for forecasting purposes [Jellema, 2008]. The most important drawbacks found where:

- The reliance for SWE on GML (Geographic Markup Language) creates a large overhead in meta-data, which results in large XML-objects to be passed, while the actual data of interest, i.e. the observations, are relatively small in package size.
- The key data content within the XML-string, i.e. the observed values, is packed in a generic XML-element which has to be interpreted based on a self-described format. A dedicated, and complex, piece of code is required to parse this element in a generic way, since it cannot be handled by an industry standard XML-parser.

3.2 Transform data

Typically, the data imported needs to be fed into forecasting models. Spatial and temporal resolutions need to be adapted and gaps need to be filled to ensure that the data fits the input requirements of the model. Within Delft FEWS the functionality required for manipulation, combination and completion of data are implemented by an interpolation module and a transformation module. The interpolation module offers a wide range of spatial and temporal interpolation algorithms, while the transformation module offers a wide range of functions to transform (complete) series. The transformation module can be considered as the work-horse of the system, providing an extensive library of data transformation options. Although some of these have a hydraulic or hydrological background, most are domain independent.

3.3 Load into models using an open interface format

Once prepared, the data can be loaded into the forecast model to run. Delft FEWS uses an open interface approach to accommodate models from various suppliers. A standardized XML-format, the so-called Published Interface, is used to exchange the relevant data with the models. From the system's perspective, the model run is managed by the so-called General Adapter (see Figure 1). For each model run, the General Adapter exports the relevant model state (in native format) and boundary conditions (in XML) to the model. A tailored Model Adapter converts this data into the native model format before the model is kicked off. After execution has completed, the Module Adapter converts the results back into the Published Interface-format. Finally, the General Adapter imports these results as well as the update (native) model state file, which is stored for reuse at the next run.

By adopting this standardized and open approach, the General Adapter only needs to know the data exchange format while all model specific knowledge can be encapsulated in the model adapter. The use of the published interface standard allows such model adapters to be established wholly independent of the Delft FEWS system itself, thus ensuring independence of the suppliers of the model code from the suppliers of the Delft FEWS system, should such independence be desirable from a contractual perspective.

Model codes from a range of model developers have been integrated with the approach described. Werner & Heynert [2006] published a list of some 19 models that were available at the time, but this list has been extended to include several other codes, such as HEC-ResSim, HEC-RAS, Hec-HMS, Modflow, Vflo, TOPKAPI, SNOW17, SAC-SMA, Ribasim, Delft3D-Flow, SOBEK-WQ and Delft3D-WQ.

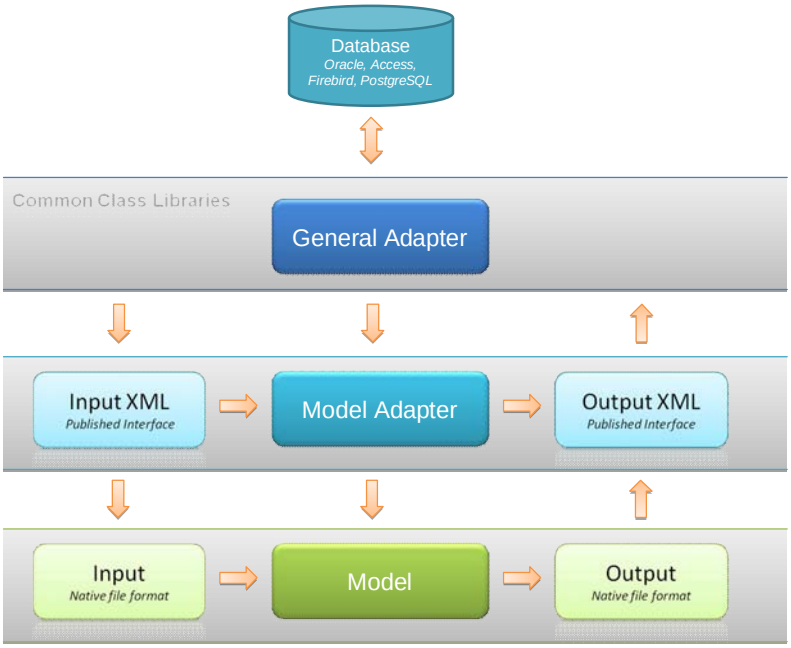


Figure 1 Open communication between Delft FEWS and Models

4. SYSTEM IMPLEMENTATION

Delft FEWS is a scalable system developed with J2EE technology. Stand alone applications utilize the Delft FEWS-shell instance as a combined GUI and computational server. This instance has its own local database. In the client-server setting, these two roles are separated. The same Delft FEWS-shell instance is then utilized on the client side as a thick client application, referred to as the Operator Client (OC) through which the data is disseminated to the forecaster, and on the served side, referred as a Forecasting Shell Server (FSS). Forecasting Shell Servers act as the computational engine which conducts all data processing tasks from importing data to running models and, if desired, producing HTML- reports. Data produced by FSSs is synchronized to a central database. From this database, data can then be distributed to the various client-applications. At the heart the system is the Master Controller (MC). This component manages the communication between all components using JMS. It contains a job scheduler to automatically initiate tasks on the FSSs. If a multiple-site system is deployed, the Master Controller ensures data replication at all sites either in duty standby mode or in replication mode to balance the load. A dedicated web-interface is available to manage the Master Controller and maintain

the system. To manage the system configuration, a separate maintenance component is available, while all system components will detect configuration updates automatically. Recently, a fully automated archiving sever has been added as well as a web server to integrate forecast reports from various forecasting agencies. The product openness has been extended with a JDBC driver to directly obtain time series or associated graphic figures. Alternatively, a web service may be requested to obtain data via the Delft FEWS Published Interface formats. The web-service can also be used to kick off a task to the Master Controller.

5. ROBUSTNESS AND CONFIDENCE

The utility of an operational system would be considered low by any operational team if the system fails when it is required most, during for example a flood event. Repeated failure of the system at such times will quickly erode any confidence in it. Such confidence can be built up by ensuring that the system runs properly with a very high availability. For example, the choice for a fat-client with a local database over a thin (web) client is guided by a desire to reduce dependency on continuos availability of high-performance internet connections. During implementation, various precautionary actions can be taken to increase system robustness, such as: ensure a continuous power supply, independent of general power failures, prevent flooding of the forecasting centre, provide back up or dual site implementation, apply proven technology, and ensure proper support arrangements. Confidence is also increased if the data provided is available, useful, frequently updated and properly flagged in terms of reliability. The latter requires thorough understanding of the forecasting process to ensure that the system does not stop should data not be available to the full extent normally expected. Fall back options to use other data sources, should be available if required.

5.1 Fall back options

Numerical simulation models are normally very strict in their input requirements. They typically cannot handle missing data in their input data sets. If data import, validation and transformation fail to provide the required input for simulation models, alternative input sources may be chosen or the model run might be skipped. In the latter case, rudimentary lookup tables or correlation functions might become the base to deliver an answer to the operational team using the best available knowledge combined with available data sources. This so-called ‘data hierarchy’ concept is crucial for system robustness as it allows alternative data sources to be used as a fallback in the forecasting process should the nominal sources be unavailable. The use of these options will ensure the continuity of the forecasting process, even if available real-time data is incomplete or inconsistent. As such robustness contributes to confidence as the operational team is aware that the system will always provide the best and latest data available, including associated quality flags.

5.2 Confidence in forecasting models

Skilful forecasts play a vital role in building confidence. Flood events often happen with model conditions that often are not very well tested due to limited data availability. Various advanced techniques, such as error correction and ensemble Kalman filtering, can be applied to improve model performance.
Methods to evaluate model performance form a key component in evaluating the performance of the forecast models. Within Delft FEWS, the concept of establishing performance indicators as an integral part of the forecast process has been introduced to compare observed and calculated time series of the same type at a given location. This accuracy can be assessed at the level of the individual forecasting modules using performance measures typically applied in module calibration (e.g. root mean square error,

Nash-Sutcliffe). In addition, the performance of the forecasting system itself can be derived to reflect the accuracy of the system in forecasting. Three types of measure are available; (i) lead time accuracy of forecast time series, (ii) accuracy of timing threshold event crossings, and (iii) accuracy and timing of peak predictions [Delft Hydraulics, 2005]. Confidence of operational teams also improves if both the false alarm rate and the number of missed events are low. For this purpose a forecasting skills concept is used. Indicators such as probability of detection and false alarm ratio are derived from a contingency table in which the number of (calculated) threshold crossings is compared against the number of observed threshold crossings.
Within Delft FEWS, both the performance indicators and the skills score table are continuously updated as a part of the operational process. Integration of such evaluation measures can be used both to target improvements where forecasting skill is considered low, as well as to increase model complexity only when proven to increase forecast skill. This allows for an optimal balance between CPU run time and the spatial model representation. Models with high spatial resolution require more computation time, but may not necessarily lead to better results for forecasting purposes. Often, it is preferable if the frequency of runs is determined by the availability of data and not by the runtime of the model.

6. LESSONS LEARNED

Operational systems for operational water management have to support the operational team with reliable data provision, both real time monitored and forecasts. Skilful forecasts and a system design accounting for robustness have a significant impact on confidence in the system. Continuous assessment of forecast skills and performance can contribute to decisions where investments in the system are most effective. Openness in system design allows agencies to reuse previously developed components, e.g. models and knowledge bases, hence enabling preservation of previous investments. The design choice to separate domain knowledge from the system modules and to separate data/communication formats from functional content is considered a major contributor to the evolution of Delft FEWS from its application in flood forecasting towards its current wide range of applications.

REFERENCES

Delft Hydraulics, DelftFEWS Configuration Guide, 2005. Also available at http://public.wldelft.nl/display/FEWSDOC/Home

Funke, R., B.Courtois and S.Soleilhavoup, Integrated hydrological information and knowledge management transfer for flood warning centres in European countries. Proc. of the 7th Int. Conf. on Hydroinformatics, 2006. Gourbesville, Cunge, Guinot and Liong (eds), Research Publishing, Vol.1, pp.163-170

Grijsen, J., Snoeker, X., Vermeulen, C., el Amin Moh. Nur, M., and Mohamed, Y., An information system for flood early warning, In: A. Saul (ed.), Floods and Flood Management, Kluwer Academic Publishers, (1992) 263–289.

Jellema, J. SWE-connected sensors: data source for geo information. European Spatial Data Research & Int. Soc. for Photogrammetry and Remote Sensing workshop on Geosensor Networks. Hannover, Germany 2008, http://www.ikg.uni-hannover.de/geosensor

OCG OpenGIS® Sensor Observation Service Implementation Specification, OGC® 06-009r1, OpenGIS Implementatoin Specification Open Geospatial Consortium Inc. 2006

Parker, D. and Fordham, M., Evaluation of flood forecasting, warning and response systems in the European Union. Water Resour. Manage. **10**, (1996), 279–302

Werner, M., M. van Dijk and J.Schellekens Delft-FEWS: an open shell flood forecasting system; 5th International Conf on Hydroinformatics. Liong, Phoon & Babovic (eds) 2004 World Scientific Publishing Company Vol.2 pp.1205-1212.

M.G. Werner J. Schellekens and J.C. Kwadijk Flood early warning systems of hydrological (sub) catchments. Encyclopedia of Hydrological Sciences, 2005. M.G. Anderson and J.J. McDonnell (eds), 349- 364

Werner, M.G.F. and K. Heynert, Open Model integration – a review of practical examples in operational flood forecasting, Proc. of the 7th Int. Conf. on Hydroinformatics, 2006. Gourbesville, Cunge, Guinot and Liong (eds), Research Publishing, Vol.1, pp.155-162

Werner, M., and D. Whitfield, On model integration in operational flood forecasting, Hydrol. Process. **21**, (2007) 1519–1521.

Whitfield D., The National Flood Forecasting System (NFFS) of the UK Environment Agency. In: Int. Conf. on innovation, advances and implementation of flood forecasting technology. Balabanis P, Lumbroso D, Samuels P (Eds). Tromsø, Norway, (2005), http://www.actif-ec.net (last accessed 7-8-2008).

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Design of a sensor network with adaptive sampling

Sergi Pons [a], **Luigi Ceccaroni** [b] **and Jaume Piera**[a]

[a] *UTM-CSIC, Passeig Marítim de la Barceloneta 37-49, E-08003 Barcelona, Spain (spons@utm.csic.es, jpiera@utm.csic.es)*

[b] *Departament de Llenguatges i Sistemes Informàtics (LSI), Universitat Politècnica de Catalunya (UPC), Campus Nord, Edif. Omega, C. Jordi Girona, 1-3, 08034 Barcelona, Spain (luigi@lsi.upc.edu)*

Abstract: This paper presents the preliminary design of a marine sensor network with high spatial-density of nodes, which uses artificial intelligence techniques to observe episodic events of dynamic natural phenomena in shallow waters. The focus of the research is on how to use intelligent data analysis and reactive behavior to improve the management of network resources, thus increasing the quality of the data collected or reducing the costs of the associated technology.

1 INTRODUCTION

In the emerging research field of marine observation systems composed of sensor networks it is useful to differentiate between deep sea and coastal networks, given the different environmental conditions and scientific interests. Important initiatives involving advanced networks in deep sea scenarios include ARENA (Kasahara et al. [2003]), NEPTUNE (Barnes et al. [2007]), ALOHA/MARS Mooring (Duennebier et al. [2006]) and ESONET (Priede et al. [2003]), while initiatives involving shallow water deployments include LOBO (Johnson et al. [2007]) and the Australian Coral Reef Sensor Network (Bondarenko et al. [2007]). The expansion of this technology is due to the fact that sensor networks facilitate and improve the coordination of sampling at several temporal and spatial scales. That is, sensor networks can be used to observe phenomena ranging from resolutions of minutes to resolutions of years and from resolutions of centimeters to resolutions of kilometers.

The control of sensors' operation generally needs some software, which can be of different degrees of sophistication. Artificial intelligence (AI) techniques have been applied to environmental problems for a long period of time with good results, as for example in environmental planning (Wright et al. [1993]), fish stock prediction (Sazonova et al. [1999]) or wastewater management (Ceccaroni et al. [2004]), just to name a few. The first reliable applications of AI to environmental issues appeared in the 1980's. More recently, AI research has been oriented towards the development of knowledge-based systems (KBSs), which, when applied to environmental issues receive different denominations, such as environmental decision support systems (Cortés et al. [1999]) or environmental KBS. An environmental KBS is a system, applied to an environmental issue, that potentially reduces the time in which decisions are made by the agents involved (e.g., the network nodes) and improves the consistency and quality of those decisions. Advancements in KBS research have the potential of benefiting many environmental fields, as well as very different disciplines. These advancements are in general results in applied optimization, process-control improvements, environmental decision-making and adaptive behavior.

In this paper, sensors capable of integrating highly sophisticated software are considered. In these cases, the software can be referred to as a KBS. Sensor networks can be connected to remote control centers or be autonomous. Data can be gathered in real time or collected at specific

times. If a network is connected to a control center, the sampling parameters of its equipment can be modified at will, to correct misbehavior or to accomplish new objectives. If a network is autonomous, to dynamically adapt to changes in the environment, reactive linear planning can be integrated into the equipment's control software in such a way that sensors:

- receive and possibly process on-line information about the process or phenomenon or parameter being sampled;
- use this information for selecting and revising models of the sampling dynamics;
- apply these models for autonomously planning the control of sampling and for supporting sensors' management by operators.

This adaptive behaviour, detailed in section 2, allows a more intelligent sampling and finally a better management of the network resources, including a reduction of energy consumption and of communication-bandwidth needs.

2 WHY IS ADAPTIVE SAMPLING NEEDED?

In the present study we aim to design an adaptive sensor network for shallow waters, with operational oceanography[1] and environmental monitoring purposes in mind; a network with nodes located a few kilometers from each other and not meant to be used under extreme environmental (especially, pressure) conditions, such as those of deep oceans.

Nowadays, the energy required to acquire data from sensors is used to: cancel noise through filters, amplify and digitize analog signals sensed, process and transport information, as shown by Raghunathan et al. [2006]. Sensors often require high-rate analog-to-digital converters with large power-demands. Power and data-transmission bandwidth are the most limited resources and should be used efficiently, as shown by Akyildiz et al. [2002]. Therefore, ideally a sensor network should guarantee that the sampling happens only if needed, when needed, where needed, and with the adequate level of fidelity. This kind of efficient sampling behavior would reduce the energy used for processing and communication.

Bio-geo-physical phenomena and most of the signals of natural origin have a highly intermittent behavior (e.g., wind gusts, turbulences, algal blooms). An intelligent analysis at network level can reactively plan the sampling through the detection and detailed characterization of episodic events. An example is sun radiation, shown in Figure 1. Cases (a) and (c) correspond to sunny days; these episodes can be easily reproducible by standard mathematical models, therefore a high temporal resolution is not needed. That is, if the interpretation of the signal can be easily inferred from a few data points, sampling can be at low rates. Case (b) corresponds to a cloudy day; it cannot be easily interpreted by current models due to its complexity, and its high variability would benefit from the use of high resolution sampling rates. Low resolution sampling rates should, then, be used in the periodic intervals with no radiation (night time). The general goal is to dedicate more resources to the most significant data intervals and to save resources when data are less relevant.

Another example is a system managing a wide-area sensor network, which could identify sub-regions of particular interest via low-resolution sampling, and then prompt a higher-resolution sampling for the sensors located in those sub-regions.

Finally, adaptive sampling could be used in such a way that sensors in adjacent regions could be reactively alerted in case of need, modifying their future behavior. For example, if a node detects a harmful algal bloom (HAB), it modifies its sampling parameters and sends a notification to neighbor nodes. These nodes use this notification as an additional input parameter for their decision systems, using the information to decide if the data they are measuring indicate the

[1] Operational oceanography is the activity of systematic and long-term routine measurements of seas, oceans and atmosphere, and their rapid interpretation and dissemination.

presence of a HAB of the same type. Figure 2 shows part of this behavior, specifically how the network reacts to HAB presence. In the figure, arrows represent data transmission between nodes. In (a) all the nodes are working at low sampling rate. In (b) a HAB appears on a region monitored by two of the nodes. In (c) the affected nodes change to high sampling rate, and alert neighbor nodes and the shore station.

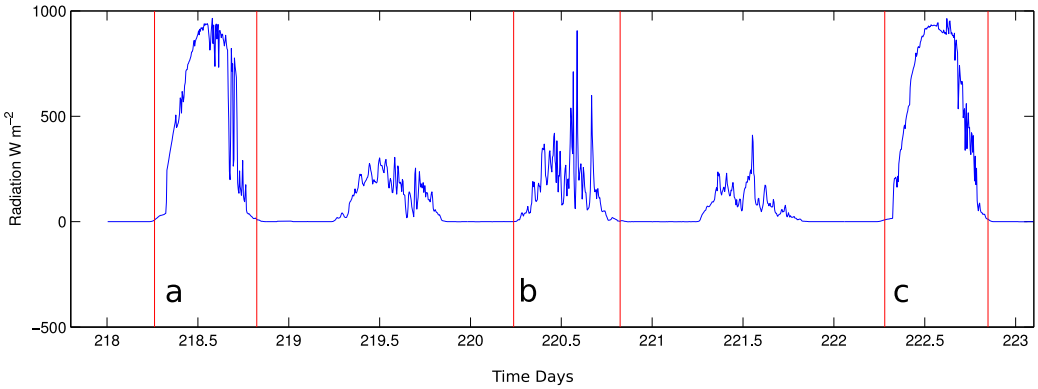


Figure 1: Example of signal (sun radiation), whose observation and post-processing would benefit from intelligent variation of sampling rates

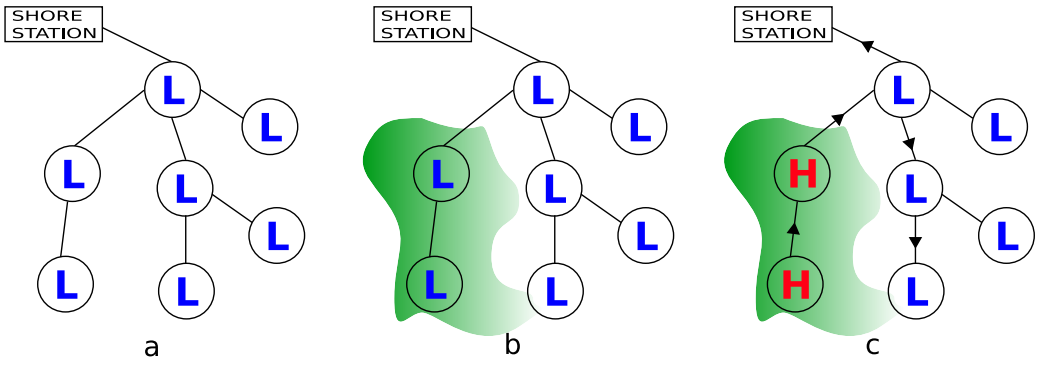


Figure 2: Network's reactive behavior. "L" indicates a node working at low sampling rate and "H" indicates a node working at high sampling rate.

Moreover, if a node has sampling equipment with motion capabilities, such as autonomous vertical profilers[2] (AVPs), it could be able to identify the points of most interest and sample at high resolution only at those points. In the case of AVPs, the sensors could move up and down the water column looking for depths with particularly interesting data (e.g., thin algal layers), and reduce its speed at these points, simultaneously increasing the sampling rate.

3 TECHNICAL CONSIDERATIONS

Three main technical topics are dealt with in this preliminary design of a marine sensor network:

[2] A vertical profiler is composed of a sampling module with positive floatability, with a winch and a cable attached to the sea floor. Releasing or winding up the cable, the system can modify its depth.

- network infrastructure;
- adaptive behavior of the network;
- intelligence at the node level.

The first topic taken into account is the **network infrastructure**. The connection among nodes can be fully cabled or acoustic/wireless. Wireless communications could present difficulties due to absorption, that increases with frequency, as shown by Akyildiz et al. [2005] and have in some case limited bandwidth; while the main drawback of cabled networks is the installation cost. Related to the needed bandwidth, when communications are in real time, a good management of the resources and the processing of the gathered data allows to use lower transmission rates. When communications are asynchronous, data can be stored locally at the nodes and sent when bandwidth is available. Even systems that deal with large amounts of data do not need constant streaming, nor constant storage. For example, in the case of high definition cameras, images are captured only when motion is detected. If not, data is discarded, neither stored nor sent.

When dealing with complex environmental problems, with processes that are not easily modeled because our knowledge is still incomplete and uncertain, environmental KBSs can be useful. Regarding the **adaptive behavior of a sensor network**, KBSs, considered as intelligent information systems that allow autonomous decisions and improve the time in which decisions can be made as well as the consistency and the quality of these decisions, represent an interesting approaches to infer spatiotemporal relationships among descriptors measured by sensor nodes. An important feature of KBSs is that they allow the use and management of specialized knowledge, which may include, among others:

- empirical knowledge about organisms and their environment;
- situational knowledge about local environmental conditions and its possible relation with the global environment;
- knowledge about human beliefs, intentions and priorities;
- theoretical knowledge about biological, physical and chemical phenomena.

The proposed research is centered in particular on *rule-based expert systems* (RBESs) and *case based reasoning* (CBR) systems. RBESs are advanced computer programs which emulate, or try to, the human reasoning and problem-solving capabilities, using the same knowledge sources, within a particular discipline. RBESs possess a fact base or ontology, a knowledge-base made of rules, and some inference and search process. The problems addressed through RBESs are very complex and related to specific domains, and they would usually need a human expert (i.e., a large amount of knowledge) to be solved. The central idea of CBRSs is that the solution from some past case is reused to solve the current problem. By recalling old solutions given to similar problems and adapting them to fit the new problems, CBR systems can improve their performance, becoming more efficient. Furthermore, they do not have to solve new problems from scratch. The memorization of past problems or episodes is integrated with the problem-solving process, which thus requires the access to past experience to improve the system's performance. Additionally, case-based reasoners, becoming more competent during their functioning over time, can derive better solutions when faced with equally familiar situations because they do not repeat the same mistakes (learning process). If nodes of a sensor network include complex systems, such as vertical profilers, the incorporation of some **intelligence** in the form of reactive behavior (see Figure 3), adopting for instance the approach of CBR systems, can optimize their processing and autonomy.

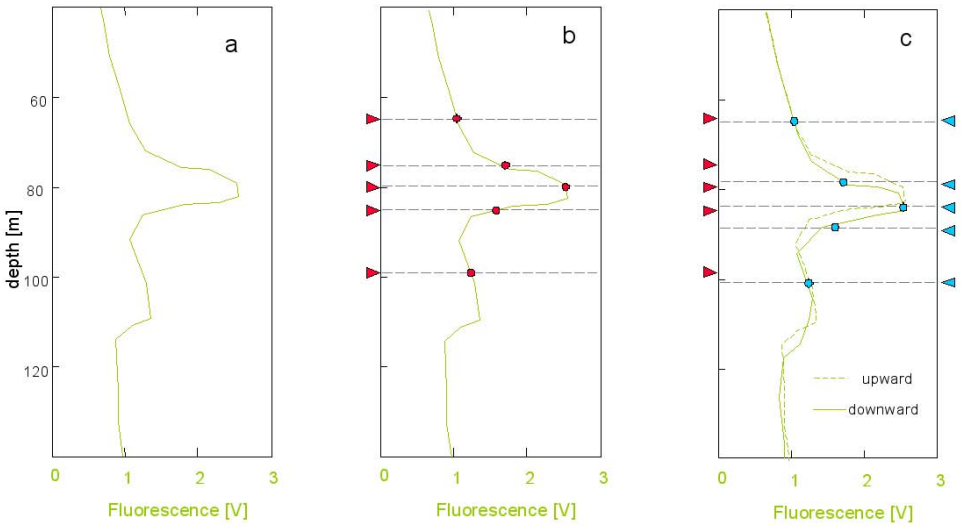


Figure 3: Example of profiler with the proposed reactive behavior. The type of data is not relevant in this case, as it is just a theoric example. (a) Going downwards, the profiler characterizes the water column through low resolution sampling. (b) Data are processed, and interesting depths (according to predefined rules) are determined for high resolution sampling on the way back up. (c) Going upwards, the water column properties could be different from the first profile, so the system, while planning for the high resolution sampling, has to react to the changes.

4 TEST BED

An excellent place to implement and test this kind of network is Alfacs Bay (Figure 4), Tarragona, Spain. This is a bay with a rhomboid shape of 11 x 5 km, a maximum depth of 7 m, an average depth of 3 m and a total water volume of 1.5 x 10^8 m^3. It acts as an estuary where fishing and seafood farming are the main economic activities and the appearance of HABs has economic negative effects. At times the bay must be closed because of HABs and the production is lost. A long term monitoring can be used to improve the understanding of the bay dynamics, while the small scale, high frequency sampling can help to detect and prevent the HABs damage and to study small scale phenomena.

The planned first phase of the project will consist of the deployment of a central node, which will be connected to a land observatory, plus two other nodes connected to it. This land line will provide energy supply and a communication channel to all the network. During the next phase, additional nodes will be added for higher spatial resolution. The proposed sensors to implant on the nodes are: conductivity sensor, temperature sensor, fluorometer, acoustic doppler current profilers (ADCPs) and hyperspectral sensors.

The sensors could be attached to vertical profilers or be fixed, close to the sea floor. They will provide information about physical and biological processes of the bay and the information they will collect could be useful for both scientific research and economic activities (e.g., detection of poisonous algae on seafood farming regions).

5 CONCLUSIONS

The work presented represents a new research line in which artificial intelligence techniques are starting to be introduced in the design and implementation of sensor networks for shallow, marine zones. The adoption of problem-specific, environmental, knowledge-based systems, in particular, could provide a qualitative and quantitative improvement in different aspects, but primarily in the

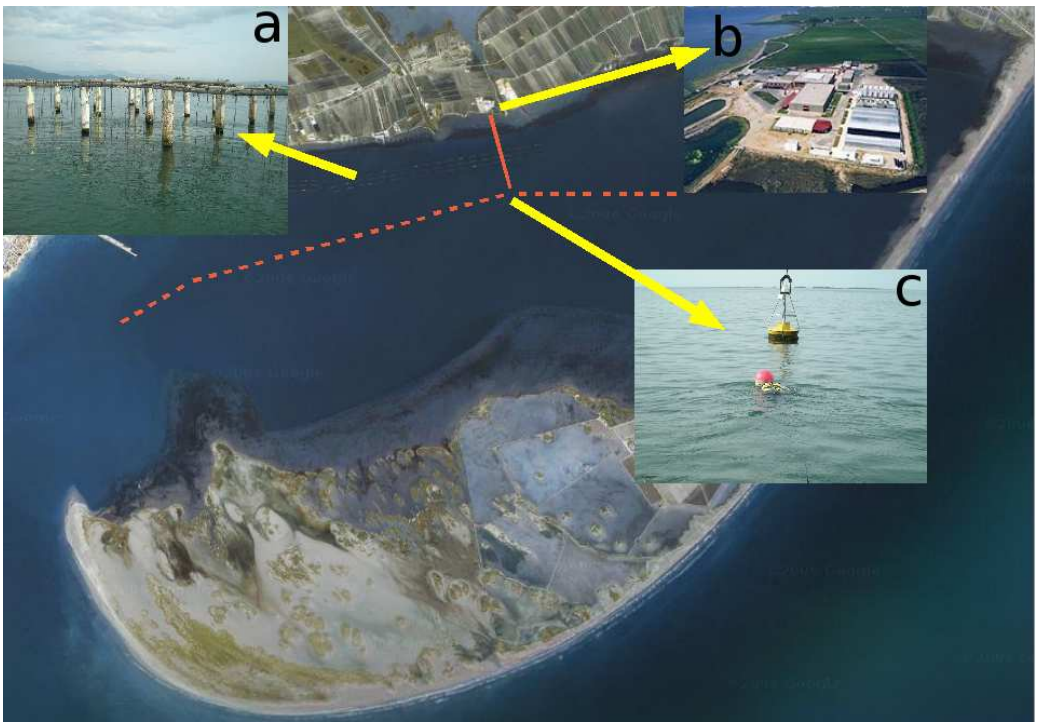


Figure 4: Proposed test bed for a sensor network at Alfacs Bay. (a) Seafood farming, an important economic activity in the region. (b) Current facilities that can act as shore station. (c) Possible emplacement of the central node, with a link to the coast.

optimization of the sampling process. We believe that the introduction of these advanced systems is essential in order to obtain intelligent, autonomous and energy efficient sensor networks. The system under development is tailored to relatively narrow environmental problems and domains, but it is applicable to a wide range of different locations and situations. Additionally, a real scenario in which to implement a network with these characteristics is proposed. The placement not only fits the environmental requirements (i.e., shallow water), but is also a region of high scientific and economic interest.

REFERENCES

Akyildiz, I. F., D. Pompili, and T. Melodia. Underwater acoustic sensor networks: research challenges. *Ad Hoc Networks*, 3(3):257–279, May 2005.

Akyildiz, I., W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *Communications Magazine, IEEE*, 40(8):102–114, 2002.

Barnes, C., M. Best, B. Bornhold, S. Juniper, B. Pirenne, and P. Phibbs. The neptune project - a cabled ocean observatory in the ne pacific: Overview, challenges and scientific objectives for the installation and operation of stage i in canadian waters. *Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, 2007. Symposium on*, pages 308–313, 2007.

Bondarenko, O., S. Kininmonth, and M. Kingsford. Coral reef sensor network deployment for collecting real time 3-d temperature data with correlation to plankton assemblages. *Sensor Technologies and Applications, 2007. SensorComm 2007. International Conference on*, pages 204–209, 2007.

Ceccaroni, L., U. Cortés, and M. Sànchez-Marrè. OntoWEDSS: augmenting environmental

decision-support systems with ontologies. *Environmental Modelling and Software*, 19(9):785–797, 2004.

Cortés, U., M. Sànchez-Marrè, I. R-Roda, and M. Poch. Towards Environmental Decision Support Systems. Environmental Decision Support Systems and Artificial Intelligence. *AAAI Workshop Technical Report (WS-99-07)*, pages 8–14, 1999.

Duennebier, F., M. Tremblay, D. Harris, J. Jolly, D. Copson, J. Babinec, R. Lukas, and D. Karl. The ALOHA Cabled Observatory. 2006.

Johnson, K., J. Needoba, S. Riser, and W. Showers. Chemical Sensor Networks for the Aquatic Environment. *Chem Rev*, 2007.

Kasahara, J., Y. Shirasaki, K. Asakawa, and K. Kawaguchi. Scientific application of arena networks. *Scientific Use of Submarine Cables and Related Technologies, 2003. The 3rd International Workshop on*, pages 272–275, 2003.

Priede, I., M. Solan, J. Mienert, R. Person, T. Van Weering, O. Pfannkuche, N. ONeill, A. Tselepides, L. Thomsen, P. Favali, et al. ESONET-European Sea Floor Observatory Network. *Proc. of the 3rd Inter. workshop on Scientific Use of Submarine Cables and Related Technologies*, pages 263–265, 2003.

Raghunathan, V., S. Ganeriwal, and M. Srivastava. Emerging techniques for long lived wireless sensor networks. *Communications Magazine, IEEE*, 44(4):108–114, 2006.

Sazonova, L., G. Osipov, and M. Godovnikov. Intelligent system for fish stock prediction and allowable catch evaluation. *Environmental Modelling and Software*, 14(5):391–399, 1999.

Wright, J. et al. *Expert systems in environmental planning*. Berlin; New York: Springer-Verlag, 1993.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Integration of sensor networks and decision support tools for basin-scale, real-time water quality management

N.W.T. Quinn [a,b]

[a]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California (nwquinn@lbl.gov) and [b]US Bureau of Reclamation, 2800 Cottage Way, Sacramento, CA 95825 (nquinn@mp.usbr.gov)

Abstract : Recent advances in sensor technologies and the steep decline in the cost of developing and operating sensor networks have led to an explosion of environmental applications. Real-time telemetry access to the environmental monitoring data produced by these sensor networks can now provided reliably and cheaply. However development of the institutional and administrative structures that turn this torrent of real-time data into information that informs decisions is more challenging - examples of successful sensor network projects that support decision-making at the basin-scale level are few. In this paper some of the challenges faced in making sensor networks an integral part of river basin management are described using an example from the San Joaquin River Basin of California. The San Joaquin River is an impaired water-body and one of the most regulated in the nation. An innovative alternative to EPA's TMDL (Total Maximum Daily Load) pollutant control policy is being investigated which relies on forecasting salt assimilative capacity in the River in real-time and coordinating salt load discharge and diluting flow releases, in response to these forecasts while meeting downstream water quality objectives. The paper discusses key elements of real-time water quality management which address the technology (sensors, telemetry, visualization, databases, quality assurance procedures); information dissemination and technology transfer to stakeholders including decision support procedures; and cost containment strategies for long-term system deployment. In the paper one important set of stakeholders – the managers of seasonally drained wetlands - are singled out to clearly illustrate the impediments and adaptive solutions to this ambitious basin-scale project. Unique features of the project include the monitoring technologies chosen to accurately measure highly variable flows in irregular channels and some of the data quality assurance challenges that have ensued. Also the institutional differences between State, Federal and private wetland management that need to be recognized in the choice of software that helps to transform field data to information that can be used to guide decisions. The paper also discusses the concept of project assurances and how these are factored into the adaptive management conceptual model being followed to realize long-term project goals.

Keywords: TMDL, sensor networks, real-time water quality management, wetlands, drainage, salt

1. INTRODUCTION

The last decade has heralded significant advances in the quality and accuracy of environmental sensors. The market has been penetrated by large numbers of small specialty firms that have recognized the enormous market potential for these devices given public demand for safe drinking water and environmental regulation of non-point source polluters. The entry of these new companies has increased competition, driven down the cost of sensor deployment and improved sensor reliability. More recent advances have introduced web-enabled sensor networks and smart sensors, capable of stopping, starting and performing elementary control

functions based on simple logic. A decade ago this technology was the domain of firms specializing in industrial control systems or in SCADA (Supervisory Control and Data Acquisition).

Environmental applications have lagged industrial and irrigated agriculture applications largely on account of the significant investments required and the technical know-how required for ongoing operations and maintenance. However, recent environmental policy in the United States directed at non-point source pollution, such as the Total Maximum Daily Load (TMDL), has encouraged watershed stakeholder groups whose operation generates much of the non-point source pollutant loads to consider real-time data acquisition and decision support technologies. The TMDL is defined as the maximum amount of a pollutant that a water body can receive and still meet water quality standards. The TMDL is typically the sum of the Waste Load Allocation (the total daily load of a contaminant from point sources entering the water body allocated among individual point sources), the Background Loading (naturally occurring loading of the contaminant) and also commonly includes a Margin of Safety (an additional allocation to account for parameter uncertainty). TMDL's are best implemented taking a "watershed approach" with its emphasis on holistic watershed planning and coordination with other watershed activities affecting water quality. In this paper some of the challenges in implementing a basin-wide TMDL in a complex river basin are described using the example of the San Joaquin River Basin of California. The San Joaquin River which provides the only drainage for the Basin is an impaired water-body and one of the most regulated in the nation. An innovative alternative to the prescribed TMDL pollutant control policy has been investigated for almost a decade – this approach relies on forecasting salt assimilative capacity in the River in real-time and developing the capacity for scheduling and coordinating discharges of saline drainage to the River, in response to these forecasts, to improve compliance with water quality standards, set a downstream compliance monitoring station.

2. ENVIRONMENTAL REGULATION

2.1 Salt Management in the San Joaquin Basin

The State of California charges government entities overseen by Regional Water Quality Control Boards (RWQCB) to regulate and enforce water quality objectives in the surface waters of the State. These water quality objectives are defined in the Water Code as "… the limits or levels of water quality constituents or characteristics which are established for the reasonable protection of beneficial uses of water or the prevention of nuisance within a specific area." Beneficial uses include an anadromous salmon fishery, riparian diversions to agricultural crops and avian wildlife resources. Federal and State environmental regulations do not always concur in the United States although federal regulations typically define the minimal level of compliance – States' can promulgate stricter standards which they sometimes are able to enforce. The TMDL for salinity in the 80,000 square kilometre San Joaquin Basin, California relies on the typical formulation which is based on a design flow condition equivalent to the 10% hydrologic low-flow condition, a factor of safety and an allowance for inter-annual variation. Because of the extremes in design flows typical of an arid zone river basin, the design flow restriction produces a TMDL which is very restrictive in allowable salt loading. During all but wet years, the TMDL would allow no salt loads to be discharged to the River above the compliance point during the irrigation season months of June, July and August when irrigation applications are highest and the need for drainage most acute. The TMDL also constrains drainage from approximately 60,000 hectares of managed seasonal wetlands in the Basin which provide wetland habitat to migratory waterfowl on the North America Pacific Flyway. Habitat is managed to mimic pre-development hydrologic conditions when native grasslands were inundated by spring flood-flows from east-side tributaries to the San Joaquin River. Constraints on seasonal drainage could impair the long-term sustainability of this important wildlife resource.

Recognizing the potential negative socioeconomic impact of the salinity TMDL on the San Joaquin River Basin, the RWQCB (CEPA, 2002) introduced a unique variant to the standard

TMDL formulation popularly known as "real-time water quality management" (RTWQM). This strategy has the potential to allow more export of salt load once implemented however it also requires the development of a watershed-wide, RTWQM system to continuously monitor assimilative capacity and prevent violation of salinity objectives at the downstream compliance monitoring station. Operation of the RTWQM system will require the formation of an institution charged with oversight of the design, installation and management of real-time flow and water quality monitoring stations. This institution would also be responsible for running simulation models of the basin to continuously forecast salt assimilative capacity and provide decision support for scheduling and coordination of saline drainage export to the San Joaquin River. A Management Agency Agreement was recently signed between US Bureau of Reclamation (the federal water agency that provides water supply Basin irrigators and wetland managers) and the State RWQCB as a first step towards providing the framework for such an institution. This agreement also sets a timetable for RTWQM implementation and helps to guide and prioritize federal funding of the first phase of activity.

2.2 Seasonal Wetland Drainage Salt Management

Although the wetland resources of the San Joaquin Basin are protected under the Endangered Species Act wetland discharges of salt load are conjoined with agricultural salt loads in the legislation that provides the RWQCB with the authority to regulate non-point source pollution (CRWQCB, 2002). The 50,000 hectare Grasslands Ecological Area constitutes the largest contiguous wetland in the State of California and is comprised State and Federal wildlife refuges and privately managed seasonal wetlands. These lands serve as important over-wintering habitat for waterfowl on the Pacific Flyway and provide sport for the State's duck hunters. These wetlands are flooded in the autumn, starting in early September, and drained in the early spring, typically between February and April in order to mimic the natural flooding of the San Joaquin River prior to irrigation development. Seasonal wetland drawdowns are timed

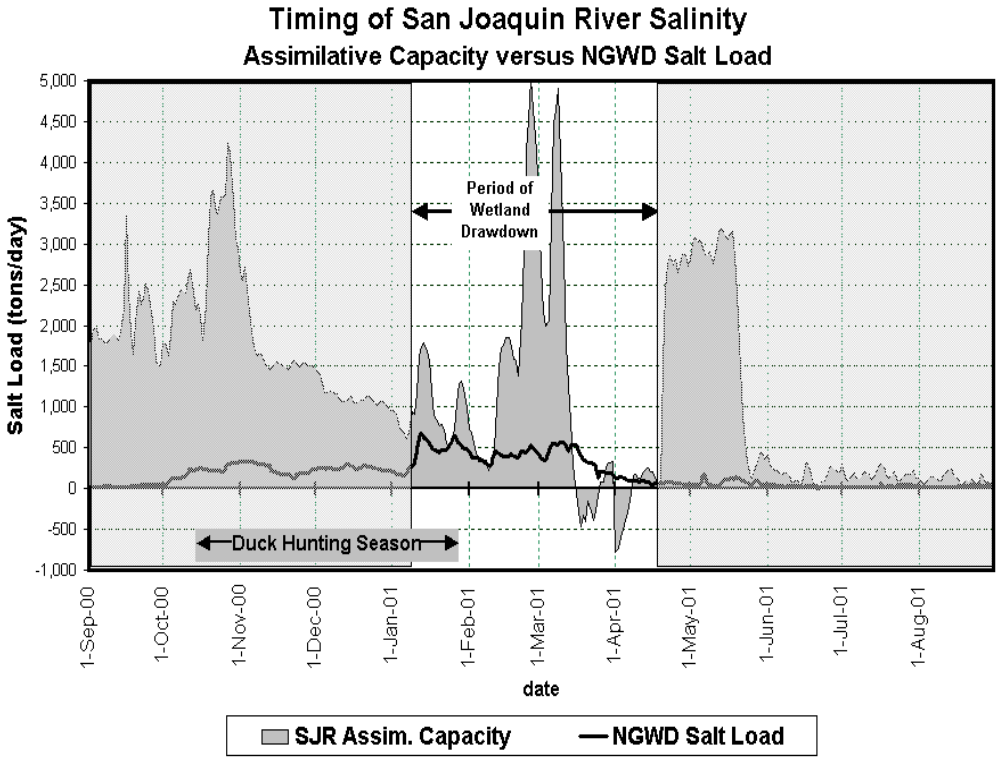


Figure 1. San Joaquin River salt assimilative capacity showing short periods in April and May 2001 when salt load discharged from the North Grassland Water District (drainage outlets for private wetlands within the Grasslands Ecological Area) exceeds river salinity objectives. Period of annual seasonal wetland drawdown coincides with downstream agricultural diversions for salt sensitive crops.

to make seed and invertebrate resources available during peak waterfowl and shorebird migrations and to correspond with optimal germination conditions (primarily soil temperature) to grow desirable moist-soil plants. The seeds produced by these moist-soil plants are recognized as a critical waterfowl food source, providing essential nutrients and energy for wintering and migrating birds (Fredrickson and Taylor 1982).

Wetland releases that contain high salt loads during the months of March and April coincide with agricultural pre-season irrigation to propagate plant seedlings. Saline water can inhibit germination and reduce crop yields. Better coordination of agricultural and wetland releases with reservoir releases of good quality snow-melt water on the east-side of the San Joaquin River Basin has been suggested as a means of improving San Joaquin River water quality for all beneficial uses (Quinn et al., 1997; Quinn and Karkoski, 1998). Management of wetland drainage, through scheduling of releases to coincide with periods of San Joaquin River assimilative capacity, is being evaluated relative to potential biological impacts of changes to traditional wetland management practices. This evaluation will identify potential impacts to seed germination rates, waterbird foraging rates, habitat availability, and species diversity and abundance brought about by manipulating wetland drawdown schedules. Drawing down wetlands earlier than optimal would be timed to coincide with assimilative capacity provided by rainfall runoff. Wetland drainage can be delayed to coincide with prescribed reservoir flow releases to benefit fish migration that are made between April 15 and May 15 each year (shown as the large step increase in river assimilative capacity in Figure 1).

The institutional constraints to wetland real-time water quality management differ between the state, federal and private entities that make up the Grassland Ecological Area. The state and federal refuges and private wetlands all have different priorities with respect wildlife diversity and species of migratory waterfowl that they wish to attract – wetland drawdown schedules can affect moist soil plant germination success and the ability of native grasses to compete with weeds that have little benefit either as overwintering habitat or forage. However since the state and federal refuges are public resources - decision making is highly centralized and management priorities are set by the Refuge Manager and are implemented by refuge staff. Private duck clubs, of which there are 160 in the Grassland Ecological Area, have their management priorities set by the board of directors of each club – these management priorities differ between duck clubs and more from those in the state and federal refuges. For example, certain duck clubs graze cattle during the spring months both for income and as a means of controlling propagation of undesirable vegetation. As a consequence these wetlands drain early so as to start cattle grazing once the soils have dried sufficiently. The Grassland Water District, which supplies water to each of these 160 duck clubs, manages a more dynamic system with many decision makers. The management priorities of these various stakeholders have implications for the design of the environmental monitoring system network and decision support software that continuously processes data from the stations that comprise the monitoring system network.

2.3 Environmental Monitoring System Design

To be eligible for the excursion from typical TMDL requirements mandated by the RWQCB, the state, federal and private entities that comprise the Grasslands Ecological Area must coordinate their drainage schedules not only among themselves but also with agricultural and municipal stakeholders within the River Basin to: (a) take advantage of periods of assimilative capacity in the San Joaquin River; and (b) limit salt loads discharged when assimilative capacity is impaired. The environmental monitoring system must be capable of : (a) providing each entity with the means of continuously quantifying salt loads contained within the many wetland impoundments that comprise the total area under management; (b) supporting daily decisions making concerning allowable salt export; and (c) providing the means of regulating these drainage discharges. Quantitatively assessing salt loads for more than 1,500 seasonally managed wetland impoundments within the Grasslands Ecological Area will require portioning the managed wetland are into sub-catchments – the outlets from which should become part of the monitoring network. The number of monitoring stations must be carefully designed to

provide critical information to support decision making without overwhelming the resources of entity responsible for water and drainage management. Since real-time water quality monitoring stations have only been operated in the Grasslands Ecological Area for a period of five years – enhancements to the existing monitoring network should be made incrementally. Simulation models of wetland hydrology, models of open water evaporation and wetland moist soil plant evapotranspiration will be needed to extrapolate monitoring station data to the entire area under management. The state and federally owned wetlands in the Grassland Ecological Area are aggregated according to their hydrology into separate wildlife management areas each with their own water supply facilities and drainage outlets. In one of the wildlife management areas all drainage return flows collect in a terminal pond which is maintained as a permanent wetland. Such a configuration greatly simplifies monitoring requirements and enhances the ability to regulate discharge of salt loads to the San Joaquin River from a single point. This configuration also lends itself to future automation where a screw gate could be raised and lowered in response to dynamic assimilative capacity allocations provided by a centrally-maintained salinity forecasting model. Implementation of RTWQM may require modifications to the design of the existing drainage network which might involve a greater level of cooperation and coordination than currently exists and possible use of shared facilities.

3. COMPONENT ELEMENTS OF REAL-TIME WATER QUALITY MANAGEMENT

3.1 Continuous Water Quality Monitoring

Monitoring and the collection of accurate, meaningful data is the most difficult and frequently under-budgeted part of watershed water quality management. Continuous monitoring augmented by periodic discrete sampling of both flow and water quality are necessary to achieve a scientifically valid estimate of salinity loading (Jarrell, 2003). Salinity is measured as specific conductance or electrical conductivity – a measure of the ionic content of the water that is closely correlated with the total dissolved solids (TDS). Because different waters contain different dissolved ions - the relationship between EC (deciSiemens/cm) and TDS (in parts per million) can vary between 0.55 and 0.95 for common waters. EC is multiplied by 0.75 for the conversion to TDS for water discharged from the Grasslands Ecological Area. Readings of EC must be corrected to compensate for the increase in ionic activity with temperature. Flow measurement in an area such as the Grasslands Ecological Area is challenging owing to the frequent low rates of discharge from wetlands ponds that can vary from 0.002 m^3/sec during the flooded season to 2 m^3/sec for short periods during wetland drawdown. Outflow from these wetlands is typically controlled using concrete culverts with sliding weir boards which are imprecise owing to the small differences in weir stage and the potential error introduced by earth tides and atmospheric pressure variations. Weir boards were replaced with v-notch weirs were tried in an attempt to improve the accuracy of measurement at low flow – these were unpopular and keeping track of adjustments in weir height became unwieldy and irritating for water managers. The best results have been obtained with acoustic Doppler flow sensors – the technology chosen manufactured by MACE Inc. (Australia). These flow sensors were attached to a 2 meter length of channel iron which was fixed in place at the bottom of each drainage culvert using a T-brace at the upstream end. Each flow sensor is polled every 15 minutes by a MACE data collection platform that has connectivity to a YSI-EcoNet monitoring network.

3.2 Sensor Network Platform

YSI-EcoNet (YSI Inc., 2005) is a sensor network platform that uses both radio and cellular telemetry to produce a low-cost, easy to use distributed system of flow and water quality sensors with a link to a centralized database that allows real-time access to the data via the world-wide web (Figure 2). The wireless mesh network topology allows "point-to-point" or "peer-to-peer" connectivity and creates an ad hoc, multi-hop network. The mesh network is self-organizing and self-healing – hence loss of one or more nodes does not necessarily affect its operation. This increases the overall reliability of the system by allowing a fast local response to critical events in the rare event of a communication problem. Radio telemetry is

used for communication between individual data nodes, which report drainage water flow, temperature, electrical conductivity, and salt load at each monitoring site every 15 minutes, and a master node which relays this data to the NIVIS Data Center on the east coast of the United States. The NIVIS Data Center provides real-time web access to all current and past data. The system allows point and click access to current monitoring data at a particular Data or Access Node within the network. Maintenance of the monitoring network can now focus on monthly sensor quality assurance checks including cleaning of sensors and checking the accuracy of gauge stage data from which flow is determined. The YSI-EcoNet website can be customized to restrict access to certain more sensitive data – such as flow totals or computed salt loads. Because the data is uncensored – problems at a site could produce alarming numbers that might be used against the stakeholders politically. However, despite this potential concern, the provision of real-time access to environmental data has had significant impact on stakeholders and wetland managers. It has provided the first educational feedback on the effects of certain management on water quality and has proved a low cost and effective way of communicating to State water quality regulators that Basin salinity problems are being addressed in a proactive manner. The YSI-EcoNet sensor network configuration is fully programmable through the website and the user interface is sufficiently user-friendly to allow sensor configuration and network maintenance to be performed by project stakeholders. Local ownership of the technology from the onset has an important factor in the progress made toward RTWQM.

3.3 Real-Time Data Quality Assurance

Real-time data is always considered to be "preliminary" until the data has undergone quality assurance checks and has been verified against gauge records or discrete samples. Until relatively recently software to assist with the process of providing continuous quality assurance of monitoring data has been lacking as have data management software packages specifically designed for dealing with time series data. We are in the process of implementing Aquatic Informatics Inc.'s Aquarius time series software for streamlining and eventually automating the data downloading, error checking, error correcting and validated data web posting. Some of the key features of the Aquarius software is its open architecture and object structure allowing customized "whiteboarding" of data processing, standardization of data flagging, data correction algorithms and data interpolation. With the next release of the software YSI-EcoNet will be a sub-object within the data import object of the software allowing direct time series data downloading from the NIVIS Data Center. This will allow full automation of the data processing function and direct reporting of validated data to stakeholders, addressing the earlier limitation of only making total flow and salt load data available to water agency stakeholders and project participants.

3.4 Information dissemination and technology transfer to stakeholder groups

Connectivity of monitoring network resources is enhanced with standardization of sensors, data collection platforms, telemetry, database and data processing functionality. One of the significant challenges for RTWQM will be real-time data acquisition from water districts, refuges, municipalities and other entities that have their own closed flow and water quality monitoring systems owing to concerns with data security. For a RTWQM monitoring network to work effectively the network database should be refreshed daily with flow and water quality data that are automatically uploaded to a central workstation where the simulation model running continuous forecasts of San Joaquin River assimilative capacity is physically located.

The formation of a stakeholder-led entity that has primary responsibility for maintaining the monitoring station network, making routine forecasts of San Joaquin River assimilative capacity and controlling daily salt loading to the San Joaquin River will have secondary benefits by promoting the transfer of technology between stakeholders and providing assurances to stakeholders who may be concerned about information sharing with government regulators. Recent experience with selenium regulation from agricultural drainage sources in the San Joaquin Basin has provided an excellent exemplar for future full implementation of RTWQM. In the selenium regulation example (Grasslands Bypass Project) monthly and

annual selenium load targets were determined for the watershed by a agency-led stakeholder committee – a local water district representing landowners whose farms produced the selenium drainage being regulated assumed responsibility for a network of monitoring stations, invested in drainage recycling technology, ran accounting models to keep track of annual and monthly

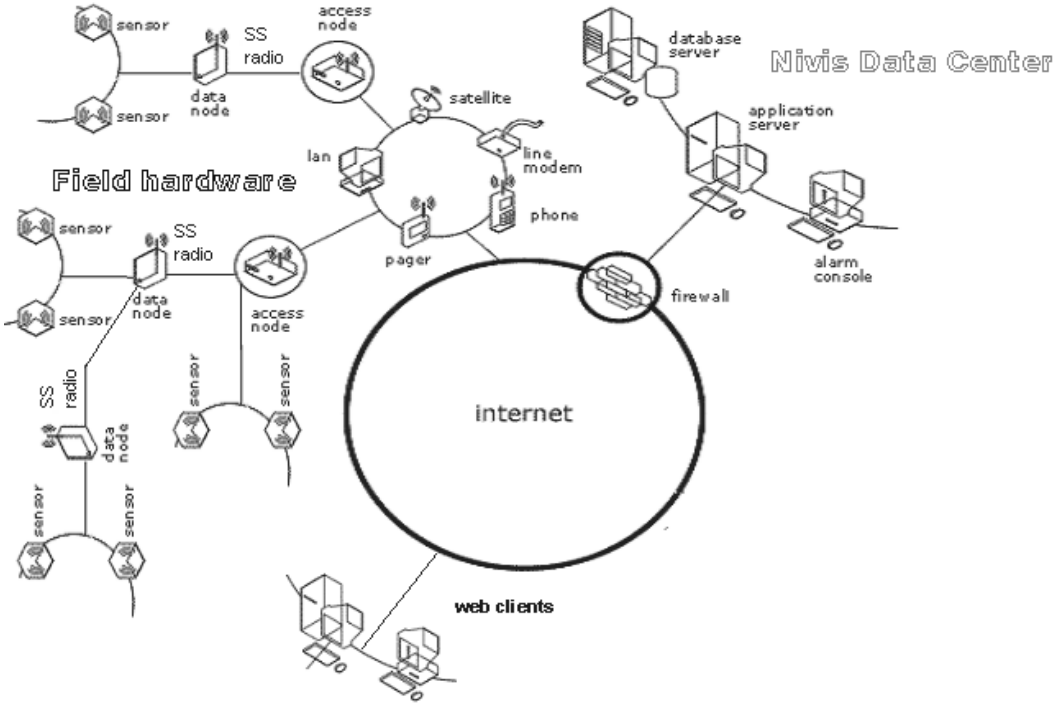


Figure 2. System architecture linking field monitoring stations with the NIVIS Data Center which stores, maintains and serves real-time flow and water quality data on public and private websites (YSI EcoNet manual, 2005).

selenium loads and managed all drainage outlets to the San Joaquin River. Except for two El-Nino years of exceptionally high rainfall the water district was able to consistently meet selenium load targets using individual drainage sump totalizing meters and a knowledge of irrigation schedules to forecast drainage flows and selenium drainage loads. Extending this model to RTWQM in the San Joaquin Basin suggests that coordination is best achieved and stakeholder assurances more likely provided by subdividing the Basin into drainage management areas that coincide with stakeholder coalitions since these groups are best suited to representing the interests of their stakeholders.

3.5 Salt Assimilative Capacity Simulation and Forecasting

Watershed management of salt loads under RTWQM requires simulation of current flow and water quality conditions in the San Joaquin River and the ability to make short-term (2 week) forecasts of salt assimilative capacity. River salt assimilative capacity forecasting is performed using a watershed-based flow and water quality mass balance model WARMF-SJR (an outgrowth of previous river water quality simulation models, Pate,2001; Kratzer et al., 1987).

WARMF-SJR is a GIS-based, public-domain, watershed simulation model that has been used extensively for TMDL planning (Herr and Chen, 2006). WARMF-SJR performs flow and salt load mass balance calculations of all inputs and diversions to the San Joaquin River between the upstream boundary at Lander Avenue downstream to the Vernalis salinity compliance monitoring station, located approximately 100 km downstream. The hydrological model simulates percolation of irrigation water through soil, evapotranspiration of water through crops, change of groundwater table, agricultural return flow, and groundwater accretion to the river reaches. River salt assimilative capacity forecasting requires prior knowledge of same data inputs used to perform the flow and water quality simulations. In the case of reservoir releases these are highly deterministic during certain times of the year when the federal

government purchases water supply to aid fish migration in the San Joaquin River and its major tributaries. However, outside of these periods, model forecast input data must be gathered from stakeholders and rely on professional judgement and experience. The WARMF-SJR model has the capability of simulating watershed processes to calculate flows and non-point source loads of pollutants from various land uses (urban, forested, and agricultural areas). The watershed simulator within WARMF is used to make daily assimilative capacity forecast in the absence of stakeholder projections of flow and water quality input parameters.

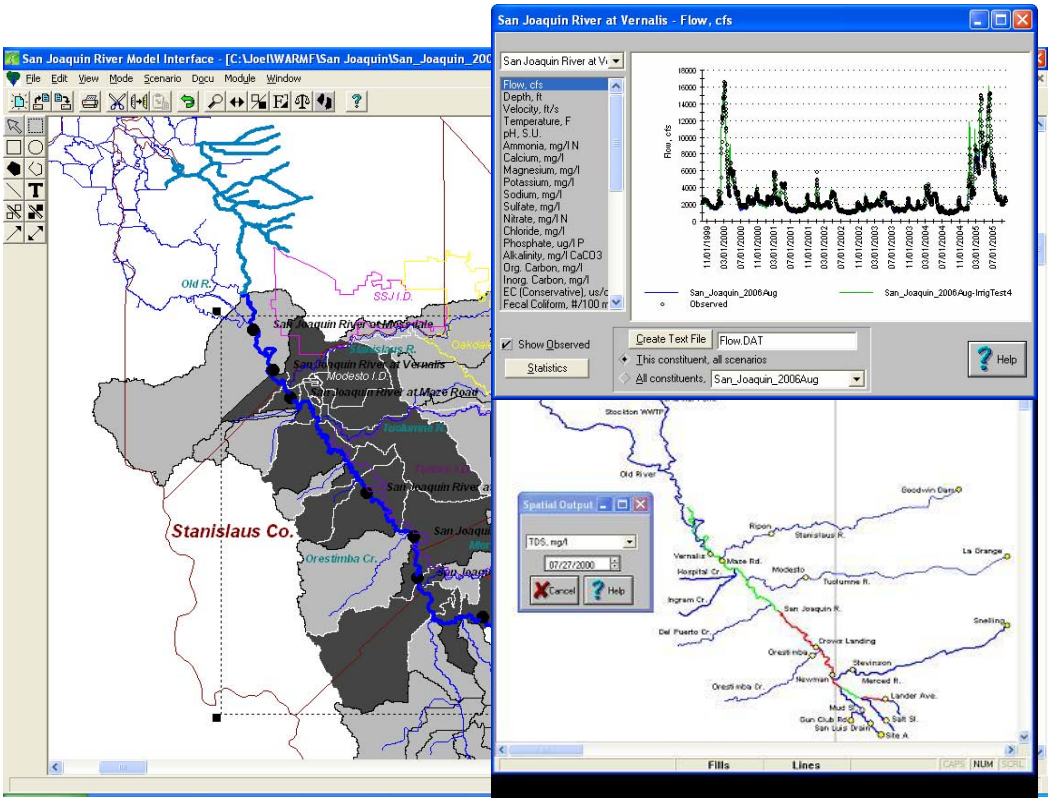


Figure 3. WARMF-SJR flow and water quality forecasting model. Interactive map allows point and click access to all input data in a Microsoft Excel$_{TM}$ format. Model hindcasts can be checked against real-time monitoring data using this interface. Model output can be exported to a graphical user interface which provides easy access to assimilative capacity forecasts and shows those portions of the River that are in compliance with salt load targets.

Work is underway to provide a forecasting input data resource module that will support the WARM-SJR salt assimilative capacity forecasting process. Daily forecasts currently use the previous week's real-time monitoring data as a training "hindcast" period, which permits validation (and occasional calibration) of model input parameters and then project ahead for a period of two weeks. An initial forecast is typically performed using input data from the historic record for "like hydrologic conditions" – the forecasting input data resource module will allow the forecaster to scan the database for similar year conditions and guide model data substitution. Meteorological forecasts can be used to overwrite hydrological input data during the winter rainy season using the watershed process module. The module provides rainfall runoff estimates and provides projections of surface drainage and short-duration groundwater accretions. The goal of the module development is to allow the WARMF-SJR forecasting model to be used effectively by all stakeholders, even if they have little expertise in San Joaquin Basin hydrology. The use of the module will promote adaptive learning – developing the user's experience with time.

Model salt assimilative capacity forecasts have been disseminated in a number of formats – web posting of the forecasts have been shown to be the most effective means of reaching the majority of stakeholders. An interactive user interface was developed to provide a visual

display of two week forecast – viewers can step through each day of the 14 day forecast and can see those reaches of the River that are out of compliance with daily salt load allocations and with the salinity objective at the Vernalis monitoring station. In the future, with the establishment of a central authority for managing San Joaquin River water quality, salt load allocations would be made routinely by this entity – the entity would also have responsibility for setting reach-by-reach salt load targets and for monitoring compliance with these daily salt load allocations.

3.6 Decision support for seasonal wetland salt load scheduling

Customized decision support tools are needed by stakeholders to perform salt management accounting within the local watershed and which will assist scheduling of salt load export. The tools, if they are to be used effectively, should be developed with the active participation of the water manager in charge of drainage scheduling. A prototype wetland water quality model (WWQM) was developed for the Grassland Water District in the Grasslands Ecological Area which tracks salinity changes in the wetland impoundments on a daily time-step over the winter season and incorporates user-defined schedules for wetland drawdown in the spring months. Input data for the WWQM include soil properties, land classifications, pond sizes, crop evapotranspiration (ET) coefficients, water supply and drainage quantity and quality, precipitation, wind velocity, water and soil temperatures. San Joaquin River assimilative capacity forecasts are supplied by the WARMF-SJR model and the proportion allocated to the Grassland Water District (which has yet to be determined) would be used to determine the need to withhold drainage discharge. In hydrologically wet and normal years there may be little need to manipulate drainage schedules. To improve salt load management and control individual wetland impoundments within the District were lumped using a GIS delineating the contributing subareas to five real-time drainage monitoring sites on the northern boundary of the District. This allowed salt loading to be updated continually during the winter and drawdown period without overwhelming the end user with information.

Under RTWQM water managers could delay the drawdown of certain wetland impoundments within each drainage subarea to control salt export to fall within the allocated portion of daily River assimilative capacity. At the present time drawdown of a wetland impoundment is initiated by pulling weir boards at the outlet, once a wetland starts to drain the boards are rarely replaced until drainage is complete. In the future weir boards at major outlets may be replaced with powered gates which could be controlled remotely using existing YSI-EcoNet technology with the addition of modular controller circuit boards. Over the longer term drainage scheduling decisions could be directly tied into the WWQM.

4. SUMMARY

Implementation of the RTWQM concept in the San Joaquin Basin is a unique and ambitious attempt to address some of the deficiencies of the EPA's TMDL policy which tend to be over-restrictive in arid watersheds and hence make sub-optimal use of the Basin's water resources. RTWQM requires Basin-wide real-time monitoring of both flow and water quality, simulation and forecasting of River salt assimilative capacity and the dissemination of these real-time forecasts to allow stakeholders opportunity time to curtail their exports of salt load so as not to exceed their allocations. An institutional entity will be formed over the next several years to manage this new strategy for Basin-level water quality management. This paper has described current progress towards meeting this ambitious goal and has described some of the decision support tools currently in service and those that have yet to be fully developed. Integration of the various modelling and decision support tools while providing an open architecture to allow decision support tools to be customized to match the unique needs of individual stakeholders will be one of the more significant challenges to be overcome in the next 5 years.

5. ACKNOWLEDGEMENTS

Funding for the research described in this paper comes from various sources including the California Department of Water Resources, US Bureau of Reclamation, State Water Resources Control Board and the CALFED Bay-Delta Program. Special thanks to Ric Ortega and Lara Sparks at the California Department of Fish and Game, David Widwell at Grassland Water District, Ernie Taylor at the California Department of Water Resources, Lee Mao and Gene Lee at the US Bureau of Reclamation and Jim Martin at the California Regional Water Quality Control Board who are working with the author on implementation of the RTWQM concept.

6. REFERENCES

California Environmental Protection Agency. Total Maximum Daily Load for Salinity and Boron in the Lower San Joaquin River. Staff report by the Regional Water Quality Control Board, Central Valley Region. Sacramento, CA. 2002.

Fredrickson, L.H.,and T.S. Taylor. Management of seasonally flooded impoundments for wildlife. Resource Publication 148. U.S. Dept. of the Interior, Fish and Wildlife Service, Washington, D.C. 1982.

Jarrell W.M. Watershed Monitoring for Watershed Planning and TMDL's. Available online at : www.ysi.com/watershed-tmdl . 2003.

Kratzer, C.R., P.J. Pickett, E.A. Rashimawi, C.L. Cross, and K.D. Bergerson. An Input-Output Model of the San Joaquin River from the Lander Avenue Bridge to the Airport Way Bridge", Technical Committee Report No. N.Q. 85-1. State Water Resources Control Board, Sacramento, CA. 1987

Pate, T. DSM2 San Joaquin Boundary Extension : in Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh, 22nd Annual Progress Report to the State Water Resources Control Board. Department of Water Resources, Sacramento, CA. 2001.

Quinn, N.W.T., and J. Karkoski. Real-time management of water quality in the San Joaquin River Basin, California. American Water Resources Association, Vol. 34, No. 6. 1998.

Quinn, N.W.T. and W.M. Hanna. A decision support system for adaptive real-time management of seasonal wetlands in California. Environmental Modelling and Software, Volume 18, Issue 6, ppg. 503-511. 2003.

Herr, J. and C. W. Chen. San Joaquin River Model Calibration Report. CALFED project ERP-02D-P63. San Ramon, CA 94583. 2006.

Quinn, N.W.T., L.F. Grober, Jo-Anne Kipps, C.W. Chen, and E. Cummings. Computer Model Improves Real-Time Management of Water Quality, California Agriculture, Vol. 51, No. 5, pp. 14-20. 1997.

Yellow Springs International Inc. EcoNet Systems operations Manual, Yellow Springs, OH 45387. 2005.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

New Sensor System for Environmental Monitoring: the Potentiometric Electronic Tongue

M. Gutiérrez[a], J. M. Gutiérrez[b], S. Alegret[a], L. Leija[b], P. R. Hernández[b], R. Muñoz[b] and M. del Valle[a]

[a] Grup de Sensors i Biosensors, Departament de Química, Universitat Autònoma de Barcelona, Edifici Cn, E-08193, Bellaterra, Spain (manel.delvalle@uab.cat)
[b] Departamento de Ingeniería Eléctrica, CINVESTAV. Av. Instituto Politécnico Nacional 2508, 07360, Mexico D.F., Mexico (rmunoz@cinvestav.mx)

Abstract: The use of electronic tongues is proposed for two different environmental monitoring applications. This approach in chemical analysis consists of an array of nonspecific sensors coupled with a multivariate calibration tool. In this case, the proposed arrays were formed by potentiometric sensors based on polymeric membranes (PVC) and the subsequent cross-response processing was based on a multilayer artificial neural network (ANN) model. Special attention was paid in compensating temperature effects and response drifts in the sensors. In addition, in order to demonstrate the viability of the proposed systems for automated remote applications, the use of data transmission by radiofrequency has been tested.

Keywords: Sensor Array; Ion Selective Electrodes; Artificial Neural Networks; Electronic Tongue; Environmental Monitoring.

1. INTRODUCTION

Environmental monitoring and surveillance of surface waters is a major issue worldwide. Particular attention is dedicated to assess the integrity of natural waters, especially near potential pollution sites. The quick determination of contaminants and other species entering into the environment is necessary for taking control measures to confine any environmental damage. There are laboratory methods available for analysis of the different substances in water, featuring high precision and confidence, however, most of them need complex sample pretreatment and/or sophisticated equipment, e.g. high performance liquid chromatography, atomic spectroscopy, fluorimetry, etc. To get on-line and real time information about actual changes of the water composition, automatic systems for on-site analysis are highly desirable. The measuring principles employed within these systems must be robust, sensitive enough and of wide applicability. These requirements are needed because the monitoring sites may present differences in sample nature, i.e. surface, ground, waste, and sewage water, etc. and may present a broad scope of inorganic as well as organic target compounds, microorganisms, etc.

The use of chemical sensors for natural water monitoring has been successful in certain cases; a progress of the use of a single sensor is the use of multisensor systems; these can be based on arrays of non-specific sensors combined with advanced data processing of their complex signals, in the approach called electronic tongue [Vlasov et al., 2005]. The sensor array takes profit of the advantages of using chemical sensors such as speed of response, low cost of analysis and on-site determination with a relatively simple measuring setup. Also, the use of a sensor array permits to widen the number of species that can be determined simultaneously, while any interference effect can be solved employing

chemometric tools as data processing element. In this field, electronic tongues employing voltammetric sensors and Principal Component Analysis (PCA) [Krantz-Rülcker et al., 2001] have been applied to detect annoyances in drinking water quality and to monitor on-line industrial processes [Winquist et al., 2005]. In the same way, electronic tongues using an array of potentiometric sensors, specifically made of chalcogenide glasses, and combined with a pattern recognition routine, were used to detect heavy metals in river water, and were suggested for environmental- and process-monitoring purposes [Legin et al., 1996]. When this type of potentiometric sensors were combined with others based on poly(vinyl chloride) (PVC) membranes, heavy metals, alkali- and alkali-earth cations plus inorganic anions could be quantified in model solutions of groundwater [Rudnitskaya et al., 2001]. In our own experience, all-solid-state potentiometric sensors based on polymeric membranes were successfully applied for the simultaneous quantification of alkaline ions in river and waste waters using a multilayer artificial neural network (ANN) model where the different crossed interference effects could be solved [Gallardo et al., 2003; Gallardo et al., 2005].

This communication describes the development, application and testing of specifically designed electronic tongues, which employs potentiometric sensors and ANN models to develop environmental monitoring systems.

2. EXPERIMENTAL SECTION

The used sensors were based on PVC membranes and provided cross response to various species in solution. The ANNs were the multivariate calibration tools used to process the original complex data, being able to quantify the target analytes for which their cross-response interference effects had to be compensated. Two different applications were developed.

2.1 First Application in the CSTR

Firstly, a Continuous Stirred-Tank Reactor (CSTR) with a simulated surface water background was inoculated with natural microorganisms in order to imitate natural biodegradation conditions. While in operation, the simultaneous monitoring of ammonium, potassium, sodium, chloride and nitrate was accomplished during a perturbation lasting for several days, in this case, the addition of a concentrated liquid fertiliser to the whole, simulating a wastewater spillage. Figure 1 shows the whole proposed manifold for this first application.

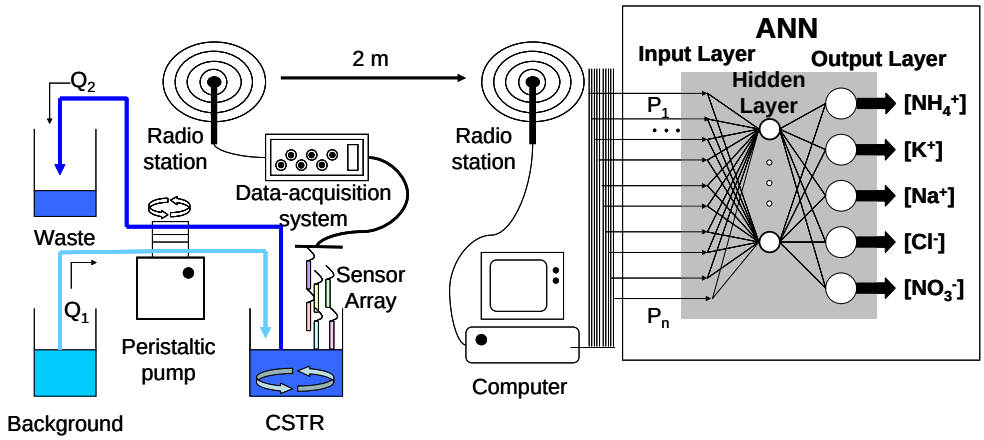


Figure 1. Block diagram of the proposed system for monitoring the concentration of ammonium, potassium, sodium, chloride and nitrate in the CSTR.

2.2 Second Application in the "Ignacio Ramírez" Dam

Secondly, an electronic tongue was optimised for on-site determination of ammonium, potassium and sodium ions in the "Ignacio Ramírez" dam (Mexico). Before this on-site application, the system was validated with a set of real water samples collected from four different points of the dam. Once filtered, different dilutions of these samples with Milli-Q water were processed by the electronic tongue. Results were compared with those provided by reference methods: ammonium was determined by the Nessler reaction, and potassium and sodium, by atomic absorption spectroscopy.

Figure 2 shows the whole proposed manifold for this second application. In both cases, special attention was paid in order to compensate temperature effects and sensor drifts. Apart, the possibility of wireless connection has been explored to demonstrate remote, unattended operation. A digital radio link was used between the monitoring site and the PC, where the data were processed employing a previously trained ANN model.

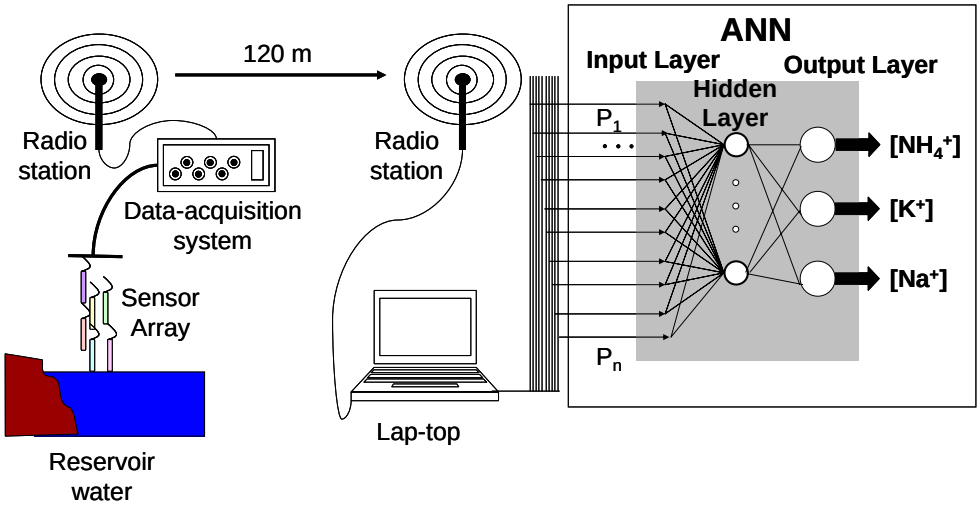


Figure 2. Block diagram of the proposed manifold for monitoring the concentration of ammonium, potassium and sodium in the "Ignacio Ramírez" dam.

3. RESULTS AND DISCUSSION

3.1 First Application in the CSTR

With the optimised ANN, we estimated the performance of the proposed system as the Relative Standard Deviation (% RSD) for a measurement interpolated in the training curve. The obtained values were 6.5% for ammonium determination, 15.2% for potassium, 9.4% for sodium, 2.3% for chloride and 2.4% for nitrate. In all cases, except for potassium, the RSD is lower than 10%, and specially good results were obtained for the determination of the two anions, chloride and nitrate.

In this first application, the concentrations levels of ammonium, potassium, sodium, chloride and nitrate, once the fertiliser was added, were continuously monitored for more than 3 days, from which the first 24 h were analysed in more detail. Figure 3 represents this detailed period of the experiment, where the concentrations of the considered ions predicted by the electronic tongue, together with the solution temperature, are shown. At the beggining of the experiment, the tank was inoculated with natural microorganisms from a stream to imitate real biodegradation conditions in a surface water. After 1 hour, the liquid fertilizer was added into the tank, and one measurement per sensor was done every 20 min.

The results demonstrated that the electronic tongue corrected the temperature effect, as observed by the smooth variation of concentrations. The Figure shows that ammonium

disappeared more quickly than the other species, that displayed equivalent behaviour. This fact suggested that a biochemical reaction was occurring inside the CSTR depleting ammonium. Considering that ammonium ion is the unique monitored substance that can be oxidized, this behaviour can be consequence of the natural biodegradation by the environmental microorganisms that were introduced in the tank.

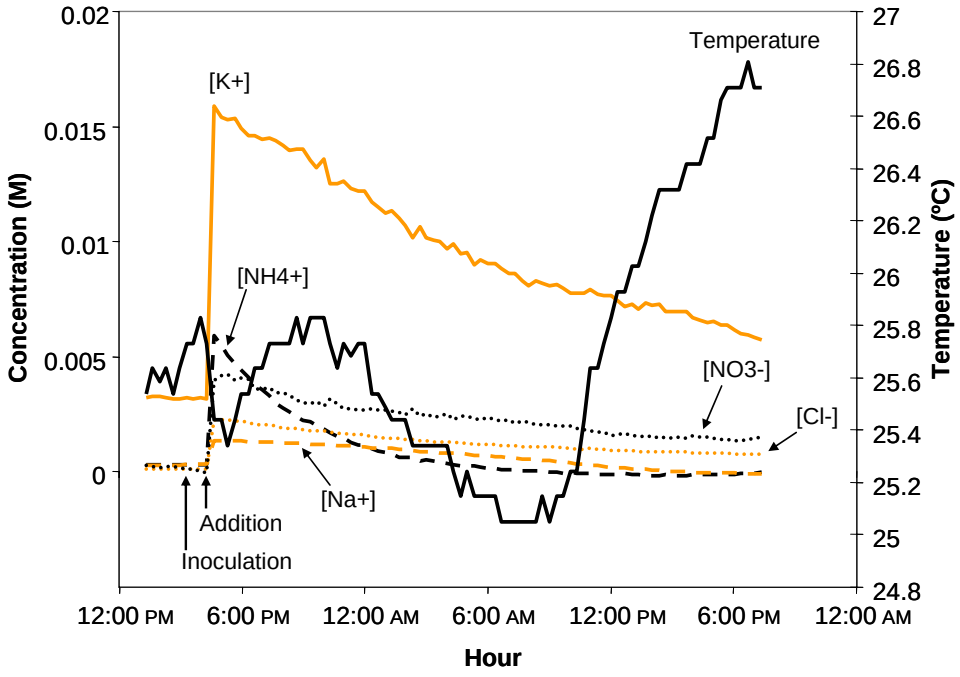


Figure 3. Representation of the concentration values predicted by the electronic tongue for the considered ions: ammonium, potassium, sodium, chloride and nitrate, in the CSTR during 24 h of continuous monitoring. The variation of temperature according to the cycle of day and night is also observed.

3.2 Second Application in the "Ignacio Ramírez" Dam

With the optimised ANN, we estimated the performance of the proposed system as before. The obtained RSD were 4.2% for ammonium determination, 2.8% for potassium and 3.1% for sodium. These values are comparable to those obtained for the used reference methods, which were between 0.8 and 1.3% for ammonium, and between 0.2 and 1.5% for potassium and sodium.

The proposed system was first applied to real samples from the dam. The predicted relative errors of the logarithm of the concentration of the three considered ions are summarized in Table 1. The errors are in specific cases higher than 10%, although lower on average. For example, the mean relative error was 3.2% in the determination of potassium. All the obtained errors for sodium and potassium were negative because of the influence of the sample matrix, which is highly saline and specially complex in this reservoir. The matrix effect is a special type of interference that caused a decrease or increase of the sensor response due to the presence of other species in solution. This effect specially affects the determination of sodium, the more concentrated ion, since a mean relative error of 8.4% was obtained.

Table 1. Obtained relative errors of the logarithm of the concentration of the three ions in real water samples from "Ignacio Ramírez" dam.

Sample	Dilution	Relative Error (%) NH_4^+	Relative Error (%) K^+	Relative Error (%) Na^+
Real 1	3:2	-1.6	-7.9	-10.7
Real 2	2:3	0.2	-2.6	-8.5
Real 3	2:3	-2.2	-2.0	-11.2
Real 4	2:3	-1.4	-1.2	-3.5
Real 1	3:2	12.9	-4.7	-9.7
Real 2	2:3	-2.2	-2.1	-7.9
Real 3	2:3	7.3	-1.1	-12.0
Real 4	2:3	11.9	-0.4	-3.6
Real 1	3:2	1.2	-8.9	-11.0
Real 2	2:3	-2.2	-2.6	-8.1
Real 3	2:3	-2.2	-3.2	-11.4
Real 4	2:3	-5.9	-2.3	-3.3
Mean Relative Error (%)		4.3	3.2	8.4

Once the system was validated, the concentration of ammonium, potassium and sodium was continuously monitored on-site for 2 h, approximately. Figure 4 shows the concentration of the considered ions that was predicted by the electronic tongue, together with the recorded dam water temperature. Sudden changes in temperature, observed in the graphs, were caused by the periodically measurement of the reference solution, that was at higher temperature. Nevertheless, the system is able to correct the effect of these temperature changes, given that alterations in the calculated concentrations of the three ions were not observed. The predicted concentration of sodium was around 0.006 M; for potassium, the obtained concentration was around 0.001 M with a light trend toward decreasing, while the ammonium concentration was between 1.5×10^{-5} and 4×10^{-5} M. The predictions of the ammonium and potassium contents were close to the concentrations determined by reference techniques for the real samples. However, the sodium concentration determined by the electronic tongue was lower. As we explained before, the matrix effect was specially strong for sodium, although its presence was correctly counterbalanced for the determination of ammonium and potassium employing the ANN model.

4. CONCLUSIONS

Two different electronic tongue systems have been developed and optimised for environmental monitoring applications using potentiometric sensors and ANN models; one, to monitor the concentration of ammonium, potassium, sodium, chloride and nitrate inside a CSTR that imitates real biodegradation conditions in surface waters. The second one, to monitor the concentration of ammonium, potassium and sodium in the "Ignacio Ramírez" dam (Mexico). For the first application, it has been demonstrated that the electronic tongue approach allows to monitor accurately the dilution of the five considered ions, as well as the effect of the natural biodegradation. For the second application, the results show that the proposed electronic tongue allows to monitor the content of ammonium, potassium and sodium in real dam water, but the calculated concentration of sodium is lower than the obtained by reference methods due to the high matrix effect. To improve the response towards this cation, an alternative would be to increase the number of sensors used in the array, incorporating new species not considered here to the model. It is important also to establish the training space before designing the electronic tongue in order to cover the variability and the complexity of the real conditions. Therefore, the calibration medium needs a exhaustive study to match better the real samples.

In any case, the two systems are able to compensate natural temperature variations by incorporating the solution temperature as input in the ANN model. Although the two considered experiments show a measuring time of several hours, in previous works we have checked that the applicability of the numerical model, in terms of aging of the electrodes, for a system like this is longer than one month [Gutiérrez et al., 2007]. The used radio link demonstrated a robust operation, so we can conclude that systems of the studied type can be applied to automatic wireless monitoring for environmental applications.

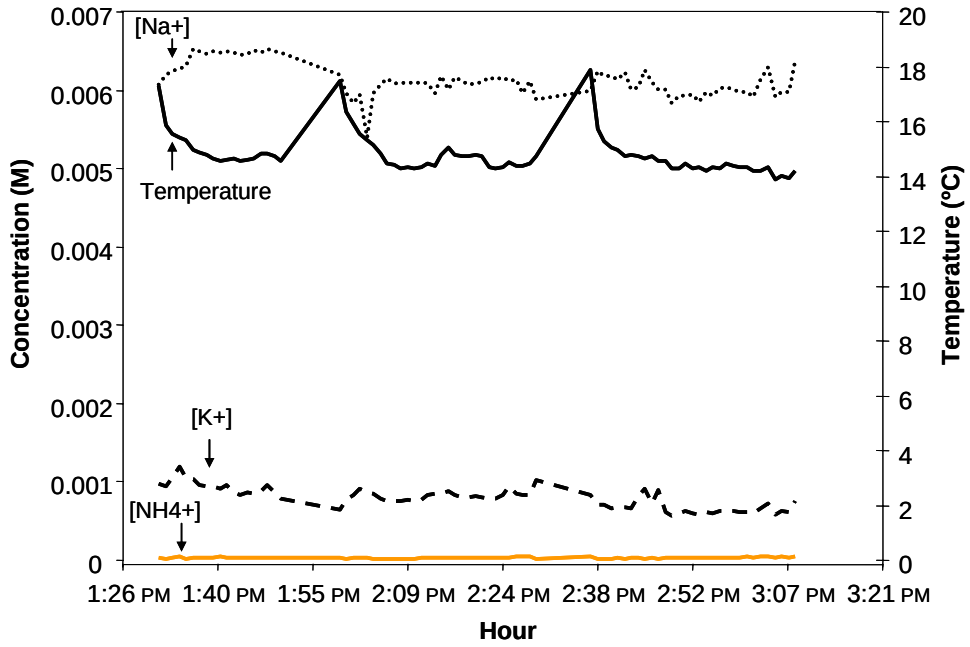


Figure 4. Representation of the concentration values predicted by the electronic tongue for the cations considered: ammonium, potassium and sodium, during the application in the "Ignacio Ramírez" dam. The monitorized water temperature is also represented.

ACKNOWLEDGEMENTS

The authors wish to thank the Spanish Ministry of Education, Science and Sport for the financial support of the project TEC2007-68012-C03-02/MIC, the Mexican National Council of Science and Technology CONACYT and the Department of Innovation, Universities and Enterprise (DIUE) from the Generalitat de Catalunya.

REFERENCES

Gallardo, J., Alegret, S., de Román, M.A., Muñoz, R., Hernández, P.R., Leija, L., and del Valle, M., Determination of ammonium ion employing an electronic tongue based on potentiometric sensors, *Analytical Letters*, *36*(14), 2893-2908, 2003.

Gallardo, J., Alegret, S., Muñoz, R., Leija, L., Hernández, P.R., and del Valle, M., Use of an electronic tongue based on all-solid-state potentiometric sensors for the quantitation of alkaline ions, *Electroanalysis*, *17*(4), 348-355, 2005.

Gutiérrez, M., Alegret, S., Cáceres, R., Casadesús, J., Marfà, O., and del Valle, M., Application of a potentiometric electronic tongue to fertigation strategy in greenhouse cultivation, *Computers and Electronics in Agriculture*, *57*(1), 12-22, 2007.

Krantz-Rülcker, C., Stenberg, M., Winquist, F., and Lundström, I., Electronic tongues for environmental monitoring based on sensor arrays and pattern recognition: a review, *Analytica Chimica Acta*, *426*(2), 217-226, 2001.

Legin, A., Vlasov, Y., Rudnitskaya, A., and Bychkov, E.A., Cross-sensitivity of chalcogenide glass sensors in solutions of heavy metal ions, *Sensors and Actuators B, 34*(1-3), 456-461, 1996.

Rudnitskaya, A., Ehlert, A., Legin, A., Vlasov, Y., and Büttgenbach, S., Multisensor system on the basis of an array of non-specific chemical sensors and artificial neural networks for determination of inorganic pollutants in a model groundwater, *Talanta, 55*(2), 425-431, 2001.

Vlasov, Y., Legin, A., Rudnitskaya, A., Di Natale, C., and D'Amico, A., Nonspecific sensor arrays ("electronic tongue") for chemical analysis of liquids, *Pure and Applied Chemistry*, *77*(11), 1965-1983, 2005.

Winquist, F., Bjorklund, R., Krantz-Rülcker, C., Lundström, I., Östergren, and K., Skoglund, T., An electronic tongue in the dairy industry, *Sensors and Actuators B, 111-112*(11), 299-304, 2005.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Particle size distributions in an office environment: An experimental approach and a source apportionment analysis through positive matrix factorization.

Saraga D.[(a)], Sfetsos A. [(a)], Andronopoulos S. [(a)], Bartzis J. [(b)], Maggos T. [(a)], Vasilakos C. [(a)]

(a) Environmental Research Laboratory, Institute of Nuclear Technology - Radiation Protection, NCSR "DEMOKRITOS", 15310 Aghia Paraskevi Attikis, Athens, Greece (dsaraga@ipta.demokritos.gr)

(b) University of West Macedonia Dept. of Eng. & Manag. of Energy Resources, Sialvera & Bakola Street, 50100 Kozani, Greece

ABSTRACT

The identification of particulate matter sources is an important step in the development of indoor air quality control strategies. Advanced modelling techniques, termed as receptor modelling, are used to identify the presence, determine the sources of aerosols and most importantly quantify the source contributions. This study presents a source apportionment analysis, supported by targeted experimental campaign, to identify the main pollution sources in an office indoor environment. For the purpose of the experimental campaign, continuous monitoring of particle (0.3 - > 20.0 μm in aerodynamic diameter) number concentrations took place during the period between 16th and 27th July 2007 in two office environments, in Athens, Greece. A portable dust monitor (GRIMM model 1.108) was used to measure in real-time the particles size distribution in 16 different size channels. Two adjacent offices -with similar characteristics except for the presence of smokers- were used. Additionally, the contribution of other possible sources (e.g. photocopy machine, air conditioner, penetration and outdoor environment) was investigated. The indoor temperature, relative humidity and light intensity levels were also recorded in a continuous basis. Finally, all activities were daily recorded in detail on a logbook. The positive matrix factorization (PMF) method has been applied on the data collected during the experimental campaign for source apportionment analysis. The output of the analysis has been related with source information provided by the occupants and other source profiles by bibliography. This study aims to elicit indoor source characteristics based on particles size distribution data, especially in situations where the indoor sources are unknown beforehand.

Keywords: particulate matter; size distribution; positive matrix factorization.

1. INTRODUCTION

Several studies underline the role of good indoor air quality, since an average person spends more than 80% of the day in an indoor environment (Robinson and Nelson, 1995) Indoor air quality depends on several factors e.g. outdoor concentrations, indoor sources, ventilation, building design etc. These factors play different roles, depending on the kind of the indoor environment (Hameri et al., 2003). For an occupied building, indoor particle sources such as smoking and resuspension from clothes etc. are often prevalent while outdoor particles are often from both mobile and stationary (natural or anthropogenic) sources. Each of these particles sources produces different size distributions that may result

in studying the mortality and morbidity effects of airborne particulate matter (PM) because they will have different penetrability and deposition patterns.

The investigation of chemical pollutants in the air of public utility buildings, including offices has been the aim of a lot of studies for several years (Brown 1999, Jankowksa et al., 2004). Smoking seems to be an important determinant of personal exposures in such an indoor environment as PM and Volatile Organic Compounds (VOC) levels have been found even 2 times higher among a tobacco-smoke exposed working group compared to a non-smoke exposed group (Lai et al., 2004). Factors related to the use of printers and photocopy machines (Guo et al., 2003), to the building and constructing materials (Pellizari et al., 1987), to outdoor sources etc have been examined in the atmosphere of an office environment. Fine particles which were found in office buildings and other non-residential buildings were most of outdoor origin (Ogulei et al., 2006 [a]). When low or no activity was present in the building, such as when the house was unoccupied, Kopperud et al., 2004 observed that up to 60 % of the observed PM2.5 originated indoors.

The identification of air pollution sources is an important step in the development of air quality control strategies. Abatement strategies may significantly improve the air quality after identifying the main aerosol sources. *Receptor modelling* has been used to determine sources of aerosols in ambient or indoor environments. Receptor models *identify the presence* and *quantify of source contributions* to the receptor.

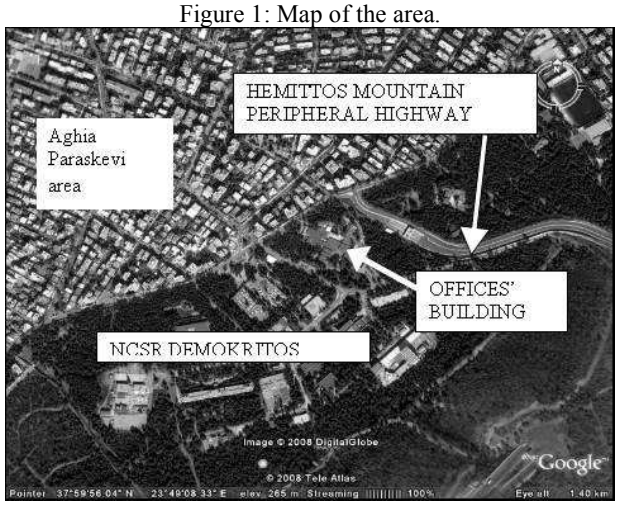
The fundamental principle of receptor modelling is based on the assumption that *mass is conserved* and that on this basis a mass balance analysis can be used to identify and apportion sources in the atmosphere. All source apportionment (receptor modelling) techniques are based on assumptions regarding the source the chemical or physical characteristics (e.g. chemical composition and size distribution) and the measurement methodology. Some of the methods require a certain degree of knowledge about the number of sources, the sources profiles and the source strength regardless the origin of the source (outdoors or indoors).

It is the purpose of this study to present the results of Positive Matrix Factorization (PMF) method application on the data collected during an experimental campaign for source apportionment analysis. The output of the analysis has been related with source information provided by the occupants and other source profiles by bibliography. This study aims to elicit indoor source characteristics based on particles size distribution data, especially in situations where the indoor sources are unknown beforehand.

2. DESCRIPTION OF THE FIELD CAMPAIGN

For the purpose of the study, two offices with similar characteristics were employed. The two offices are located in the same building, in NCSR ″Demokritos″ which is situated in Aghia Paraskevi, a north-eastern suburb of Athens (fig.1). The experimental campaign took place between 16th and 27th July 2007. As it has been reported, the kind of building and furniture materials, the equipment and the external environment is similar and the major factor which differentiates the two environments is smoking activity. In the first office, where the smoking is permitted, there are two employees; a smoker and a passive smoker. The second office is used by two, non-smoker employees and smoking activity is not allowed. Both offices have an air conditioning system, two computers, a printer and a photocopy machine. It has to be mentioned that both offices are daily open to public, while during weekend no activity is occurred. Important information is also that all employees were following the same working time schedule, as recorded in questionnaires.

Figure 1: Map of the area.



2.1 Experimental results

For the particle number measurements/size range an automatic portable aerosol spectrometer (GRIMM 1.108) was used in the offices. This unit is based on the principle of light scattering and can give the size distribution of dust particles in μg/m^3 or in counts/L.
Regarding particle concentration (counts/L) measurements, a different picture is observed between the two offices. For the smokers' office, the number of smaller particles (with diameter less than 0.40 μm) reached the value of 6,439,100 particles per litre when smoking activity occurred and windows were kept closed. During hours with open windows and the same intensity of smoking activity, the number of these particles was lower (1,084,730 particles/L). Furthermore, in the early morning hours (before 8.00am) the average concentration was 63,111 particles/L while in the afternoon hours -and after the occupants' leaving- the concentration was 94,287, following a decreasing trend during night. During weekend, the size distribution of the number particles concentration presents a different picture, possibly due to the absence of the main anthropogenic sources. In particular, the concentration of the smaller particles (<0.40 μm) presented a peak two orders of magnitude lower than this in weekdays (87,510 particles/L) and did not follow a variation because of the stable activity and ventilation conditions.
In non-smokers' office, smaller particles' number concentration during occupants' absence was in average, 90,920 particles/L (background). During working hours, concentration reached the value of 449,379 particles/L but significantly decreased when windows were open (190,150 particles/L). The same picture as in smokers' office was noticed for larger in size particles which were much fewer in number, reaching the order of 10 particles/L (for particles >20 μm).

3 PMF & data handling

All data obtained by GRIMM were averaged to 5 min resolution so as to eliminate wide changes in concentrations that could severely distort results of any apportionment study. As size distributions from a specific source are expected to remain reasonably stable over a short time interval, averaging the number or volume concentrations in this way was not expected to considerably affect the outcome of this study (Ogulei et al., 2006).

In this study, PMF was used with the GRIMM data. PMF is a new variant factor analysis method and is described in detail by Paatero (1997). Only a brief description of the technique is given here. PMF uses a weighted least-squares fit with the known error estimates of the elements of the data matrix used to derive the weights. The factor model (PMF2) can be written as

$$X=GF+E, \quad (1)$$

where X is the known $n\times m$ matrix of the m measured chemical species in n samples. G is an $n\times p$ matrix of source contributions to the samples (time variations). F is a $p\times m$ matrix of source compositions (source profiles). Both G and F are factor matrices to be determined. E is defined as a residual matrix, i.e., the difference between the measurement X and the model Y as a function of factors G and F.

$$e_{ij} = x_{ij} - y_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj} (i = 1, \ldots, n;\ j = 1, \ldots, m;\ k = 1, \ldots, p). \quad (2)$$

The objective of PMF is to minimize the sum of the squares of the residuals weighted inversely with error estimates of the data points. Furthermore, PMF constrains all of the elements of G and F to be non-negative; meaning that sources cannot have negative species concentration ($f_{kj} \geqslant 0$) and sample cannot have a negative source contribution ($g_{ik} \geqslant 0$). The task of PMF analysis can thus be described as to minimize Q, which is defined as

$$Q(E) = \sum_{i=1}^{n}\sum_{j=1}^{m} (e_{ij}/s_{ij})^2 \quad (3)$$

with $f_{kj} \geqslant 0$; $g_{ik} \geqslant 0$ and s_{ij} is the error estimate for x_{ij}. The solution of Eq. (3) is obtained by a unique algorithm PMF2 in which both matrices, G and F, are adjusted in each iteration step. The process continues until convergence (Paatero 1997).

A critical step in PMF analysis is determination of the number of factors. It is a fact that choosing too few factors may lead to non-well separated sources, whereas too many factors may essentially lead to split up of a true source into two or more non-existing sources.

The primary consideration is basically to obtain a good fit of the model to the original data. The theoretical Q-value should be approximately equal to the number of degrees of freedom, or approximately equal to the number of entries of data array, provided that correct values of s_{ij} have been used in equation 3. Often there is no reliable information on s_{ij} and the presence of outliers (even in the robust mode) complicates the situation. Then it may be impossible to determine whether the observed value of Q is normal or too large. It is helpful to examine the distributions of scaled residuals (e_{ij}/s_{ij}). In a well-fit model, the residuals e_{ij} and the error estimates s_{ij} should be about equal and the ratios e_{ij}/s_{ij} should fluctuate between ±2 (Juntto and Paatero, 1994).

4. RESULTS AND DISCUSSION

In the frame of this study, one day's PMF application results are presented as six factors were successfully identified. The selected day was Thursday, 19th July 2007 where the measurements were conducted in smokers' office. However, information from PMF application results from other days of the experimental campaign was used. In this work, the number of factors that was examined ranged from four to seven, although a number between five and six seems more appropriate. The calculated Q-values obtained from trials with five or six factors did not show that either was much better than the other.

The final number of factors, and therefore existing sources, were identified by relating the obtained source contributions to the source information provided by the occupants (logbook with detailed activities). The subjective analysis together with the application of the robust PMF mode and with rotation resulted in six factors for all experimental days. The *source contributions* and *time series concentrations* are presented in Table1 and figures 2a- f, respectively.

Table 1. Percentage contribution of particle channel to each Factor

Channel	Size range (μm)	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	sum
1	**0.30-0.40**	4.6	10.2	9.8	5.7	1.1	**68.6**	100%
2	**0.40-0.50**	5.4	11.5	5.1	25.3	0.5	52.4	100%
3	**0.50-0.65**	4.0	19.0	1.8	**39.4**	1.8	33.9	100%
4	**0.65-0.80**	0.9	35.0	5.1	29.8	6.8	22.3	100%
5	**0.80-1.0**	0.0	**44.7**	13.0	8.9	13.1	20.3	100%
6	**1.0-1.6**	25.9	33.8	27.6	0.6	0.0	12.1	100%
7	**1.6-2.0**	0.0	24.4	45.1	1.6	21.1	7.8	100%
8	**2.0-3.0**	23.0	6.0	**47.3**	0.0	21.5	2.2	100%
9	**3.0-4.0**	29.4	0.6	28.8	1.1	40.0	0.0	100%
10	**4.0-5.0**	24.8	0.0	20.5	0.0	53.4	1.4	100%
11	**5.0-7.5**	21.0	0.5	8.8	1.9	**67.4**	0.5	100%
12	**7.5-10**	72.5	0.0	0.0	0.0	24.4	3.1	100%
13	**10.0-15.0**	**84.0**	1.3	12.8	0.0	0.0	1.9	100%
14	**15.0-20.0**	68.8	0.0	0.0	11.8	19.4	0.0	100%

Factor 6 has been identified as **tobacco smoking**. This factor has its strongest contribution at smaller particles (0.30-0.40μm and 0.40-0.50μm), that is in ultra fine particles as expected (He et al., 2004). Time series concentration shows that this factor exists only during working hours and presents picks which match with the periods that smoking activity actually occurred, as recorded in the logbook. Additionally, concentration picks seem to decrease with low rate because cigarette smoke tends to remain even for few hours after cigarette smoking (Halios et al., 2005).

Factor 3 has a number mode at 2.0-3.0μm and is suggested to represent **primary PM emissions from office equipment and furniture**. It is remarkable that this factor remains almost constant; presenting a slight fluctuation during day and night, a fact that amplifies the above suggestion. As it can be noticed, a more intense fluctuation can be observed during working hours, where human activities occurred and air streams could affect concentration at the receptor site.

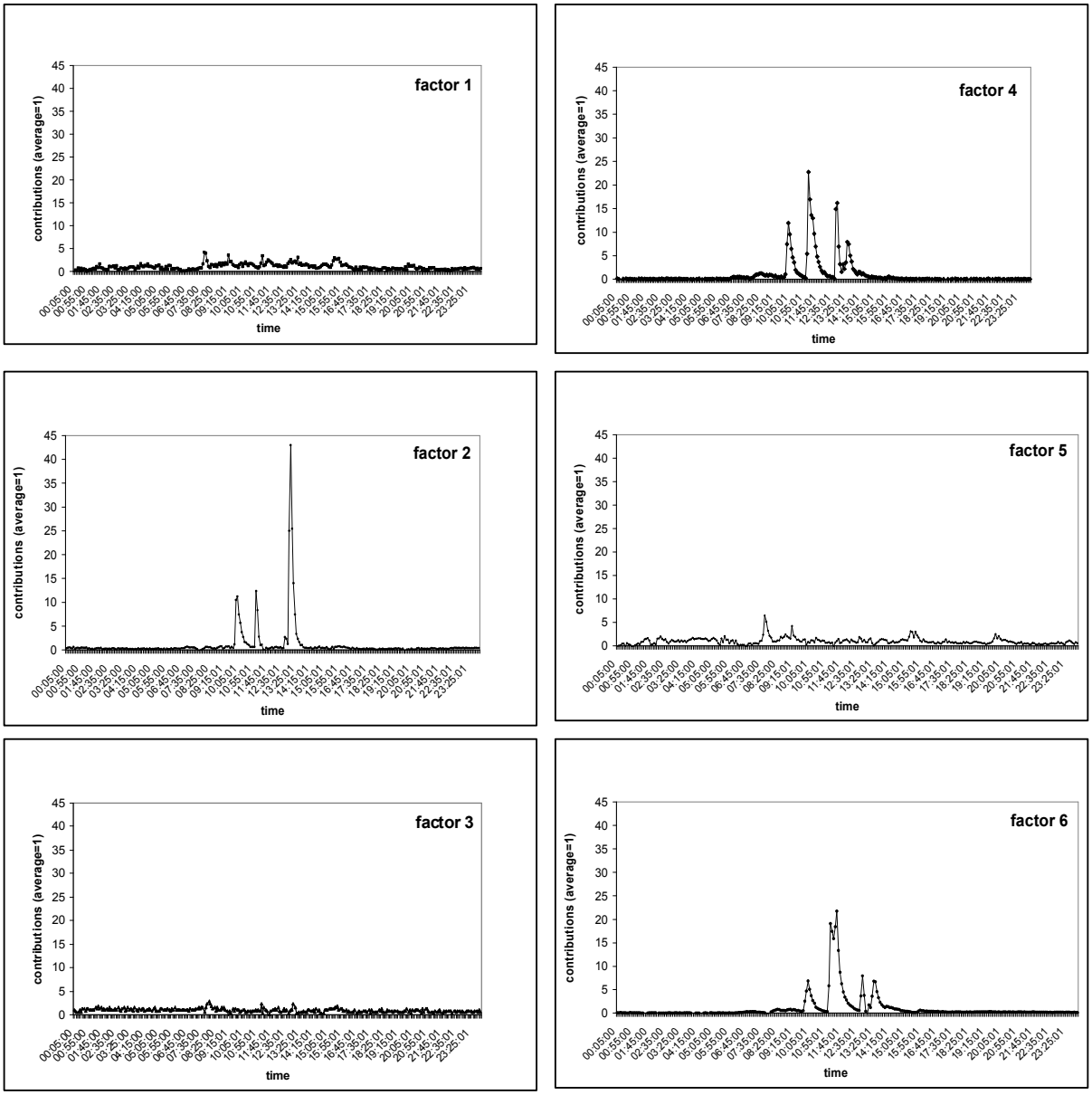
Factor 2 has its strongest contribution at particle size range of 0.80-1.0μm. Time series concentration presents picks during working hours that do not match in number and time with the real cigarettes smoked. This fact, in combination with information received from the other experimental days should lead to another factor. Actually, in the same office, during weekend (when no human activity occurred) a test of air conditioner 12-hour function took place. PMF results (not shown here) presented a new factor, which has a number mode at 0.80-1.0μm, too. Furthermore, similar picks were found during this 12-hour interval. Taking all the above into consideration, we can assume that Factor 2 is associated with **air conditioner emissions**, as its presence match in time with the air conditioner function as recorded in questionnaires.

Factors 1 and 5 are suggested to represent **outdoor sources** e.g. vehicles exhausts emissions, construction activities, pine trees etc as they are associated with larger particles, 10-15.0μm for factor 1 and 5.0-7.5μm for factor 2. For both factors, time series present small fluctuations during all day except for the period of working hours, when opening of windows occurred several times and air exchange took place. Outdoor environment contribution could possibly happen either by air exchanging through the window or by infiltration (cracks, small openings in doors and windows etc) or by the combination of all these. It is noticeable that these factors appeared in all days as expected, since outdoor environment is a constant source. Especially for factor 1 (major contribution at the size

range of 10-15μm), an additional observation amplifies the assumption that it is associated with vehicles emissions, possibly from Hemittos mountain peripheral highway (see fig. 1), an avenue in short distance to the building (100m). To be more specific, time series of this factor during all days suggests a frequent activity from 6.30am to 11.00pm that is the hours that vehicles circulation occurs. Furthermore, picks are presented during rush-hours, even on evening indicating the external origin of the source.

Factor 4 presents its major contribution at the size range of 0.50-0.65μm and can attributed to **penetration**. In particular, it can be associated with particles originated outdoors with have been transferred and remained indoors. Penetration sources can mainly associated with accumulation mode particles probably because those particles can persist in the air since they are too small for inertial deposition and too big for diffusive removal processes and, hence, are capable of penetrating into buildings and remaining airborne for longer periods of time. Otherwise, this factor can be associated with the **unapportioned** PM fraction which represents the fraction that could not be resolved by the model. It is possible that this fraction may represent background levels and/or other unresolved sources.

Figures 2 a- f Time series concentration (19/7/2007)



5. CONCLUSIONS

PMF method has been successfully applied on the PM data collected during the experimental campaign in an office environment for source apportionment analysis. Six size distribution profiles were resolved, using information kept in log books by the occupants. In general, the profiles identified are in agreement with literature references. The six profiles included tobacco smoking, office equipment emissions, air conditioner emissions, penetration, vehicle exhausts emissions, other outdoor sources.

In a source apportionment study one should take under consideration some errors that can be the result of several critical problems. In a study like this, someone must be very careful about the information collected on the logbook because either participants can very subjectively estimate the duration of smoking/other activities or minutes of smoking can not be a good indicator for PM emission from smoking (Yakovleva et al., 1999).

Finally, another source apportionment method (e.g. chemical mass balance method) using results from the chemical analysis of particles collected in such an experimental campaign could provide an additional aspect of the study. As reported (Ogulei et al., 2006 [b]), once particles are emitted from a given source, their size, number, and chemical composition change by several mechanisms until they are ultimately removed by natural processes. However, the particle size distribution would be expected to remain approximately stationary at some appropriate distance from the emission source. Thus, properties of the observed size distribution at the receptor site can be used along with chemical composition information to identify the emission sources. This approach has been recently explored by Zhou et al., 2005.

REFERENCES

Brown S.K. Assessment of pollutant emissions from dry-process photocopiers. *Indoor Air 9:259-267*, 1999.

Guo H., Lee S.C, Li W.M., Cao J.J. Source characterization of BTEX in indoor microenvironments in Hong Kong. *Atmospheric Environment 37,* 73-82, 2003.

Halios, C.H., Assimakopoulos, V.D., .Helmis, C.G., Flocas, H.A.. Investigation cigarette-smoke indoor pollution in a controlled environment. *Science of the total environment 337*, 183-190, 2005.

Hameri, K., Lahde, T., Niemela, R., Korhonen P. Fine aerosols indoors and outdoors in downtown Helsinki. Abstracts of the European Aerosol Conference 2003

He, C., Morawska L., Hitchins J., Gilbert D. Contribution from indoor sources to particle number and mass concentrations in residential houses. Atmospheric Environment 38 3405-3415, 2004.

Jankowksa, E., D., Kondej, D., Jankowski T., Glinski M. Fine particles suspended in office air. Abstract of the European Aerosol Conference 2004.

Juntto S., Paatero P. Analysis of daily precipitation data by positive matrix factorization. Environmetrics 5 (2),127-144, 1994

Kopperud, R.J., Ferro, A.R. and Hildemann, L.M. Outdoor versus indoor contributions to indoor particulate matter (PM) determined by mass balance methods, *J. Air Waste Manage. Assoc*., 54, 1188-1196, 2004.

Lai, H.K., Kendall, M., Ferrier, H., Lindup, I., Alm, S., Hanninen, O., Jantunen, M., Mathys, P., Colvile, R., Ashmore, M.R., Cullinan, P., Nieuwenhuijsen M.,J. Personal

exposures and microenvironment concentrations of PM2.5, VOC, NO2 and CO in Oxford, UK. *Atmospheric Environment 38*. 6399-6410, 2004.

Ogulei, D., Hopke P.K., Wallace L.A. Analysis of indoor particle size distributions in an occupied townhouse using positive matrix factorization. *Indoor Air; 16:* 204-215, 2006 [a].

Ogulei, D., Hopke, P.K., Zhou, L., Pancras, J.P., Nair, N., Ondov, J. Source apportionment of Baltimore aerosol from combined size distribution and chemical composition data. *Atmospheric Environment 40* S396-S410, 2006 [b].

Paatero, P. Least squares formulation of robust non-negative factor analysis. *Chemometr. Intelligent Lab. Syst., 6037,* 23-35, 1997.

Pellizari, E., Leaderer, B., Zelon, H., Sheldon, L. Emissions of volatile organic compounds from building materials and consumer products. *Atmospheric Environment 21,* 385-393, 1987.

Robinson, J., Nelson WC., National human activity pattern survey data base. *Research Triangle Park, NC: United States Environmental Protection Agency*, 1995.

Yakovleva, E., Hopke, P., Wallace L. Receptor modeling assessment of particle total exposure assessment methodology data. *Environ. Sci. Technol., 33,* 3645-3652, 1999.

Zhou, L., Hopke, P.K., Stanier, C.O., Pandis, S.N., Ondov, J.M., Pancras, J.P. Investigation of the relationship between chemical composition and size distribution of airborne particles by partial least squares (PLS) and positive matrix factorization (PMF). *Geophysical Research, VOL. 110, D07S18, doi:10.1029/2004JD005050,* 2005.

.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Development of an Information System for the Administration of Environmental Monitoring Data

J.Ráček[a], T. Ludík[a], V. Kubík[b], J. Ministr[c], J. Sedláčková[d] and J. Fiala[c]

[a] *Faculty of Informatics, Masaryk University, Brno, Czech Republic*

[b] *Faculty of Science, Masaryk University, Brno, Czech Republic*

[c] *Faculty of Economics, VŠB-Technical University, Ostrava, Czech Republic*

[d] *Faculty of Information Technology, University of Technology, Brno, Czech Republic*

Abstract: This paper presents the current state of the information system for administration of environmental monitoring data developed by Masaryk University in Brno (MU). The rising environmental database of MU can be considered as unique for the type and the quantity of stored data. The information system running over this database allows the management of monitoring projects, localities and samples and viewing of various data as well as exporting and simple analysis of the data. The paper describes the basic architecture of the developed information system and shows the most important parts of its data model. The system prototype is implemented as a web application for internal use at MU.

Keywords: Database system; Environmental data processing; Software development.

1. INTRODUCTION

To the main activities of the Research Centre for Environmental Chemistry and Ecotoxicology (RECETOX) of Masaryk University in Brno (Czech Republic), belong long-term environment pollution monitoring and ecotoxicological analysis of real matrices. The contaminants content and many different other parameters are measured in five basic matrices – soils, sediments, water, air and biota. A major part of the data set is formed from POPs and heavy metal (HMs) concentrations, physico-chemical parameters and the results of different ecotoxicological tests. The data collection has been proceeding in many parallel campaigns since the eighties, thus lot of data is available. Therefore a central information system for environmental monitoring data is needed. RECETOX started with the development of new data management system in 2006 and its purpose is to standardise and store data in a unified data format and allows data visualization in the form of predefined tables, graphs and maps.

2. RANGE OF ICT SUPPORT OF ENVIRONMENTAL MONITORING

Environmental projects of the RECETOX deal not only with sampling at selected localities and the laboratory analysis of samples, but they also include the data management, statistical support and visualisation of the results. These main points are fundamental to implement for each environmental monitoring program:

- Development of monitoring plan
- Cost calculation
- Identification of appropriate and feasible monitoring locations
- Identification of equipment requirements, procurement, and training
- Analytical laboratory assessment

- Sophisticated statistical analysis
- Data Quality Assurance/Quality Control (QA/QC)

A separate problem of the environmental monitoring is how to collect and manipulate the data or outputs (GIS maps, tables, graphs) from partial stages of the monitoring process and access to the data is according to property rights commonly differentiated into several levels.

Our information system does not support any aspects of the environmental monitoring. It is mostly oriented to the administration of monitoring projects data, localities data and samplings data. It also includes the set of basic statistical functions for data analysis, data interface to professional statistical software (SPSS and Statistica) and interface to the GIS (ArcGIS and MapInfo).

The following figure shows the recommended process for the editing, processing and visualization of data using the information system of the RECETOX.

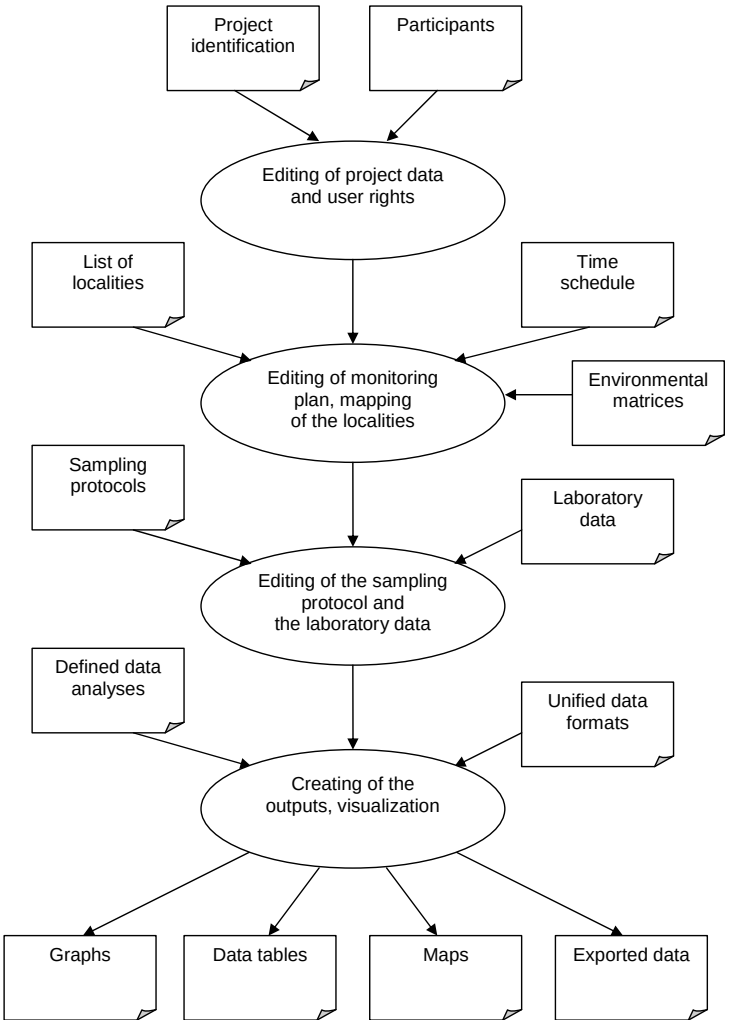


Figure 1. The main system process

We recommend inputting the data of the new monitoring project at first if you like to use the RECETOX information system for supporting your monitoring process. Then it is necessary to assign user rights for the new project. After it, the localities, matrices and sampling schedule can be assigned to the monitoring project.

Users send the data of sampling protocols (related to conditions at selected localities) together with laboratory results to the information system during the whole monitoring season. The last step of data processing is the creation of outputs in the form of graphs, maps, data tables and exchange data formats.

3. DATA MODEL

The ICT solution developed by Masaryk University is based on the relation data model. The data entities are separated into three basic parts:

- Part of sampling data – the data of individual samplings.
- Part of projects and localities – the data about monitoring projects and localities (areas) of samplings.
- Administration part – the data about users and executed changes in the database (logs).

Just the one locality for any sample has to be in the system. Localities and their samplings should be assigned to one or more monitoring processes. In the next articles are shown two basic parts of the database information system used for the processing of environmental monitoring data.

3. 1 Sampling part

The data entities of the sampling part, together with some entities from the part of projects and localities, are the core of the system data structure.

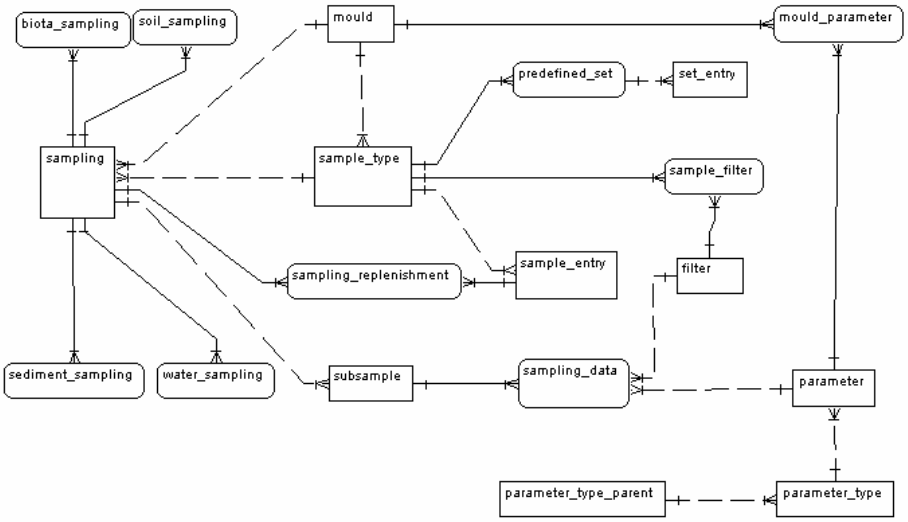


Figure 2. The data model of the sampling part

The most important data entities of this system part are:

- *Sampling* - This is the one of the most important entities in the system data model. It represents the realised samplings in the particular projects in the particular areas (localities). There exists a real sampling protocol for any sampling data set in the database.
- *Biota_sampling*, *Soil_sampling*, *Sediment_sampling*, *Water_sampling* - These entities contain the extensible data of sampling protocol of every particular (environmental matrices). It is the sampling depth for the soil mould, the biological material type for the biota mould and the sample amount and the sampling technology for the water mould. The entity of sediment contains data about the sampling amount, sample description and description of the sampling method, also for example temperature, pH, water flow, depth and sampling place and metrological conditions.

- *Sampling_data* - There are individual relationships *subsample – parameter – filter* and mainly measured value in this entity. The measured values of sampling are identified by the univocal key from attributes *filter_id*, *parameter_id* and *subsample_id*.
- *Subsample* - This entity represents subsamples with different modification and manipulation in a laboratory.
- *Sample_type* – The specific sample types of the particular mould are stored in this entity.
- *Parameter* - This entity is used as the set of all measured sample parameters (chemical element, chemical compound, etc.). It also contains the parameter name and its unit of measurement.
- *Filter* –The filters are created mainly for air mould.

3.2 Project and locality part

The data structure of the monitoring projects and localities is shown in the next figure.

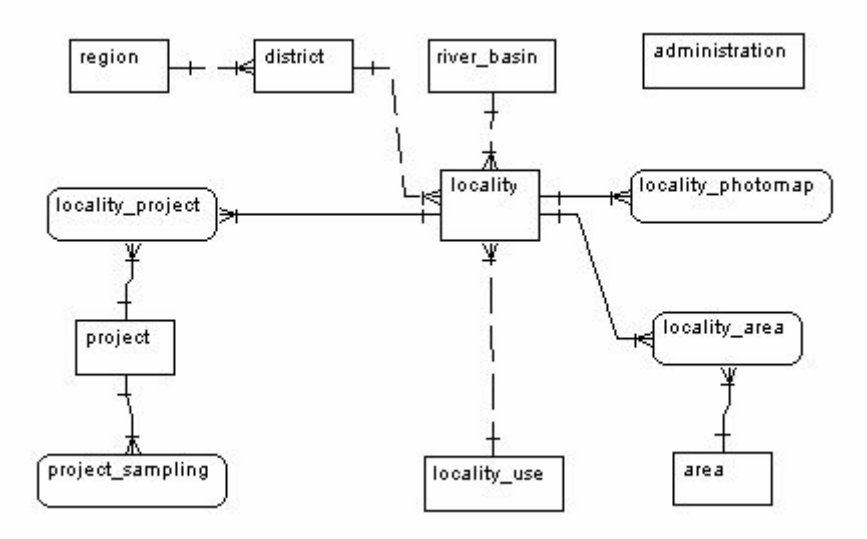


Figure 3. The data model of the project and locality part

The most important data entities of this part are:

- *Project* – There is the data about the particular project of the environmental monitoring in this entity.
- *Locality* - The entity stores information about localities where some monitoring is done and where any sampling campaign is made.
- *Locality_photomap* – The system is able to save photos and maps of the locality. It contains the picture name, its description, suffix, relationship to the picture and description of the type (for example satellite map).
- *Region*, *District*, *River_basin*, *Area* - The entities represent the geographical affiliation of the locality. The region is the first level of the public administration segmentation and the district is the second level. The area is defined as a group of localities from the point of view of the concrete user (for example group of localities in mountains or localities beside a highway).
- *Administration* - This entity saves the personal information of the project administrators.
- *Locality_use* - It is the codebook describing how the locality is used and what is there (built-up area, forests, vineyards, rocks, watercourses, meadows, technical areas, etc.).

4. IMPLEMENTATION

The system is implemented as a web based application and is accessible in the intranet of Masaryk University. PHP and Java are used as the two main programming languages and Oracle 9 is used as the database platform.

Behind the administration of the data of samplings, projects and localities the system provides other services. It can import data directly from selected laboratory machines and export data to the external data analysis. Currently tolls for the GIS connection and visualisation of the measured data are being developed.

The system also enables the functions of the predefined data analysis. There is shown an example of this analysis on the next figure.

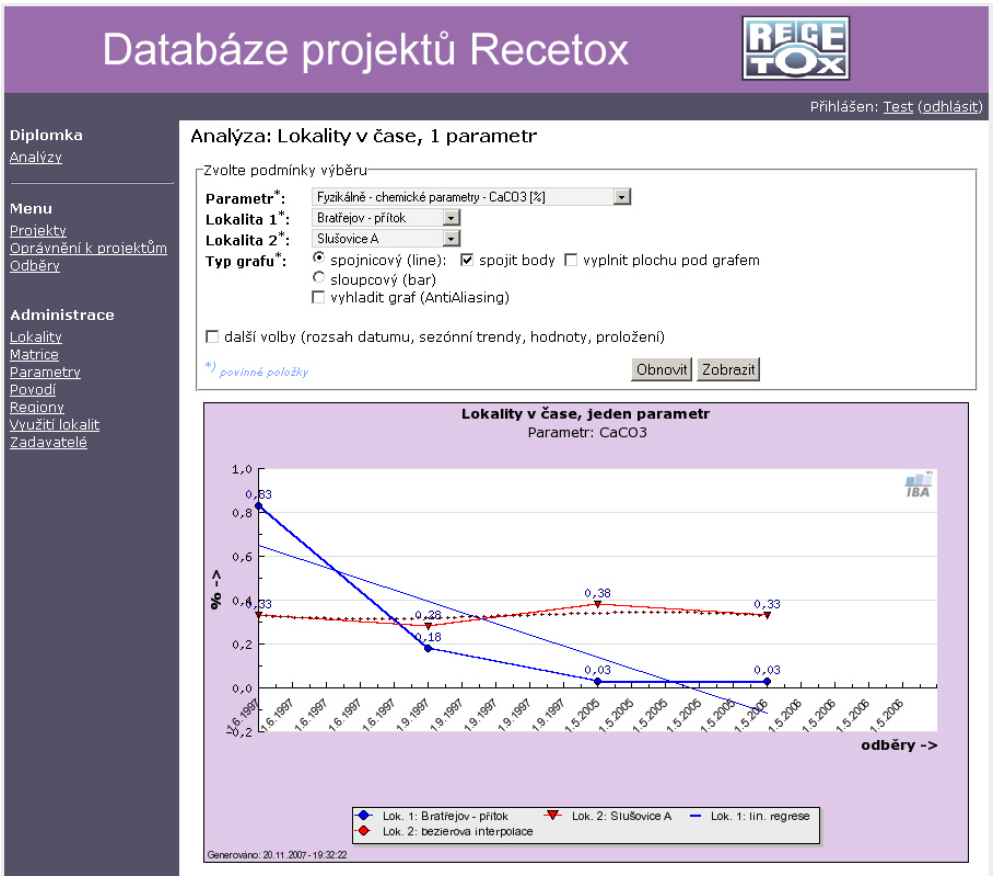


Figure 4. Tool of predefined data analysis (in Czech)

Figure 4 shows the screenshot of the comparison of one physical parameter at two different localities in the selected time interval. Using the predefined forms the user can combine the number of parameters (up to 5), the number of localities (up to 5), the type of the graph and the type of the approximation function.

A large set of the actual system data is targeted to the Zlín region (Czech Republic) or more precisely to the real contamination of the Zlín region. A multimedia sampling of sediment, soil, ambient air, surface water and biota, as the key components of the environmental system, has already been undertaken for many years. The selected priority pollutants from PAHs, PCBs, OCPs, PCDDs/Fs and HMs at chosen sampling sites are measured and various ecotoxicological tests with sampled matrices are being performed. Whereas disaster floods affected this region in July 1997, we have available very valuable data that allows us to study the impact of the floods on contaminant distribution in the environment.

The development of the described information system is continuing. The system prototype is actually running at the server of the Masaryk University but is accessible only for university employers at the address http://projects.cba.muni.cz/inchembiol/.

ACKNOWLEDGEMENTS

This paper was supported by the research plan INCHEMBIOL (no. MSM-0021622412) and GEOKRIMA (no. MSM-0021622418) of the Czech Ministry of Education, Youth and Sports.

REFERENCES

Hřebíček, J., Sluka, J. 2003. Model of the Object-Oriented Environmental Information - A New Approach in the Environmental Data and Information Management. In 17. International Conference Informatics for Environment Protection. The Information Society and Enlargement of the European Union. Marburg : Metropolis Verlag, Germany. 381-388.

Hřebíček, J., Pitner, T., Benko, V. 2003. Standardization of Environmental Data and Information Management in the Czech and Slovak Republics. In 17. International Conference Informatics for Environment Protection. The Information Society and Enlargement of the European Union. Marburg : Metropolis Verlag, Germany. 117-124.

Jarkovský, J., Ráček, J., Némethová, D., Pavliš, P., Brabec, P., Hřebíček, J., Hodovský, J. 2006. Information System of ARROW Project: Assessment of Ecological State of Surface Waters. In Managing Environmental Knowledge, EnviroInfo 2006, 20th International Conference on Informatics for Environmental Protection. Graz, Austria: Shaker Verlag, 353-360.

Ráček, J., Hřebíček, J., Pitner, T. 2006. Process Analysis of Environmental Reporting. In Proceedings of the iEMSs Third Biennial Meeting "Summit on Environmental Modelling and Software". Burlington : International Environmental Modelling and Software Society, 320-324.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Shared Data Model to Support Environment Sensor Network Data in Hydrologic Models

M. Kumar[a], **and C.J. Duffy**[a]

[a] *Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA (muk139/cxd11@psu.edu)*

Abstract: With the advancement of wireless communication and miniaturization of digital electronics, long term observation of remote hydrologic systems using adaptive sensor networks at high spatio-temporal resolutions and across multiple scales, has become a reality. However, for large spatial scales embedded multi-sensor networks with fine temporal sampling rates, the amount and distribution of data generated by these networks becomes unmanageably large. While the sensor network installation itself is generally supported by basic data management software, in the hydrologic sciences there is little support available to directly incorporate the data generated into the hydrologic model. We contend that a seamless transfer of the observed data to the model can be achieved by developing a shared data model which will standardize storage and management of data both at the sensor base station and the hydrologic model. This will lead to enhanced data transfer integrity and will also result in direct input of the sensor network data to the model in realtime without having to go through intermediate pre-processing steps which are error prone. Here we present the shared Data Model structure along with its design considerations in terms of data types, identification of data-classes, relationships and constraints.

Keywords: Sensor Networks; Hydrologic Model; Data Model; CZO-Net.

1. INTRODUCTION

In order to advance our understanding of multiscale coupling of hydrologic processes, new observation systems that capture the spatio-temporal dynamics need to be designed. The motivation is improved predictability of the terrestrial water cycle as well as addressing the problem of closing water, energy and solute budgets. Such an observing system can be expected to perform synergistic measurements of the atmosphere (e.g. water vapor, winds, thermodynamics, cloud-radiative forcing, and precipitation), the near-surface (e.g. surface exchange fluxes of heat, momentum and moisture, including transpiration, along with radiation balances, vegetation dynamics, precipitation, and runoff), and the subsurface (e.g. soil moisture, temperature, pressure profiles, water table and baseflow). The installation of such an observing system requires the explicit intersection of the terrestrial scales associated with hillslopes, watersheds, and river basins, with ecological regions, estuaries, subsurface linkages, and meso-scale weather.

A modern observing platform, comprised of distributed "intelligent" sensors with small, automatic, low-cost, energy-efficient, non-invasive, computationally-capable, and communicative sensor nodes, is able to collect long term data from remote locations at scales and resolutions. These systems are already being used in a variety of environmental monitoring applications [Cerpa et. al., 2001; Cardell et. al., 2005, Hart and Martinez, 2006]. On the one hand, each sensor node and type provides a localized measurement of the hydrologic states (much like observations from data loggers), the network reveals information that is more than sum of its parts, since it is able to measure distributed heterogeneity, localized anisotropy and spatially derived variables (e.g. fluxes) by virtue of its node topology. These networks are flexible and robust due to adaptivity of the individual nodes in their work assignment and communication topology in response to changes in

environment conditions (e.g. events), health of the network (e.g. node failures) and project needs (individual nodes are mobile). The network can be designed in different configurations that recognize the natural landscape boundaries and scales where the atmosphere, vegetation and subsurface partitions interact.

Depending on the node density, coverage area, number of states observed (sensors) at each node, and the sampling rate, the amount of heterogeneous data generated can be very large. Typically, the observed data generated by the sensor network will be used in a numerical model as a parameter, forcing, initial state, or as a validation set. This necessitates intensive data development, organization and topological definition of the multi-sensor data vis-à-vis a hydrologic model discretization grid. A seamless transfer of data directly from the sensor network to the model grid can be achieved by the development of a shared "data model" that will provide a standard structure for storage, sharing and exchange of data independent of the software environment and programming languages [McKinney and Cai, 2002].

In this work we discuss the details of a "shared" data model that can be used to directly assimilate observed data generated by the sensor network with the hydrologic model. To guide this discussion we use a sensor network under development at Shale Hills and Shavers Creek Watershed in central Pennsylvania referred to as CZO_Net. The network employs Crossbow motes® to map hydroclimatic variables such as temperature (atmosphere/ground), relative humidity, incoming solar radiation, barometric pressure, wind speed, wind direction, soil moisture, soil matric potential, and groundwater level. Before discussing the data structure at each nodes and its object oriented classification, the architectural framework of the sensor network is developed.

2. ARCHITECTURAL FRAMEWORK OF CZO-NET

The architecture of the network is three "tiered". At the lowest and least power-intensive layer lie the sensor nodes or "motes". The sensor nodes observe states in its immediate vicinity and communicate the stored data after "limited" signal preprocessing to the neighboring nodes. The nodes are composed of four primary components [Raghunath et. al. 2002] viz. a) a microcomputer that supports a processor, a memory unit and a controller to execute power scheduling and communication protocols, b) a transceiver that is essentially a short range radio to transmit and receive data from/to other nodes and gateways, c) a sensing hardware that is a collection of sensors that measures a set of state variables in the immediate vicinity and d) a power supply which can be a battery or a solar/wind power scavenger. The sensor nodes are often deployed in localized cluster patches which communicate between themselves through a gateway. The gateway periodically downloads the data to the remote base station database server through local area network (LAN) connectivity. The logged information is disseminated over the web from the base station over wide area network (WAN). Figure 1 shows the sensor network framework.

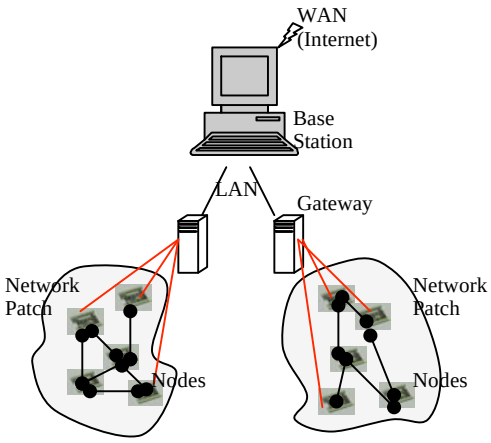


Figure 1: 3-tier sensor network architectural framework

3. DESIGN CONFIGURATIONS

A generic sensor network design is determined by architectural factors such as fault tolerance; scalability; production costs; sensor network topology; hardware constraints; transmission media; and power consumption [Akyildiz et. al., 1999]. For purposes of hydrologic research in addition to the architectural limitations, the network design would be driven by the science goals. We use the watershed as the organizing principal at the regional scale. Physical properties of the watershed (topography, slope, hydrogeologic and landuse/landcover heterogeneity) and hydrologic process dynamics are the basis for sensor deployment. At each node, sensor systems are deployed in 3D domain, which extends from the base of active groundwater circulation, through the soil, vegetation and the top of the atmospheric boundary layer. At any particular nodal location on the land surface, topology of the sensors is designed to capture the direction and magnitude of boundary fluxes across the faces of a 3D control volume. We note that the numerical model also evaluates the same interfacial fluxes by forming semi-discrete balance equations over a unit discretized domain obtained by integration of coupled process differential equations over the projected control volumes. In this way the model serves as both a conceptual tool that explicitly defines the particular interface for which the instrument should measure the flux, and as a constraint on the overall energy/moisture budget itself. The basic instrument configuration is illustrated in Figure 2, and includes a "whole canopy" micrometeorological tower configuration, boundary-layer profilers, as well as surface, soil and groundwater observations.The flexible design is be able to take advantage of the natural scales of motion for water, energy, and should be adaptable to most physiographic and climatic settings.

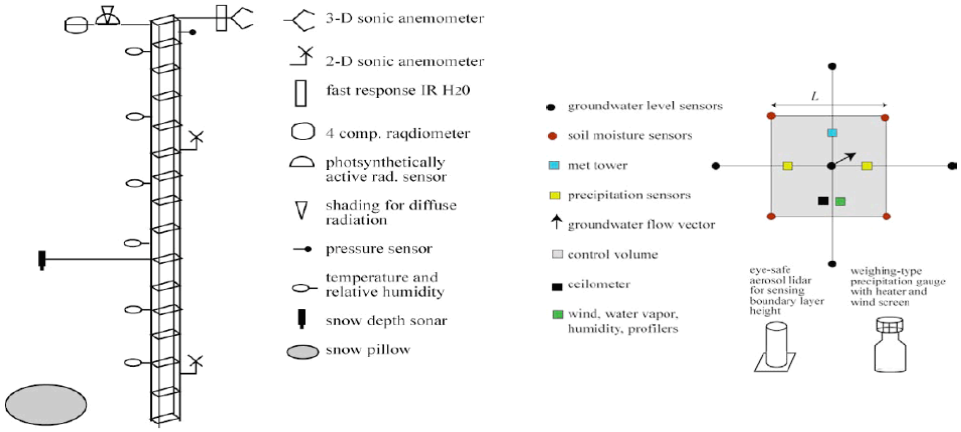


Figure 2: Flux Tower (left) and Subsurface and land-surface instrumentation (right) that constitutes a local sensor array

At the watershed scale, process interactions, hydrogeologic and climatic heterogeneities vary spatially, and thus the placement of sensors would be designed to capture the gradient variability of phenomena of interest.

One approach to an optimal sensor deployment capable of heterogeneous sampling in localized region of the watershed uses Delaunay triangulation. Figure 3 shows three different potential configurations of the sensor network design where sensor node placements reflect a) the boundary between characteristic hydrodynamic descriptors like hypsometry/ vegetation/ soil property, b) a nested local zone of interest, and c) a new state such as a river. One representative example of the latter case is measurement of stream temperature [Troch, 2008] in Valles Caldera, NM. We note that a higher nodal density in any of the shown sensor mesh configurations can be hierarchically obtained by application of "incremental-insertion" algorithm [Lawson, 1977], assuming that the rest of the architectural constraints are satisfied.

We assume that any site chosen would have a completed (or anticipated) digital watershed survey available for soils, geology, vegetation, high resolution topography, in addition to the hydroclimatic database.

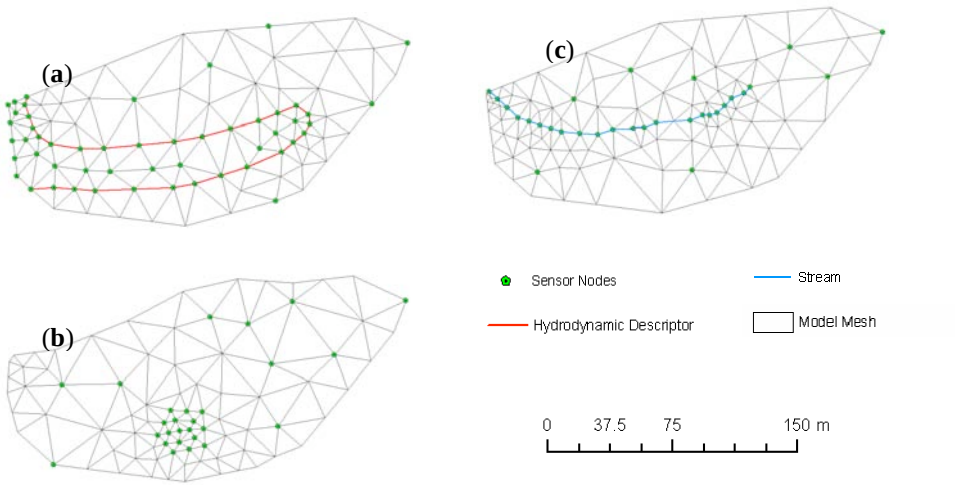


Figure 3: Three different sensor network configurations mapped on Shalehills Watershed

4. SENSOR NETWORK DATA MODEL

The first prerequisite to optimal data model design is accurate assessment of all types of data and data formats. As shown in Figure 2, each sensor node measures a range of of data types including precipitation, net solar radiation, wind speed, relative humidity, air and ground temperature, soil moisture and matric potential. Nonetheless, all the observations are essentially time series. We also note that for some data types, multiple observations are needed at precise separation in order to calculate derived fluxes (such as for ground water head and soil moisture) or for uncertainty estimation (in case of precipitation). The topology of the network is mapped by tracking its neighbors. The sensor network data model is shown in Figure 4. The designed data model follows the standard object oriented representation in UML 2.0. Data types are first organized into different classes. The classes interact with each other through standardized relationship definitions. The advantages of using this strategy are the potential to incrementally enrich the data model, the ability to construct complex objects (extensibility), robustness, and adaptability to changing hydrologic conditions by using different instances of a single object (reusability), and by using the constructs of inheritance, polymorphism and encapsulation [McKinney and Cai, 2002].

Generalization relationship between any two classes means that one of the classes (Child class) is derived from the other (Base class). This relationship is inherent to object-oriented modeling through the “inheritance” mechanism. This relationship markedly simplifies and clarifies the data model and minimizes redundancy in definitions, access and storage. Generalization is denoted by a solid line with a closed arrowhead pointing to the super class. Figure 4 shows that Solar Radiation, Precipitation, Temperature etc. inherit the properties of “Time Series” class.

Association is the most common relationship in a class diagram. Associations can connect classes both in time and in space. They are denoted by an optional arrowhead on one end of the line. An Association linkage without an arrowhead is a bi-directional Association, which means that both of the connecting classes are aware of the relationship. Single ended arrowhead relationships are unidirectional Associations that link the classes in which only one knows about the relationship. The class from which the arrow invocation emerges is the class which has knowledge of the relationship. One other type of association that has been implemented in the developed data model is Reflexive association. This

linkage represents the association of the class to itself. This essentially means that another instance of class is associated with the present one. We note from Figure 4 that each Time Series class is associated to a "Sensor Node" class which is essentially the location at which it is observed.

Aggregation relationships explain the interaction of individual parts/components (Simple Objects) to a Complex Object. The relationship is denoted by a white diamond (for the Aggregate class) on one end of the link and arrow (for the "part" class) on the other . Sensor Node Aggregate to form Sensor Patch.

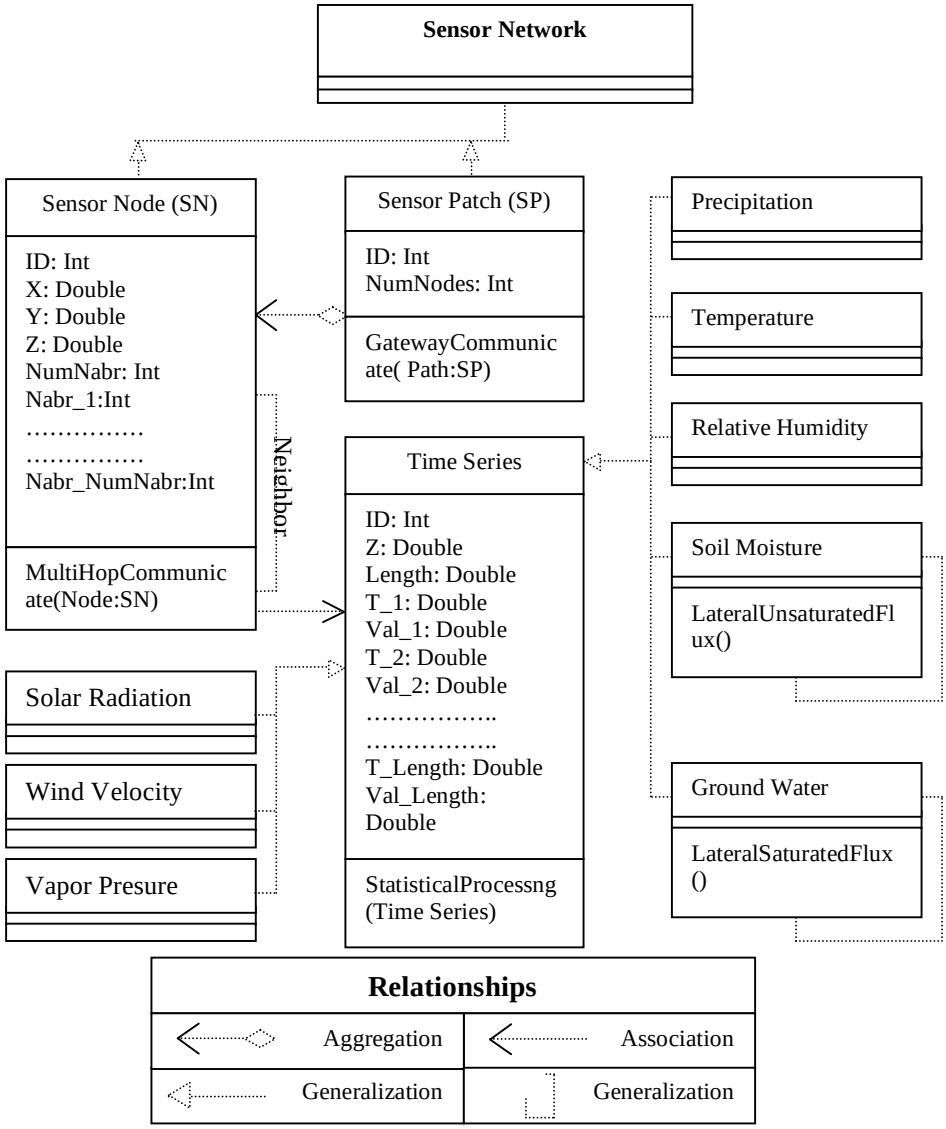


Fig. 4: Sensor Network Data Model in UML 2.0. Note that all the hydroclimatic data measured at the sensor node is "associated" to it. The operators in the bottom compartment for each individual class are basic schema processing that are either possible at the node itself or at the base station.

The operation that is carried out on each Class Object is shown in the lowest compartment. Operations such as SignalProcessing() on each Time Series is carried out before the sensor nodes use MultiHopCommunication() protocols to direct it through the gateway to Base

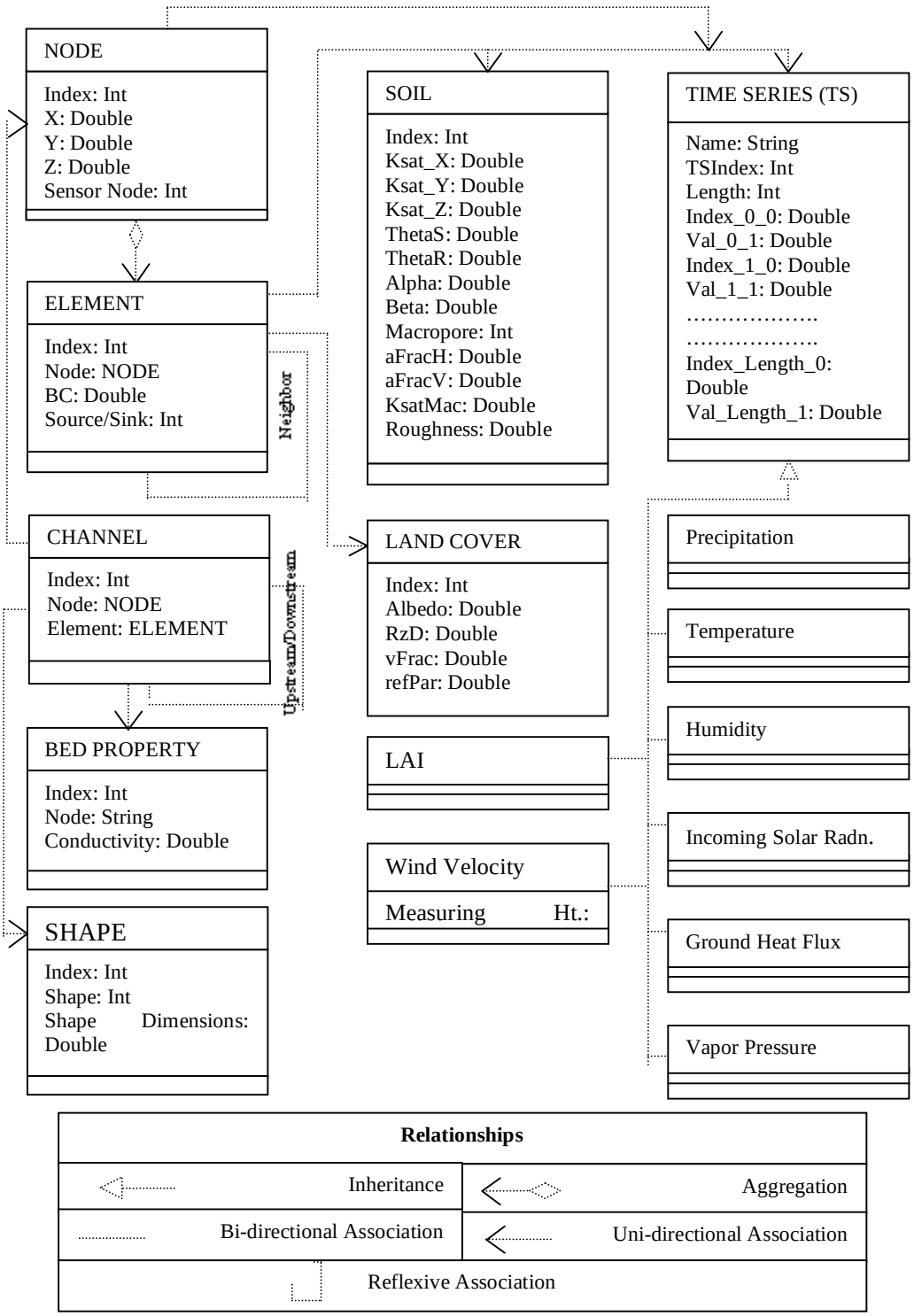


Figure 5: Shared Data Model for sensor network and hydrologic model. All the sensor node attributes and relationships are assigned to mesh nodes.

Station. In order for sensor network data to be used seamlessly in hydrologic modeling, the sensor data model constructs –classes and relationships- need to be supported in the hydrologic model data structure. By generating mesh decomposition using points and lines as constraints (shown in Figure 3, more details in Kumar et. al. [2008]), nodes and edges of the triangles in unstructured mesh decomposition of the model domain automatically

represents the sensor nodes and its neighbors respectively (shown in Figure 3). This means that the relationships and classes corresponding to each sensor node can be directly transferred to the data structure, relational attributes and topology information associated with discretized unstructured domain nodes.

5. SHARED "HYDROLOGIC MODEL"–"SENSOR NETWORK" DATA MODEL

The developed data model is shown in Figure 5. The classes identified to describe the hydrologic system and processes are: Node, Element, Channel, Soil and Time Series. Each node is uniquely identified by its coordinate location and a sensor node ID if it exists. The data model supports Aggregation, Uni-directional Association, Reflexive Association and Generalization relationships between the objects. An Element class represents a discretized triangular element in 2D and a prismatic element in 3D and is defined by six nodal locations listed in clockwise direction at two levels. The prismatic element has five neighbours- three on the sides and one at the top and bottom. We note that neighbours of an element also belong to an Element class and this recursive relationship is captured by Reflexive association. A Channel class is defined by the two end nodes and neighbouring elements on the either side of channel. Each channel segment is also composed of an upstream and downstream channel segment which is captured by a Reflexive association. Channel is also Bi-directionally associated to each Element. Bi-directionality ensures that both Element and Channel is aware of this topological relationship. These relations are fundamentally important for spatial integrity of the hydrologic modeling framework. Each Element class is also associated with Soil class and Time Series. This ensures proper, clean and efficient assignment of properties to each Element. Similarly Channel is associated to Bed Property and Shape classes. Soil Class contains several attribute fields such as for Hydraulic conductivities and van-Genuchten equation parameters. We note that Precipitation, Temperature, Humidity, Incoming Solar Radiation, Ground Heat Flux, Vapour Pressure, LAI, Vegetation Fraction, Wind Velocity, Time dependent boundary conditions and the observed and simulated state variables are all instances or child objects to the Time Series Class. We note that the shared data model (shown in Figure 5) supports all the data types and the relationships that are used to store sensor the sensor network data (shown in Figure 4).

6. CONCLUSIONS

This paper presents the design and details of a shared data model which supports coupling of sensor network data and a hydrologic model. The data model incorporates representation of a wide range of data types, feature objects and relationship between classes. The data model is rich yet flexible in terms of its extensibility and simplicity. The conceptualization and characterization of this coupling strategy can be used with other physically distributed models and can well be extended to management, visualization and decision support tools.

REFERENCES

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., Wireless sensor networks: a survey ,Computer Networks, Amsterdam, Netherlands, 1999

Cardell, R.,Smettem, K.,Kranz, M., and Mayer, K., A Reactive Soil Moisture Sensor Network: Design and Field Evaluation, International Journal of Distributed Sensor Networks, pp. 149 – 162,1,2, 2005

Cerpa,A., Elson,J., Estrin,D., and Girod, L., Habitat Monitoring: Application Driver for Wireless Communications Technology. In Proceedings of the ACM SIGCOMMWorkshop on Data Communications in Latin America and the Carribean, pages 20–41, 2001

Hart, J.K. and Martinez, K,, Environmental Sensor Networks: A revolution in the earth system science?, Earth-Science Reviews 78, 177–191, 2006

Kumar, Bhatt, Duffy, An efficient domain decomposition framework for accurate representation of geodata in distributed hydrologic models, International Journal of GIS, 2008 (submitted).

Lawson. C. L., Software for C1 Surface Interpolation, Mathematical Software III (John R. Rice, editor), pages 161-194. Academic Press, New York, 1977.

McKinney, D.C., and Cai,X.,Linking GIS and water resources management models: an object-oriented method. Environmental Modeling and Software. v17, 5, 413-425. 2002

Raghunathan, V., Schurgers, C., Park. S, and Srivastava, M.B., Energy-aware wireless microsensor networks," IEEE Signal Processing Magazine, 19, 2 , 40-50, March 2002

Troch, 2008, http://www.hwr.arizona.edu/~surface/fieldsites/valles.html

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Primary Environmental Data Quality Model: Proposal of a Prototype of Model Concept

M. Hejč[a] and J. Hřebíček[b]

[a] *Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic (hejc@ecomanag.cz)*

[b] *Institute of Biostatistics and Analyses, Masaryk University, Kamenice 126/3, 625 00 Brno, Czech Republic (hrebicek@iba.muni.cz)*

Abstract: There is the short overview of the terms data quality and environmental information, following up the model definition in the paper. It introduces one such prototype of model – easy to implement, covering all modelling niches in environmental informatics and promising easy model knowledge sharing – it will be suitable to make a basis for appropriate model library. There is also discussed the use of the model and its role in the process of measurement of data quality. The concept is illustrated by the case study of the South Moravian region waste management data evaluation (realized by authors) and compared with the approach of the Ministry of Environment of the Czech Republic. Short conclusion suggests the future exploitation of possible new ways of dealing with the primary data uncertainty.

Keywords: Environmental Data; Data Quality; Data Uncertainty; Data Quality Model; Waste Management; Data Validation.

1. INTRODUCTION

Primary environmental data are monitored and collected by different ways: technically (e.g. using sensors, monitoring devices, people, etc) and by organizations Eurostat, European Environment Agency (EEA) and U.S. Environmental Protection Agency (EPA). The quality of the data is also variable (depending on many factors) or the data can be incomplete [Eurostat1, 2002]. These primary data are processed and they form required environmental information. If we want to determine the reliability of such information, its necessary to measure the quality of the primary data and even make changes to them (add, change or delete values with poor quality).

Measurement of the quality of the primary data can be made in various manners [Eurostat1, 2002], Hejč et al. [2007], and Pipino et al. [2002]. Very often it is not made or used at all. Sometimes it is (or it can be) judged by more or less experienced administration authority, but this evaluation process costs time and money or other resources. Therefore, it is not suitable for mass data analysis – this is the reason, why it's not often accomplished. Another reason lies in the lack of proper knowledge about the primary data monitoring and collecting system. It is necessary to use some techniques, which will be suitable for automatic computer processing.

This paper proposes such new technique – the new model for describing and managing environmental data quality. It will allow better results of waste management evaluation done by national, regional and local governments in the Czech Republic.

2. ENVIRONMENTAL INFORMATION

A short excursion into the field of the term "Environmental Information" gives us the EEA definition: "Knowledge communicated or received concerning any aspect of the ecosystem, the natural resources within it or, more generally, the external factors surrounding and affecting human life."

The definition of environmental information is very broad and includes these types of information: the ***state of elements*** of the environment – such air, water, soil, land, landscape and natural sites, flora and fauna, including cattle, crops, genetically modified organisms, wildlife and biological diversity – and it includes any interaction between them; the ***state of human health and safety***, conditions of human life, the food chain, cultural sites and built structures, which are, or likely to be affected by the state of the elements of the environment and the interaction between them; any ***factor*** such as substances, energy, noise, radiation or waste, including radioactive waste, emissions, discharges and other releases affecting or likely to affect the state of the elements of environment or any interaction between them; ***measures*** and activities affecting or likely to affect, or intended to protect the state of the elements of the environment and the interaction between them. This includes administrative measures, policies, legislation, plans, programs and environmental agreements; ***emissions***, discharges and other releases into the environment; ***cost benefit*** and other economic analysis used in environmental decision making [EC2003, 2003].

Pick [2007] distinguished only state of environmental elements, factors, measures and effects.

Another current research defines Single Information Space for the Environment in Europe (SISE) specified in the Work Programme for ICT research in FP7 for 2007 and 2008 as the common platform of all kinds of environmental information [Schouppe, 2008]. This is also the common research topic of the research group, in which are the authors of this paper [Nagy, Legat, and Hrebicek, 2007].

2.1 Non-environmental Data

During the process of environmental information evaluation there are often used some non-environmental information (e.g. subject addresses, names and other mostly personal, society or business data) [Eurostat1, 2002]. Actors and their descriptive data are also playing important role in the process of environmental information evaluation.

2.2 Uncertainties

Uncertainties in the scientific sense are the component of all aspects of the environmental modelling process. They describe lack of knowledge about models, their parameters, constants, data, information and beliefs [Jolma and Bortin, 2005].

Data (or information) quality is the measure of the data (or information), which measures uncertainties. The quality of data (or information) is high when the present uncertainties are low and vice-versa.

We will not cover any of these terms in to much detail in this paper, as the detailed description can be found in [Hejč and Hřebíček, 2006], Hřebíček et al. [2006] and [Olson, 2003].

3. DATA QUALITY MODEL

The primary environmental data values are often simulated by a model, but there is no well known and respected standard of the model. Most of experts use their own models and their own concept of the data quality model Hřebíček et al. [2006], [Olson, 2003], Pipino et al. [2002].

The new prepared standard ISO 25012 is an attempt to give the framework of primary environmental data quality model, but it is not still accepted by all ISO countries. The purpose of this standard is to prompt creators of large and small scale databases of primary environmental data to observe predefined criteria which will enable them to evaluate and test the quality of data, set up integrated and interoperable databases, reduce ambiguity, avoid redundancy, promote ease of data maintenance and promote reliable, secure databases. We have taken into account this standard in the development of our model.

3.1 New Definitions

We have to define some new terms and assumptions, which we will use in the supposed model concept implementation. These additional definitions and assumptions form the model concept. The *primary data set* (database) is the real input of the model. This data can be characterized by the structure of data model (e.g. ERD, etc.). This data model contains all the dependencies of the attributes and tables. One "piece" of the *data set*, (we will call it an *item*), is formed by the attribute name, its value and its relation to the rest of database.

It is very important to distinguish two kinds of model. The *data quality model* is the main model and (as its name implies) it is used to enumerate the data quality. The *data value model* (as illustrated in Hejč et al. [2007]) is the model, used to enumerate the value of the data, when the value is not known from some reasons.

The *enhanced primary data set* is the *data set*, which contains *tags* for every its *item*. The proposed model makes the use of two new *tags*: the *first tag* represents the *probability of item value to be true* and the *second tag* the *probability of the item value to be useful*. These are two different *tags*, because the value of the data can be false, but we know it's close to the true value and on the opposite hand we can have the true value which is from some reason completely useless for us.

The *item* can be accompanied by more *primary tags*, because it has more interesting characteristics from the standard ISO 25012 (accessibility, accuracy, currency, completeness, credibility, etc) and even other characteristics, not fully covered in ISO 25012. The proposed model uses them (and stores them if they are available) but not for the reason of the final evaluation of the data. This decision is motivated by practice. These *primary tags* are mapped into 2 above-mentioned new *tags*.

The most important characteristics of environmental data are: time and the field of environment (elements, factors, etc.) being described. It is very difficult to identify the most important *primary tags* as the process of mapping of these *primary tags* into *new tags* is different for each evaluation procedure. Generally, the *anomaly* can be used, when the data are right, but from some reason they are useless (e.g. the production of waste in municipality affected by extraordinary flood is not suitable for the computation of the average waste production in municipalities). The standard ISO 25012 credibility and also the difference from *data value model* can be used as main *primary tags* to evaluate the data with high influence of uncertainty, which has been raised from the human factor (or similar factors, e.g. complete fail of measuring device, etc.), Hejč et al. [2007].

Table 1. Data quality model.

Enhanced Primary Data Set (for the given purpose)						
Primary data	Tags		Primary tags			DVM
	P(TRUE)	P(USEFUL)	*Anomaly*	*Credibility*	…	
Item 1 (key, attr., val.)	t1	u1	a1	c1	…	x1
Item 2 (key, attr., val.)	t2	u2	a2	c2	…	x2
Item 3 (key, attr., val.)	…	…	…	…	…	…
…	…	…	…	…	…	…

The whole concept and new terms are illustrated by the Table 1, where every item consists from the key (the relation to the rest of data in the given data model), the attribute name and the attribute value.

The tag P(TRUE) means the probability of item to be true, P(USEFUL) the probability of item to be useful and DVM means the data value model in Table 1. There are illustrated only few primary tags, but it is possible to present more of them.

The purpose of the enhanced primary data set is given by the evaluated case and reflected in different values of tags P(USEFUL).

3.2 Mapping

Mapping of *primary tags* into two *new tags* during the evaluation procedure is the key part of the *data quality model.*

We suppose to react only on some kinds of data uncertainty (as mentioned above). There are many sources of uncertainty, including: uncertainty in scientific constants, observation error, implementation uncertainty, etc., see Hřebíček et al [2006], but we suppose they can be solved separately by other models or by EEA, Eurostat or EPA procedures and they can be later incorporated into new model (or vice-versa).

Our model is defined as a function of several parameters. Often a very complex function (with a lot of exclusions), but not always – sometimes can be simple. Function value represents *new tag's* value. Input parameters of the function include various knowledge about the item (*primary tags*), the value of the item itself and the value of the *data value model*. Different (in the sense of *primary tags* used) *data quality models* can be easily combined as functions do the same.

When we get the data, we have to fill, look for or compute the values of all *primary tags*. It can be done very simply (by setting some default values) or by application of some rules (e.g. all the data from some sources are more suspicious of being wrong – that means setting their credibility lower then the others). Application of rules may be cumulative and this implies the need for some arithmetic to compute the final value of the *primary tags*.

The last application of the rules would be the comparison with the *data value model*. If the value of the item is not far from the value suggested by the *data value model*, the probability of the item value to be true is high, similar rules apply for usefulness.

Finally the mapping of *primary tags* into *new tags* (probability and usefulness) is done for all items and for given purpose (type of evaluation of the data). This is the new approach.

In rare cases we can employ some optimization function which recognizes the information quality by some independent (this means different than the application of the above mentioned *data quality model*) method. This gives us the possibility of feedback for the correctness of the *data quality model* (and thus possibility of automatic model shaping through mapping changes). The only way to demonstrate the correctness of the *data quality model* is in the other cases the independent study, made by some expert in the field.

3.3 Data Changes

We will get the data with some new attributes and we can decide what to do further – we can define some rules. Either we can replace item values with low probability by the *data value model* values or we can exploit some values with high usefulness.

Other possibilities lie in the filling of the gaps of data. First we have to identify them (by the *data value model*) and then we need only to deliver appropriate *data value model* values into data set. Again the set of rules would be useful in the process of data quality evaluation.

Finally we have the data set with some quality evaluation and enhancement and we can use it for information retrieval. In the same time we are aware of the information quality, as it is tightly bound to the data quality evaluated before.

4. USE OF THE MODEL

The use of new *data quality model* has been tested during annual evaluation of waste management indicators in the South Moravia Region since 2004. The evaluation of waste management indicators in the Czech Republic is usually done by the different approach of the Ministry of Environment (MoE), Hejč et al. [2007]. So-called null-variant approach is used in some other regions than South Moravian. This null-variant means simply the direct evaluation of primary data without any pre-processing treatment. Other types of approaches (hypothetical full variant by EPA and the compromise approach of statistic offices), which are described in [Eurostat1, 2002], Hejč et al. [2007], will not be compared with the above approaches.

4.1 Comparison

The differences between MoE and our approach are shown by the specific example. We choose as the example an evaluation of the household waste production at municipalities. In Visegrad countries, the household waste is collected and separated into waste containers depending on the collection system of the given municipality. The amount of the household waste production and disposal of the given municipality is announced / reported in compliance with the national legislation of the Czech Republic to MoE through local state administration bodies and the Centre of Waste Management (http://ceho.vuv.cz/). All available annual reports are evaluated and the overall production of municipal household waste is aggregated into the final environmental reports of the Czech Republic to EEA and Eurostat.

There is the common part for both approaches – the primary data about the municipal household waste production are collected and evaluated. However, there are differences in the types of data collection and their processing. When the null-variant comes into play, annual reports of municipalities are just collected and the plain summary is processed and evaluated. Sometimes some most flashy cases of errors are filtered (by means of interval arithmetic). The approach of MoE is closer to null-variant than to any of the others. Interval arithmetic is the only strong tool. A lot of knowledge is not used in the evaluation process.

But we know more about the nature of these data. The municipal household waste production strongly depends on the number of inhabitants and the standard of living. Then there are some other dependencies on the size of the community, the type of housing, unemployment rate, time series of waste production, etc. All these dependencies can be incorporated into the *data value model* of these primary data. Such model forms the knowledge and can be used for the verification of the data or to replace the gaps of the data (as statistics approach does).

We describe formally a simple model of waste production as the function of appropriate variables and bellow is presented as an example of *data value model*:

$$P = F(\#inh, spec, std, sz, unemp, hsg, heat),$$

where are defined: P is the amount of the waste production per year; *#inh* is the number of inhabitants; *spec* is the specific waste production coefficient (reference values of other coefficients), measured in kg; *std* is the standard of living coefficient; *sz* is the size of the community coefficient; *unemp* is the unemployment rate coefficient; *hsg* is the type of housing (recreation, blocks of flats, empty houses…) coefficient and *heat* is the type of heating coefficient.

In this case the function *F(#inh, spec, std, sz, unemp, hsg, heat)* can be defined, see Fig. 1, and we can write:

$$P = \#inh \cdot spec \cdot std \cdot sz \cdot unemp \cdot hsg \cdot heat / 1000 \text{ [t]}.$$

Further, we can compute some coefficients x of function F, x belongs to {*#inh*, *spec*, *std*, *sz*, *unemp*, *hsg*, *heat*}, by the expression:

$$x = (act / ref)^{cx},$$

where *ref* means a reference value; *act* an actual value and *cx* is the compensator (given by optimization process) of the considered coefficient *x*.

The model of the standard of living value (as one example of numerous sub-models) is used to compute the actual and the reference value of the considered coefficient:

$$stdV = Rinc \cdot Rsz,$$

where *stdV* is the standard of living value; *Rinc* the average income in the given region and *Rsz* the size of the community in region coefficient.

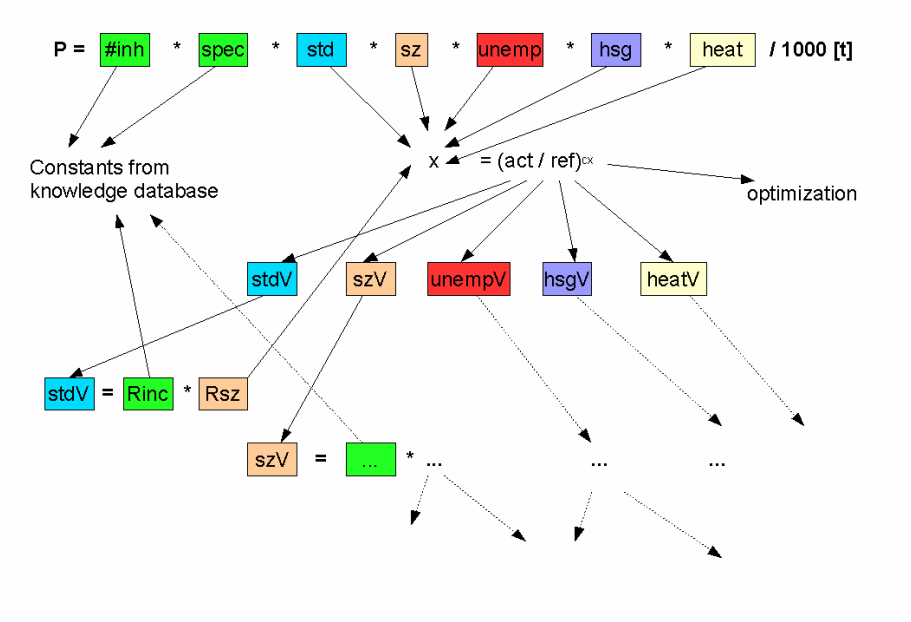


Figure 1. Model of waste production in communities (Hejč and Hřebíček, 2007).

Different colours of cells of Figure 1 are used to distinguish main areas of interest – number of inhabitants and specific waste production (green), standard of living (light blue), size of the municipality (orange), unemployment (red), housing (purple) and heating (yellow).

4.2 Details of Use

All data are enhanced by all available *primary tags*. For example: the data source district credibility; the data source subject credibility (both taken from time series); the value from *data value model*; cross-reference (by reason that subject and its partner are always reported), etc.

Other (non-environmental) data are collected (mostly for the purpose of the *data value model*) and they are enhanced by appropriate *primary tags*. For example demographic data, addresses, economic data, etc.

The appropriate mapping of *primary tags* on 2 *new tags* (see above) is used (assigned) for each type of evaluation. This mapping is fine-tuned by an optimization process in some

rare cases. The example of possible case is the overall waste production when the comparison of waste production volume and waste treatment volume is available.

The second phase of data processing uses only the primary data (without *primary tags*) and 2 *new tags* mapped by appropriate mapping which conforms to the purpose of evaluation (a specific indicator). This approach makes possible to attain better results in the process of final evaluation of the indicator. It also gives possible records (warning) for stakeholders – they are warned of problems in the data by easily understandable way.

4.3 Practical Experiences

The practical experiences with the use of the *data quality model* have been already very promising. Authors used the presented concept of the *data quality model* for the evaluation of the waste production data of the South Moravian Region in the Czech Republic. It was used in the years 2004, 2005 and 2006. Some basic experiences have been obtain also from the evaluations of the waste production data in the years 1999, 2000, 2001 and 2002, but these evaluations have not used the later proposed the *data quality model.* However, the experiences from these first years have been very useful to develop it. Table 2 illustrates some interesting statistics from the processes of evaluation of the waste production data in the years 2004, 2005 and 2006. The data of the year 2007 will be evaluated in the summer of the year 2008.

Table 2. Statistics from the processes of evaluation of the waste production data.

	Database			**Errors**				
Year	Items	Plants	Subjects	found	suspected	Hit rate	estimated	Hit rate
2004	145 068	22 428	16 783	75	228	33%	1015	7%
2005	166 501	28 815	21 413	63	130	48%	749	8%
2006	176 676	31 439	22 551	44	429	10%	530	8%

It is clear from Table 2, that the amount of data in databases is growing and the number of estimated error is decreasing. The lower hit rate of the number of found errors vs. the number of suspected errors in 2006 is due the short time for the confirmation of suspects, while the same hit rate in 2005 is higher due the short time for the preparation of suspects (with the strategy of finding only flashy ones). The 2007 year promises a good increase of hit rate of found errors vs. estimated errors, because there will be devoted more time by the local government for the whole evaluation process of the waste production data.

5. CONCLUSIONS

We have presented short overview of the terms in data quality area and we also enhanced their capabilities by defining some new ones. We defined the new model of data quality and introduced it on the example of the waste production data of the South Moravian region of the Czech Republic. The main advantage of the new model prototype lies in the representation of the data and easy implementation and sharing of the modelling results. The further property of the model is its ability to locate the data uncertainty when any other ways of uncertainty measurement are not present [Hejč and Hřebíček, 2006]. Future research will be trend towards refining of the model and also incorporating it in the framework of broader research interest of authors – environmental information space [Schouppe, 2008].

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Education of the Czech Republic for the financial support of the project No MSM 0021622412 INCHEMBIOL and the Ministry of

Environment of the Czech Republic for the financial support of the project No SPII2/f1/30/07.

REFERENCES

EC2003, Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information. Official Journal of the European Union. L 41/26. 2003.

EEA, European Environment Agency, <http://www.eea.europa.eu/>, 2008/04/30

EPA, U.S. Environmental Protection Agency, http://www.epa.gov/>, 2008/04/30

Eurostat, <http://ec.europa.eu/eurostat>, 2008/04/30

Eurostat1, Quality in the European statistical system – the way forward. 2002 Edition. Office for Official Publications of the European Communities, Luxembourg, 2002.

Hejč, M., and Hřebíček, J., Solving Waste Management Data Uncertainties. Case Study of South Moravian Region. 20th International Conference on Informatics for Environmental Protection. Managing Environmental Knowledge. Shaker Verlag, Aachen, 405-409, 2006.

Hejč, M., Hlaváček, M., and Hřebíček, J., eGovernment Services in Environment - Automate Data Quality Assessment - Czech Republic Approach, In EnviroInfo 2007. 21st International Conference on Informatics for Environmental Protection. Environmental Informatics and System Research. Volume 2. Workshop and application papers. Warsaw, Poland: Shaker Verlag, Aachen, 159-166, 2007.

Hřebíček, J., Hejč, M., and Holoubek, I., Current Trends in Environmental Modelling with Uncertainty, Proceedings of the iEMSs Third Biennial Meeting "Summit on Environmental Modelling&Software". International Environmental Modelling and Software Society, Burlington, 342-347, 2006.

Jolma A., and J., Bortin, Methods of uncertainty treatment in environmental models, *Environmental Modelling & Software*, 20(8), 979-980, 2005.

Nagy, M., Legat, R., and Hrebicek, J., Electronic Access to Environmental Information – an Important Fundament for E-Democracy and Environmental Protection, CAHDE (2007) 23 E. <http://www.coe.int/t/e/integrated_projects/democracy/02_activities/002_e-democracy/>, 2008/04/30

Olson, J. E., Data Quality: the accuracy dimension. Morgan Kaufmann Publishers, San Francisco. 2003.

Pick, T., From Aarhus to INSPIRE: Putting Environmental Information on the Map. In EnviroInfo 2007. 21st International Conference on Informatics for Environmental Protection. Environmental Informatics and System Research. Volume 2. Workshop and application papers. Warsaw, Poland: Shaker Verlag, 239-246, 2007.

Pipino, L., Lee, Y., and Wang, R., Data Quality Assessment. *Communications of the ACM,* 45(4), 2002.

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Implicit Large Eddy Simulation - a Promising Method for Turbulence Modelling in High Resolution Models

V. Fuka [a], **J. Brechler**[a]

[a]*Dept. of Meteorology and Env. Protection, Charles University, V Holešovičkách 2, Prague 8, 180 00, Czech Republic (vladimir.fuka@gmail.com, josef.brechler@mff.cuni.cz)*

Abstract: Implicit Large Eddy Simulation is an emerging tool for turbulence modelling. Instead of explicitely computing subgrid stress models, numerical dissipation of nonlinear schemes can be used as a main control mechanism of turbulent energy transfer. In this contribution we test performance of a high-resolution incompressible projection method in modelling of the Taylor-Green vortex flow, which stands as a prototype of a simple free flow with a transition to turbulence and a turbulence decay. Several criteria are used to asses the model, including the kinetic energy dissipation rate, kinetic energy spectra and probability density functions of velocity gradients.

Keywords: large eddy simulation; turbulence decay; turbulence simulation

1 INTRODUCTION

At the present time a new field of turbulence modelling called Implicit Large Eddy Simulation (ILES) (for a comprehensive review see Grinstein et al. [2007b]) enables turbulent flow computations with methods most commonly used for compressible fluid dynamics with shocks. These so called shock-capturing schemes can describe flows with very strong gradients using convection terms with nonlinear numerical diffusion. This numerical dissipation is stronger in areas with larger gradients and smaller in smooth parts. Nonlinearity can be achieved using explicit numerical diffusion terms (similar to traditional subgrid stress models) like Jameson et al. [1981] or implicitly using flux or slope limiters [Harten, 1983; van Leer, 1979]. Especially the second way proved to be efficient way of simulating compressible turbulent flows. Because of monotonicity preserving of these schemes, this kind of LES is also known as MILES - monotonically integrated large eddy simulation). To the most used methods belongs flux-corrected transport (FCT) [Boris and Book, 1997], PPM [Colella and Woodward, 1984] or MPDATA [Smolarkiewicz and Margolin, 1998]. MPDATA is used mainly for geophysical applications, which is also aim of our work.

Most of present ILES methods were developed in the framework of compressible flow solvers. Our aim is to develop and test similar method in traditional incompressible code. Schemes for incompressible flows mostly employ artificial compressibility method or projection methods. In artificial compressibility methods schemes for compressible fluid dynamics can be used in a straightforward way.

Projection methods, also known as fractional step methods [Brown et al., 2001], can be divided into two groups – exact projection methods and approximate projection methods. Approximate projection methods mostly employ cell centered grids and enable easier usage of high resolution methods [Almgren et al., 1996, 1998] at the cost of more difficult treatment of velocity – pressure coupling. Exact projection methods [Kim and Moin, 1985] provide very velocity – pressure coupling on staggered grids, but the usage of high resolution methods is complicated.

First example of an exact projection method on a staggered grid with high resolution advective scheme was a method by Tau [1994]. This scheme used a projection method of Bell et al. [1989] with Godunov method for advective fluxes modified for a staggered grid in 2D. This unique combination was emphasised by Rider [1998]. We choose this method as a base for our 3D model for incompressible implicit large eddy simulation.

As a first test case we computed Taylor-Green vortex flow, which serves as a simple example of free flow with transition to turbulence and subsequent turbulent decay. This flow was used for this purpose by several other authors, for example [Bensow et al., 2007; Garnier et al., 1999; Grinstein et al., 2007a]. The comparison is usually done with direct numerical simulation (DNS) data of Brachet et al. [1983] and Brachet [1991].

2 NUMERICAL METHOD

2.1 Projection methods

Our code solves the incompressible Navier-Stokes equations in the form

$$\frac{\partial u}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p + \nu\Delta\boldsymbol{u}, \tag{1}$$

$$\nabla\cdot\boldsymbol{u} = 0, \tag{2}$$

where $\boldsymbol{u} = (u, v, w)$ is the velocity vector, p is the pressure and ν is the coefficient of the kinematic viscosity.

The core of numerical method we used is a second order accurate projection method by Bell et al. [1989], adapted to 2D staggered grids by Tau [1994]. The main stages of one time step can be written as

$$\frac{\boldsymbol{u}^*_{i,j,k} - \boldsymbol{u}^n_{i,j,k}}{\Delta t} + \nabla_{i,j,k}p^{n-\frac{1}{2}} = -\left[(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}^n\right]^{n+\frac{1}{2}}_{i,j,k} + \frac{\nu}{2}\left(\Delta_{i,j,k}\boldsymbol{u}^n + \Delta_{i,j,k}\boldsymbol{u}^*\right), \tag{3}$$

$$\tilde{\boldsymbol{u}}_{i,j,k} = \boldsymbol{u}^*_{i,j,k} + \nabla_{i,j,k}p^{n-\frac{1}{2}}, \tag{4}$$

$$\Delta_{i,j,k}\varphi = (\nabla\cdot)_{i,j,k}\tilde{\boldsymbol{u}}, \tag{5}$$

$$\boldsymbol{u}^{n+1}_{i,j,k} = \tilde{\boldsymbol{u}}_{i,j,k} - \Delta t\nabla_{i,j,k}\varphi, \tag{6}$$

$$p^{n+\frac{1}{2}}_{i,j,k} = \varphi_{i,j,k}, \tag{7}$$

where $\boldsymbol{u}^*$ does not obey the continuity equation and in the next stage is projected onto a solenoidal field using the so called pressure form of the exact projection. The diffusion terms are treated by means of the implicit Crank-Nicolson method in the equation (3).

2.2 High-resolution advection method

The advection terms $[(u\cdot\nabla)\boldsymbol{u}]^{n+\frac{1}{2}}_{i,j,k}$ have to be second order in time, and therefore the Taylor expansion in space and time is used. Derivatives in time are replaced by derivatives in space using the equation (3). Because of the staggered grid, we have to use different treatment of spacial discretization of advective terms. We define time centered values of velocity in cell centers (with integer indexes) and cell corners (with half indexes), so that

$$u^{n+\frac{1}{2},\mathrm{E}}_{i,j,k} = u^n_{i+\frac{1}{2},j,k} - \frac{\Delta x}{2}(u^n_x)_{i+\frac{1}{2},j,k} + \frac{\Delta t}{2}(u^n_t)_{i+\frac{1}{2},j,k}, \tag{8}$$

$$u^{n+\frac{1}{2},\mathrm{W}}_{i,j,k} = u^n_{i-\frac{1}{2},j,k} + \frac{\Delta x}{2}(u^n_x)_{i-\frac{1}{2},j,k} + \frac{\Delta t}{2}(u^n_t)_{i-\frac{1}{2},j,k}, \tag{9}$$

$$u^{n+\frac{1}{2},\mathrm{SB}}_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} = u^n_{i+\frac{1}{2},j,k} + \frac{\Delta y}{2}(u^n_y)_{i+\frac{1}{2},j,k} + \frac{\Delta z}{2}(u^n_z)_{i+\frac{1}{2},j,k} + \frac{\Delta t}{2}(u^n_t)_{i+\frac{1}{2},j,k}, \quad (10)$$

$$u^{n+\frac{1}{2},\mathrm{NT}}_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} = u^n_{i+\frac{1}{2},j+1,k+1} - \frac{\Delta y}{2}(u^n_y)_{i+\frac{1}{2},j+1,k+1} - \frac{\Delta z}{2}(u^n_z)_{i+\frac{1}{2},j+1,k+1} + \\ + \frac{\Delta t}{2}(u^n_t)_{i+\frac{1}{2},j+1,k+1}, \quad (11)$$

where W,E,S,N,B and T denote directions $-x$, $+x$, $-y$, $+y$, $-z$ and $+z$ respectively.

The spatial derivatives in eq. (8-11) are obtained using piecewise linear reconstruction in computational cells. Computed derivatives are then limited using a slope limiter

$$(u^n_x)^{\mathrm{L}}_{i+\frac{1}{2},j,k} = \frac{u^n_{i+\frac{1}{2}} - u^n_{i-\frac{1}{2}}}{\Delta x}, \quad (12)$$

$$(u^n_x)^{\mathrm{R}}_{i+\frac{1}{2},j,k} = \frac{u^n_{i+3/2} - u^n_{i+\frac{1}{2}}}{\Delta x}, \quad (13)$$

$$(u^n_x)_{i+\frac{1}{2},j,k} = \Gamma((u^n_x)^{\mathrm{L}}_{i+\frac{1}{2},j,k}, (u^n_x)^{\mathrm{R}}_{i+\frac{1}{2},j,k}), \quad (14)$$

where Γ is the slope limiter, that will be described later. The procedure is similar for other directions and velocity components.

The time derivatives in eq. (8-11) are expressed as

$$(u^n_t)_{i+\frac{1}{2},j,k} = \Delta_{i+\frac{1}{2},j,k}u^n - \nabla_{i+\frac{1}{2},j,k}p^n - (u_x)^n_{i+\frac{1}{2},j,k}u^n{}_{i+\frac{1}{2},j,k} - \quad (15)$$

$$-(\hat{u_y})^n_{i+\frac{1}{2},j,k}\bar{v}^n_{i+\frac{1}{2},j,k} - (\hat{u_z})^n_{i+\frac{1}{2},j,k}\bar{w}^n_{i+\frac{1}{2},j,k}, \quad (16)$$

where $\bar{v}$ and $\bar{w}$ are an averages of v and w in cells adjacent to $u_{i+\frac{1}{2},j,k}$

$$\bar{v}^n_{i+\frac{1}{2},j,k} = \frac{v^n_{i,j+\frac{1}{2},k} + v^n_{i+1,j+\frac{1}{2},k} + v^n_{i,j-\frac{1}{2},k} + v^n_{i+1,j-\frac{1}{2},k}}{4}, \quad (17)$$

and normal derivatives $u_{\hat{y,z}}$ are computed as:

if $v_{i+\frac{1}{2},j,k} \geq 0$

$$(\hat{u_y})^n_{i+\frac{1}{2},j,k} = \frac{u_{i+\frac{1}{2},j,k} - u_{i+\frac{1}{2},j-1,k}}{\Delta y} + \quad (18)$$

$$+\frac{1}{2}\left(1 - \frac{\bar{v}^n_{i+\frac{1}{2},j,k}\Delta t}{\Delta y}\right)\frac{(u^n_y)_{i+\frac{1}{2},j,k} - (u^n_y)_{i+\frac{1}{2},j-1,k}}{\Delta y}, \quad (19)$$

or if $v_{i+\frac{1}{2},j,k} \leq 0$

$$(\hat{u_y})^n_{i+\frac{1}{2},j,k} = \frac{u_{i+\frac{1}{2},j+1,k} - u_{i+\frac{1}{2},j,k}}{\Delta y} + \quad (20)$$

$$+\frac{1}{2}\left(1 - \frac{\bar{v}^n_{i+\frac{1}{2},j,k}\Delta t}{\Delta y}\right)\frac{(u^n_y)_{i+\frac{1}{2},j+1,k} - (u^n_y)_{i+\frac{1}{2},j,k}}{\Delta y}. \quad (21)$$

(22)

One unique value of $u_{i,j,k}$ and $u_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}$ have to be obtained from reconstructed values from eq. (8-11). For cell centered values we used the same upwind procedure, as Tau [1994], while for

the corner values we used simple averaging in this version of the code

$$u_{i,j,k}^{n+\frac{1}{2}} = \begin{cases} u_{i,j,k}^{n+\frac{1}{2},\mathrm{W}} & \text{if } u_{i,j,k}^{n+\frac{1}{2},\mathrm{W}} \geq 0 \text{ and } u_{i,j,k}^{n+\frac{1}{2},\mathrm{W}} + u_{i,j,k}^{n+\frac{1}{2},\mathrm{E}} > 0, \\ 0 & \text{if } u_{i,j,k}^{n+\frac{1}{2},\mathrm{W}} < 0 \text{ and } u_{i,j,k}^{n+\frac{1}{2},\mathrm{E}} > 0, \\ u_{i,j,k}^{n+\frac{1}{2},\mathrm{E}} & \text{otherwise.} \end{cases}, \tag{23}$$

$$u_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} = \frac{(u^{\mathrm{SW}} + u^{\mathrm{SE}} + u^{\mathrm{NW}} + u^{\mathrm{NE}})_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}}}{4}. \tag{24}$$

Finally, we can compute the advective terms in the equation (3) as

$$\begin{aligned} \left[u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right]_{i+\frac{1}{2},j,k}^{n+\frac{1}{2}} &= \frac{u_{i+1,j,k}^{n+\frac{1}{2}} + u_{i,j,k}^{n+\frac{1}{2}}}{2}\frac{u_{i+1,j,k}^{n+\frac{1}{2}} - u_{i,j,k}^{n+\frac{1}{2}}}{\Delta x} \\ &+ \frac{v_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + v_{i+\frac{1}{2},j+\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}} + v_{i+\frac{1}{2},j-\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + v_{i+\frac{1}{2},j-\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}}}{4} \cdot \\ &\cdot \frac{u_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + u_{i+\frac{1}{2},j+\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}} - u_{i+\frac{1}{2},j-\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i+\frac{1}{2},j-\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}}}{\Delta y} \\ &+ \frac{w_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + w_{i+\frac{1}{2},j-\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + w_{i+\frac{1}{2},j+\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}} + w_{i+\frac{1}{2},j-\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}}}{4} \cdot \\ &\cdot \frac{u_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} + u_{i+\frac{1}{2},j-\frac{1}{2},k+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i+\frac{1}{2},j+\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}} - u_{i+\frac{1}{2},j-\frac{1}{2},k-\frac{1}{2}}^{n+\frac{1}{2}}}{\Delta z}. \end{aligned} \tag{25}$$

2.3 Slope limiters

Slope limiter $\Gamma(a,b))$ is a nonlinear function, that determinates the slope of reconstruction of a variable in the computation cell, in such a way, that no new extrema are produced. In this study we used following TVD limiters:

minmod	$\Gamma_{\mathrm{MM}}(a,b) = \begin{cases} \min(a,b) & \text{if } a>0 \wedge b>0, \\ \max(a,b) & \text{if } a<0 \wedge b<0, \\ 0 & \text{otherwise} \end{cases},$
generalised minmod	$\Gamma_{\mathrm{EMM}}(a,b) = \Gamma_{\mathrm{MM}}(\theta a, \theta b, \frac{a+b}{2}),$
van Albada	$\Gamma_{\mathrm{VA}} = b * \frac{r+r^2}{1+r^2},\ r = a/b,$
van Leer	$\Gamma_{\mathrm{VL}} = b * \frac{r+\lvert r\rvert}{1+r},\ r = a/b.$

For the extended minmod limiter we used the value $\theta = 2$, which is the maximum that holds the TVD property.

3 RESULTS

One of the most often computed cases of turbulent decaying flow is the Taylor-Green vortex. This flow is defined in a box of $(0, 2\pi)^3$ by initial conditions for the velocity in the form

$$u = \sin(x)\cos(y)\cos(z), \tag{26}$$

$$v = \cos(x)\sin(y)\cos(z), \tag{27}$$

$$w = 0, \tag{28}$$

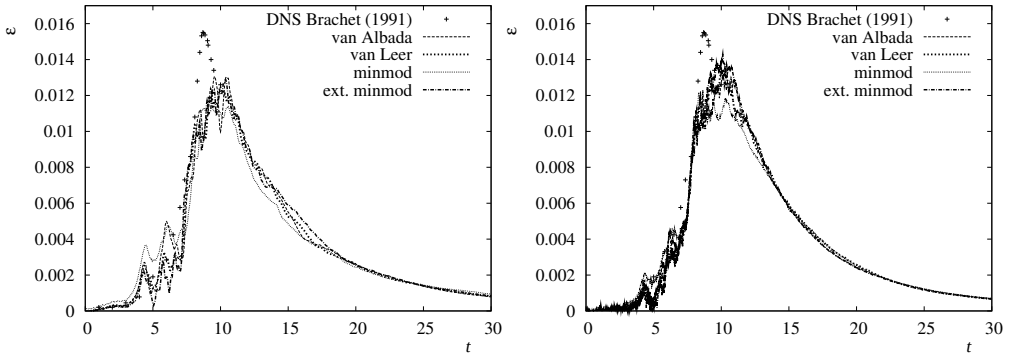


Figure 1: Plots of kinetic energy dissipation at resolution 64^3 (left) and 128^3 (right).

and by the pressure given by a solution of a pressure Poisson equation

$$p = p_0 + \frac{1}{16}(2 + \cos(2z))(\cos(2x) + \cos(2y)), \tag{29}$$

where p_0 is arbitrary (in the case of incompressible flow).

Taylor-Green vortex shows complex behaviour. At the beginning the flow is purely two-dimensional. The flow continues with forming of vortex sheets, breaking into smaller eddies, and after $t \approx 5$ the flow becomes turbulent. According to DNS of Brachet [1991] at $t \approx 9$ is the peak of the enstrophy and kinetic energy dissipation. After this peak the turbulent eddies dissipate in the self-similar energy cascade.

We performed our calculations on uniform grids with 64^3 and 128^3 cells.The molecular diffusion was set to a negligible value $\nu = 1/50000$, which results in the Reynolds number unresolvable exactly on current grids. Therefore numerical diffusion of the scheme had to act as an implicit subgrid stress model to transfer the energy to the unresolved scales of motion. For comparison we will use the DNS results of Brachet [1991] at $\mathrm{Re} = 5000$, that are very close to results at $\mathrm{Re} = 3000$ and therefore also close to the Reynolds number independence.

In figure 1 is plotted time history of kinetic energy dissipation, defined as

$$\varepsilon = -\frac{\mathrm{d}K}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\frac{\langle \|\boldsymbol{u}\|^2 \rangle}{2}, \tag{30}$$

scaled for direct comparison with Brachet [1991]. From the graph it is clear that the basic features of the time dependence are captured by the calculations. The 64^3 runs produced the broader and lower peak, than in 128^3. In both cases, however, the top of the main peak was lower, than in the DNS. This is in a good agreement with the ILES simulations in [Grinstein et al., 2007a]. The peak was also a little late. The differences between individual limiters are clearly visible, at least in 64^3 case. The most distinct one is the minmod limiter, which turns out to be overly diffusive. From the other limiters, extended minmod seems to produce result closest to the DNS. The time history of the kinetic energy K for the van Albada limiter (which is representative for the other) is depicted in figure 2. For a short time interval around $t = 10$ there can be found an area with scaling approximately $K \sim t^{-1.2}$, but for $t > 12$ there is a clear scaling $K \sim t^{-2}$. This is in agreement with Grinstein et al. [2007a], but our $t^{-1.2}$ interval seems to be too short.

At the time $t = 30$ we also computed the probability density functions (PDFs) of the velocity gradients and the pressure end the 3D kinetic energy spectra. In figure 3 are plotted examples of the PDFs for the extended minmod limiter (there were almost no visible differences between limiters).

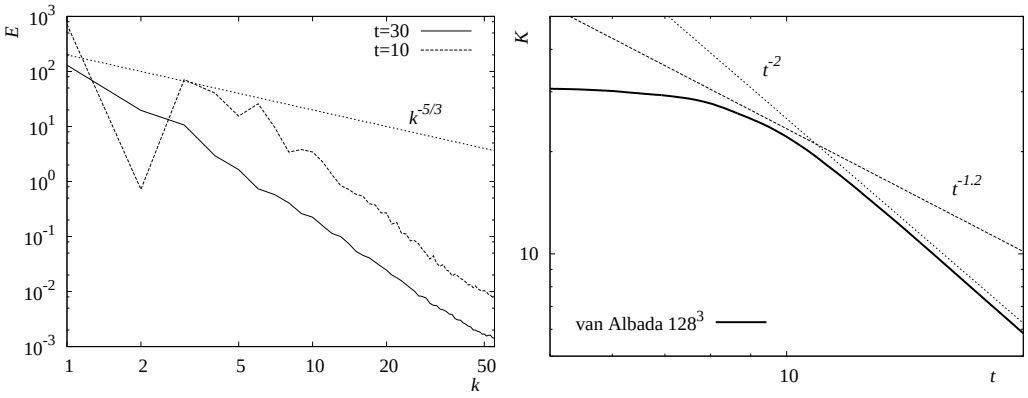


Figure 2: a) 3D spectra of kinetic energy and b) time history of total kinetic energy for van Albada limiter at resolution 128^3

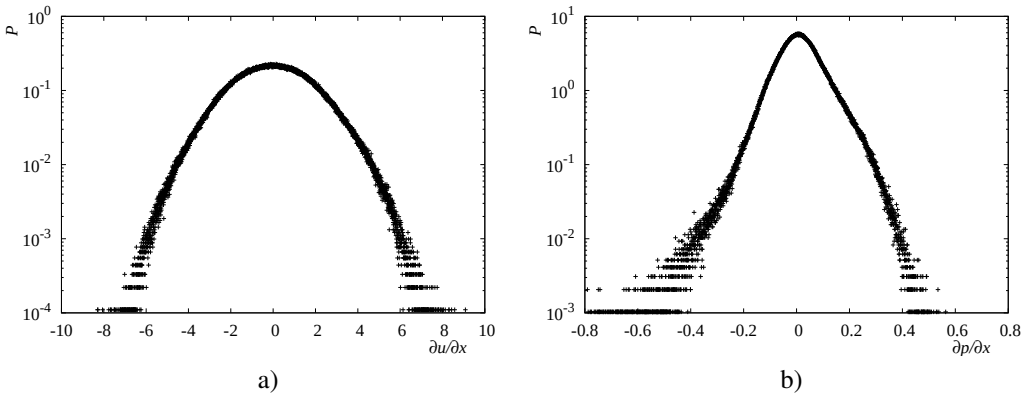


Figure 3: PDFs of a) $\frac{\partial u}{\partial x}$, b) $\frac{\partial p}{\partial x}$ for extended minmod limiter at resolution 128^3.

For both the velocity and the pressure the tails of the distributions show almost exponential nature as expected. However, the energy spectrum (for the van Albada limiter in figure 2) for $t = 10$ and $t = 30$ does not contain a clear inertial subrange with the Kolmogorov $k^{-5/3}$ power law. For $t = 30$ the spectrum scales as k^{-3} instead. All important wavenumbers are probably affected by the numerical dissipation of the scheme. Similar results were reported by Garnier et al. [1999].

In figure 4 are plotted isosurfaces of the pressure and the second largest eigenvalue of velocity gradient tensor [Jeog and Hussain, 1995]

$$\lambda_2 = \|\nabla \times \boldsymbol{u}\| - \sqrt{\|\nabla u\|^2 + \|\nabla v\|^2 + \|\nabla w\|^2}. \tag{31}$$

Both variables visualize vortex cores with pressure showing only the largest eddies and λ_2 more of the smaller eddies. Correlation between pressure lows and maxima of λ_2 is apparent.

4 CONCLUSIONS

We have developed a 3D model for incompressible flows using a projection method and a high-resolution method for the advective fluxes. We tested capability of this model for turbulent flow as an implicit large eddy simulation model using the Taylor-Green vortex. The model was capable to describe correctly many features of turbulent flow including the kinetic energy dissipation. One of

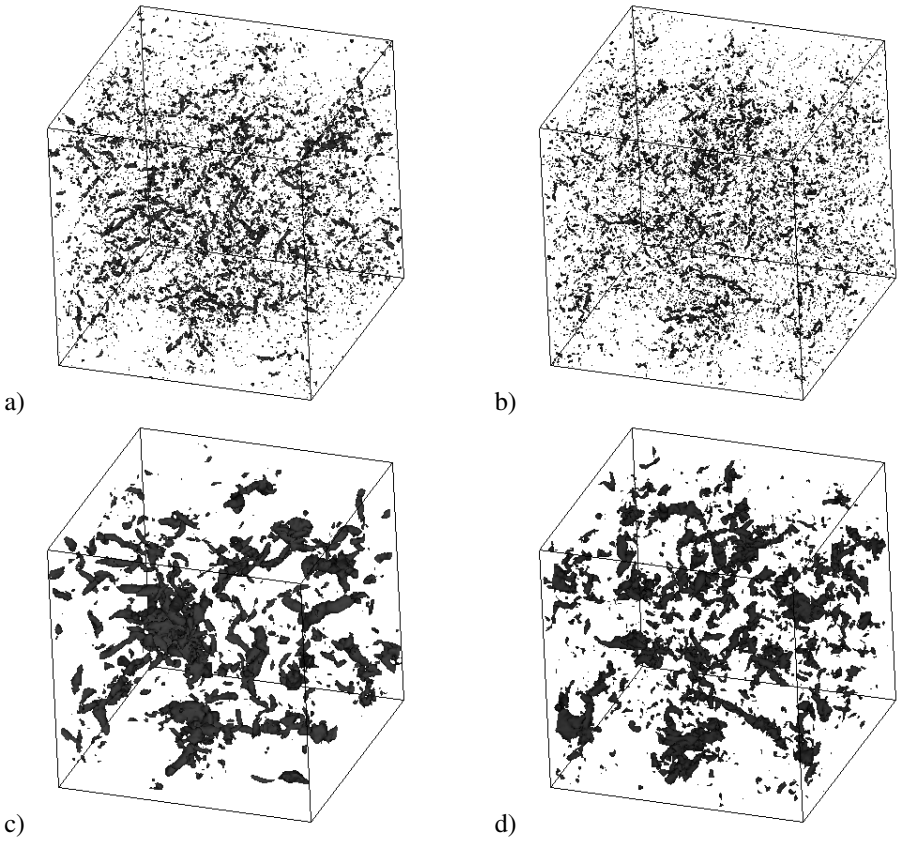


Figure 4: Volume renderings of a) λ_2 for minmod limiter, b) λ_2 for extended minmod limiter, c) pressure for minmod limiter, d) pressure for extended minmod limiter. All snapshots are at $t = 30$.

the remaining problems is the slope of the 3D energy spectrum, which does not contain well developed inertial subrange with $k^{-5/3}$ scaling. The development of the model will continue. We will try several other possibilities in the advective fluxes discretization (for example upwinding also in eq. (24). The next step will be simulations of wall bounded flows in a simple geometry (fully developed channel flow etc.). If the model will perform well we will continue with computations of separating flow, with the urban boundary layer as our main aim.

ACKNOWLEDGEMENTS

This research was supported by the Grant Agency of the Czech Academy of Sciences, grant no. T400300414, by the Grant Agency of the Czech Republic, grant no. 205/06/0727 and by the Czech Ministry of Education, Youth and Sports in the framework of the research plan MSM0021620860.

REFERENCES

Almgren, A., J. Bell, P. Colella, L. Howell, and M. Welcome. A conservative adaptive projection method for the variable density incompressible Navier-Stokes equations. *J. Comput. Phys.*, 142:1–46, 1998.

Almgren, A., J. Bell, and W. Szymczak. A numerical method for the incompressible navier-stokes equations based on an approximate projection. *SIAM J. Sci. Comput.*, 17:2, 1996.

Bell, J. B., P. Colella, and H. M. Glaz. A second-order projection method for the incompressible Navier-Stokes equations. *J. Comput. Phys.*, 85:257–283, 1989.

Bensow, R. E., M. G. Larson, and P. Vesterlund. Vorticity-strain residual-based turbulence modelling of the Taylor-Green vortex. *Internat. J. Numer. Methods Fluids*, 54:745–756, 2007.

Boris, J. P. and D. L. Book. Flux-corrected transport I. SHASTA, a fluid transport algorithm that works. *J. Comput. Phys.*, 135(2):172–186, 1997.

Brachet, M. E. Direct simulation of three-dimensional turbulence in the Taylor-Green vortex. *Fluid Dyn. Res.*, 8:1–8, 1991.

Brachet, M. E., D. I. Meiron, S. A. Orszag, B. G. Nickel, R. H. Morf, and U. Frisch. Small-scale structure of the Taylor-Green vortex. *J. Fluid Mech.*, 130:411–452, 1983.

Brown, D. L., R. Cortez, and M. L. Minion. Accurate projection methods for the incompressible Navier-Stokes equations. *J. Comput. Phys.*, 168:464–499, 2001.

Colella, P. and P. R. Woodward. The piecewise parabolic method (PPM) for gas-dynamical simulations. *J. Comput. Phys.*, 54:174–201, 1984.

Garnier, E., M. Mossi, P. Sagaut, P. Comte, and M. Deville. On the use of shock-capturing schemes for large-eddy simulation. *J. Comput. Phys.*, 153(2):273–311, 1999.

Grinstein, F., L. Margolin, and W. Rider. chapter Numerics for Iles, Limiting Algorithms. Cambridge University Press, Cambridge, 2007a.

Grinstein, F., L. Margolin, and W. Rider. *Implicit Large Eddy Simulation. Computing Turbulent Fluid Dynamics*. Cambridge University Press, Cambridge, 2007b.

Harten, A. High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys.*, 49: 357–393, 1983.

Jameson, A., W. Schmidt, and E. Turkel. Numerical simulation of the euler equations by finite volume methods using Runge-Kutta time stepping schemes. In *AIAA 5th Computational Fluid Dynamics Conference (1981)*, pages AIAA Paper 81–1259, 1981.

Jeog, J. and F. Hussain. On the identification of a vortex. *J. Fluid Mech.*, 285:69–94, 1995.

Kim, J. and P. Moin. Application of a fractional-step method to incompressible Navier-Stokes equations. *J. Comput. Phys.*, 59:308–323, 1985.

Kolditz, O. *Computational Methods in Environmental Fluid Mechanics*. Springer, Berlin, 2002.

van Leer, B. Towards the ultimate conservative difference scheme V. a second-order sequel to Godunov's method. *J. Comput. Phys.*, 32:101–136, 1979.

Rider, W. Filtering non-solenoidal modes in numerical solutions of incompressible flows. *Internat. J. Numer. Methods Fluids*, 28:789–814, 1998.

Smolarkiewicz, P. K. and L. G. Margolin. MPDATA: a finite-difference solver for geophysical flows. *J. Comput. Phys.*, 140(2):459–480, 1998.

Smolarkiewicz, P. K., L. G. Margolin, and A. A. Wyszogrodzki. Implicit large-eddy simulation in meteorology: From boundary layers to climate. *J. of Fluid Eng.*, 129:1533–1539, 2007.

Tau, E. Y. A second-order projection method for the incompressible Navier-Stokes equations in arbitrary domains. *J. Comput. Phys.*, 115(1):147–152, 1994.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Iterative Inverse Modelling Method for Locating a Source of Air Pollution and Its Robustness

B. Rajkovic[a], M. Vujadinovic[a] and Z. Grsic[b]

[a] *Institute of Meteorology, Faculty of Physics, Belgrade University, Dobracina 16, 11000 Belgrade, Serbia (bora@ff.bg.ac.yu)*
[b] *Institute of Nuclear Sciences Vinca, P.O. Box 522, 11001 Belgrade, Serbia (grsa@vin.bg.ac.yu)*

Abstract: Detection of a possible source of air pollution as a combination of measurements and inverse modelling, based on Bayesian statistics, has been proposed. The simplicity of the approach and its numerical efficiency qualifies this approach for the problem, especially when it is used in the operational mode. The method has been examined in its simplest form, with a single source and the implicit assumption that we know the moment of the release. The position of the possible source has been found as the maximum of the probability density function from an ensemble of possible sources that cover substantial part of the model domain. Members of the ensemble have been generated using a puff model. Search for the position consists of series of iterations converging toward the position of the source.

***Keywords**:* point source; puff model; Bayesian statistics; probability density function.

1. INTRODUCTION

Detection of air pollution is a multi aspect problem. The first step is to design an optimal network of measuring points (stations), relative to the envisaged or possible sources of pollution. Then, based on measurements, quick and, as accurate as possible, detection of an accidental release should follow. It is important to emphasize that only measurements are not sufficient to create (analyze) concentration field, even if we have quite large number of sampling points. The main reason is that, generally, concentration fields have large gradients. We show this in Appendix, through series of idealized experiments. The only way to overcome this situation is by introducing Bayesian statistical method, as a component of the solution [Fuentes and Raftery 2001, Wikle et al. 2001, among others]. Its additional advantage is taking explicitly into account the measurement errors which are unavoidable and depended on the methodology, as well as on the quality of the instruments that are involved in the process.

Beside these, relatively straightforward ways of estimating quality of information from small number of monitoring points, there are several other approaches, although of higher sophistication. One is to use the spatial covariance structure of the quality patterns from the so called "sphere of influence" around each site as in Langstaff et al. [1987], Buell [1975] and Van Egdmond and Onderdielinden [1981], among others. Another approach is to generate data in a relatively large set of points (dense grid) and then analyze average concentration, its clustering, frequency and clustering of average concentration, as in Tseng and Chang [2001]. In this paper, we will concentrate only on the detection of possible source position, as an aspect of the inverse problem, assuming that there exist measurements of the concentrations of a passive substance.

2. MODEL

The central point is to calculate probability density function (PDF) as a most complete description of the system. It is defined as a conjunction of three spaces, space of possible values of a parameter (m), space of simulated measurements (d_m) and space of actual measurements (d). Using Bayesian statistics we combine then into a single PDF

$$p(m,d,d_m) = const \cdot \frac{\rho(m,d) \cdot f(m,d_m)}{\nu(m,d)} \quad , \tag{1}$$

where ρ and f are PDFs defined on Cartesian product of parameters and real measurements space and parameters and simulated measurements space, respectively. *Const* is the normalization constant

$$const = \int_V \frac{\rho(m,d) \cdot f(m,d_m)}{\nu(m,d)} dV \quad , \tag{2}$$

where $\mathbf{d}V = \mathbf{d}m \cdot \mathbf{d}d \cdot \mathbf{d}d_m$. Function $v(m,d)$ represents the homogeneous state of information. Its particular form doesn't play a role, except in some highly degenerated problems that are not considered here [Tarantola, 2005]. After taking into account that set of actual and simulated measurements are independent, we get the final form for PDF as

$$p(m,d,d_m) = const \cdot \frac{\rho_M(m) \cdot \rho_D(d) \cdot f(d_m / m)}{\nu_D(d)} \quad . \tag{3}$$

The model that we used to construct probability density function of the possible source is a puff model [Grsic and Milutinovic 2000, Rajkovic et al. 2008]. The model starts with measured (10m) winds and extrapolate it to the height of each puff's center. In several previous simulations, measured winds were weaker than 1m/s, but results were still quite reasonable. The model has capability to calculate the rise of a warm gas to its neutral buoyancy height, but in these runs, we assumed that the gas has the ambient temperature. Vertical mixing is calculated using the Pasquill-Gifford stability classes. The effects of the relief are included according to the stability class. In the case of A, B, C and D class, a puff ascends over an obstacle with the same height as in front of the obstacle. In the case of stable stratification (E, F), a puff moves up the obstacle at fixed height of 10m. At the moment, the model does not estimate the stochastic part of the motion of particles. The grid had 301x301 points with spatial distance of 60 meters thus spanning the area of about 40 square kilometers. The time interval between two consecutive puffs was one minute.

3. METHODOLOGY AND RESULTS

Relative position of the assumed source and measuring point are presented in the Figure 1, upper panel. The first step was to generate field of a passive substance by its release during 120 minutes. Values at the measurement points were then randomly perturbed by 5%, mimicking the measurement errors. The second step was to create the cluster of points, possible sources, with center positioned in the measuring point with the highest concentration observed. From the cluster points we calculated PDF of the source position, assuming its Gaussian shape [Tarantola, 2005]. In the second iteration, we translated the center of a cluster to the position of the first PDF's maximum. Again, we calculated PDF and its maximum and translated the center to the position of the new maximum. Since now the cluster member with the maximum PDF was at the inside area of the previous cluster, we decided also to halve the spread among the third cluster's members. In the next two iterations we again translated the cluster and halved the distance between its members.

The error in the calculated distance after these five iterations was about 120 meters (two grid points). The whole procedure takes 12 minutes on a comparatively new personal computer thus clearly fulfilling the efficiency requirement. The accuracy is also acceptable.
After this, we increased margin of errors from 5% to 30% and increased variability of the Gaussian distribution. In both cases convergence was very close to the first case. In the case

of the increase of margin of errors, found position was one grid point away from the original margin of errors. In the case of σ, the result was the same as in the standard case.

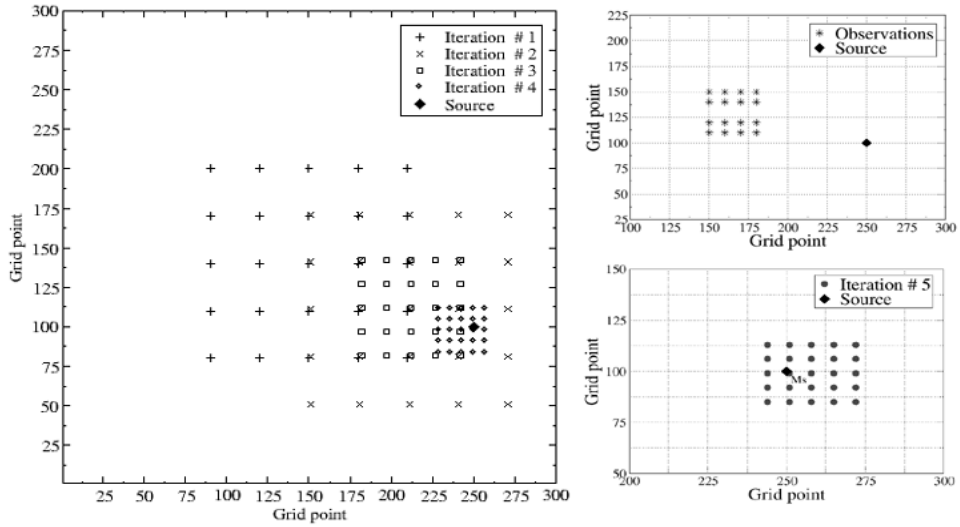


Figure 1. The right upper panel shows relative positions of the source and observation points. The left panel presents changes of the cluster position during four iterations, while the position of the final, fifth, cluster is shown on the right lower panel, where Ms denotes modelled position of the source

4. CONCLUSIONS

Combination of simulated measurements and Bayesian statistical approach through several iterations results in a very accurate yet efficient method of detecting a point like source of pollution. Sensitivity to the assumed margin of errors and σ, Gaussian distribution parameter, was quite small.

5. APPENDIX

Here we will show that even high number of measurements can not solve the problem of finding source(s) position. Let us generate concentration field with a release of seventeen sources. The geometry resembles the situation in the petrochemical zone of city Pancevo nearby Belgrade. In Figure A1, on the right panels, is a field of 24 hour-average concentration field obtained with a Gaussian plum diffusion model [Rajkovic at all, 2008] for the continuous source, using 10 minutes averages of the wind field and stability parameters with Pasquill-Gifford stability categories. The only deviation from the actual situation is the assumption that all sources are at the same height. The grid used for calculations was very dense one, with 90601 points (301x301). The idea is that such high density grid will reduce problems with the accuracy of the model and thus calculated field can be regarded as an "observed" one, i.e. we regard values in grid points as the measured ones. Formation of the isolines (graphical representation of the field) was done using the Kriging method. In this case of very dense grid we can clearly see positions of all sources as the positions of the local maximums.

If we now reduce the number of grid points, and therefore number of "sampling" points, by two orders of magnitude (31x31), we get situation depicted in the first row of Figure A1, left panel. We use again Kriging method to form the "continuous" field. Comparison with the starting situation, right panel of the same figure row, shows that we still have very high resemblance with the original field. This situation is still far from a realistic one regarding possible density of actual sampling points and tells us more about the possibility of reduction of computation effort.

Grid with 36 points (6x6) is closer to the regulatory recommendations and economical feasibility. Again, with the help of Kriging method, we arrive to the result in Figure A1, the second row. Same as in the first row, on the left we have the original concentration filed and on the right is the result from 36 points. Dots on the left panel are positions of the "sampling" locations. We now see that the distribution field is very different from the starting one, showing only the gross characteristics of the pollution concentration fields, but with complete loss of its local details. This still may be acceptable as a very rough assessment of the long term influences of pollution sources for that area, but totally insufficient in an accident situation.

Finally, we have created irregular set of "sampling" stations that qualitative represents the actual situation in that area. As it is often in the real life situations, several points are very close to each other, while others are relatively far away (Figure A1, the third row, left panel). Using the same analysis method, we get the distribution shown in Figure A1, the third row, right panel. The general structure is similar to the previous case with a few exceptions. This distribution is somewhat closer than the one with 36 points, due to the fact that several points "samplings" are very close to the position of the sources and thus accidentally giving better result. It is also possible that this particular wind direction has "helped" in this case.

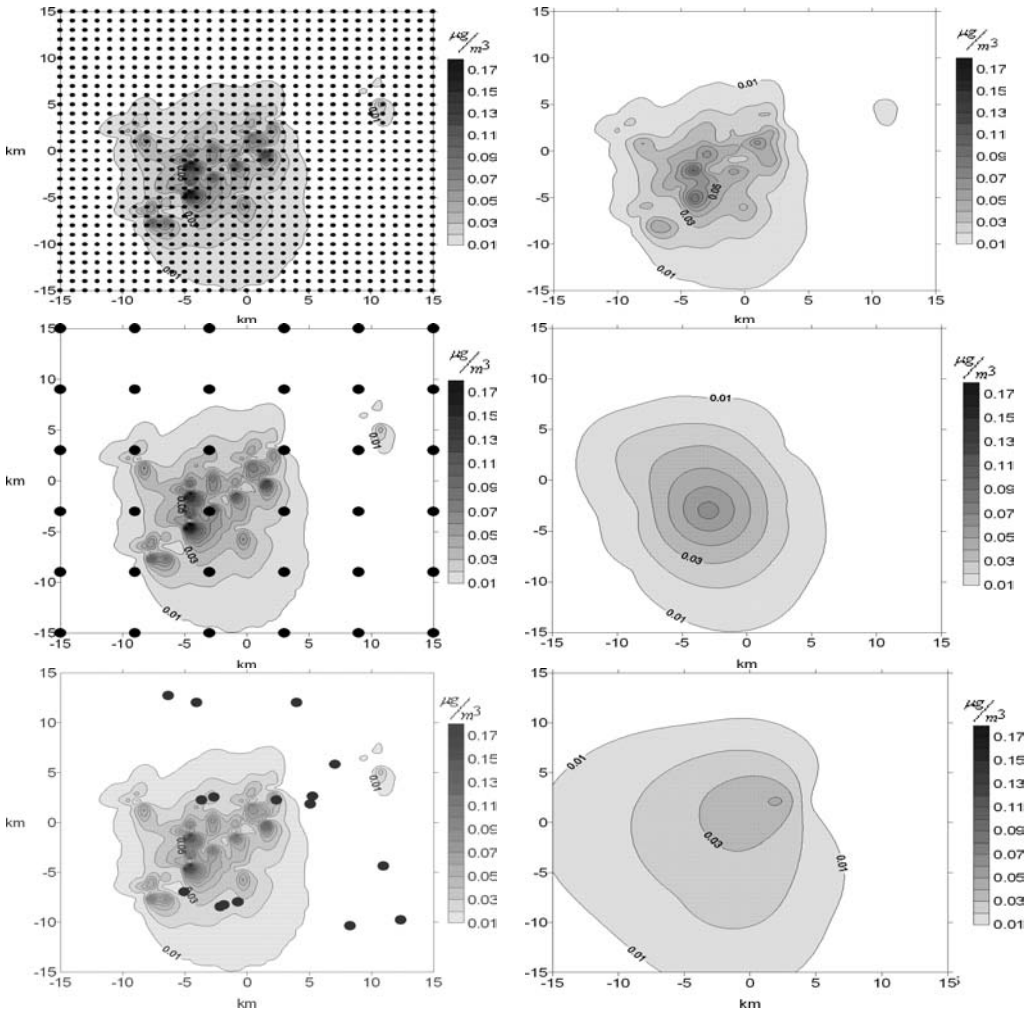


Figure A1. The right panel in every row shows the same concentration field, obtained from a release of 17 sources using a Gaussian plum model with very dense grid (301x301 grid point). Black dots on each right panel represent "sampling" points. In the first row there is 31x31 "sampling" point, in the second row 6x6, and in the third row, there is 16 irregularly spaced "sampling points". Left panels show concentration fields interpolated, using Kriging method, from "sampling" points on the right.

ACKNOWLEDGMENTS

This work has been partially funded by Serbian Ministry of Science, Technology and Environment, grant no. 1197 and Italian Ministry of Environmental and Territories through its two projects, SINTA and ADRICOSM-STAR. The third author is partially funded through the project of the decommission of the nuclear reactor in Vinca.

REFERENCES

Buell, C.E., The topography of the empirical orthogonal functions. *4th Conference on Probability and statistics,* American Meteorological Society, Boston, 188-193, 1975.

Grsic, Z., and P. Milutinovic, Automated meteorological station and the appropriate software for air pollution distribution assessment, *Air Pollution Modelling and Its Application XIII*, 2000.

Fuentes, M. and A. Raftery, Model evaluation and spatial interpolation by combining observations with outputs from numerical models via Bayesian Melding, *Technical report,* No 403, 2001.

Langstaff, J., Seigneur, C., Liu, M.K., Behar, J.V. and J.L. McElroy, Design of an optimum air monitoring network for exposure assessments, *Atmospheric Environment*, 21, 1393-1410, 1987.

Rajkovic, B., Arsenic, I. and Z. Grsic, *Fluid Mechanics of Environmental Interfaces*, Taylor and Francis, 348 pp., London, 2008.

Tarantola, A., *Inverse problem theory and methods for model parameter estimation*, SIAM, 342 pp., Philadelphia, 2005.

Tseng, C.C and N.B. Chang, Assessing relocation strategies of urban air quality monitoring stations by GA-based compromise programming, *Environment International*, 26 (7), 523-541, 2001.

Van Egmond, N.D. and D. Onderdelinden, Objective analysis of air pollution monitoring network data; spatial interpolation and network density, *Atmospheric Environment*, 15, 1035-1046, 1981.

Wikle, C.K., Milliff, R.F., Nychka, D. and L.M. Berliner, Spatiotemporal hierarchical Bayesian modelling: Tropical ocean surface winds, *Journal of the American Statistical Association*, 96(454), 382-397, 2001.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Evolvable Biological Interfaces: An Outline of the New Computing System

I. Balaz[a], **D.T. Mihailovic**[b]

[a]*Scientific Computing Laboratory, Institute of Physics, P.O.Box 57, 11001, Belgrade, Serbia (ibalaz@polj.ns.ac.yu)*
[b] *Faculty of Agriculture, University of Novi Sad, Dositej Obradovic Sq. 8, 21000 Novi Sad, Serbia (guto@polj.ns.ac.yu)*

Abstract: In this paper, we represent architecture of a new individual-based model of evolvable biological interfaces. The model is based on three parts: mutable coding sequence, individual agents produced by the coding sequence and external environment represented by various virtual substances. Agents are able to recognize external substances and to perform transformations on them. However, unlike general computational scheme currently used in evolutionary computation which is based on predefined list of actions, where each action have predefined fitness value defined strictly within current problem, the recognition and corresponding transformation in this model are organized in completely different way. Here we merged input-action-output system into one unified interdependent whole where each element's role is relative to its surrounding, and instead of unambiguous symbols we introduced alphabet where each symbol is characterized by a set of characteristics. As a result, we obtained system which is able to eventually, through evolution, recognize and assimilate unlimited number of new substances and to perform on them non-fixed set of transformations.

Keywords: Biological interface; Evolvability; Individual-based model.

1. INTRODUCTION

In most general terms interface can be defined as a surface that forms a common boundary between two media (objects, phases or systems). We can see that environment is abundant with various interfaces and that any kind of functioning or processing that can be described with input/output schema, by definition deals with interfaces. Such interfaces have been defined as interfaces between the two either abiotic or biotic environments which are in a relative motion exchanging energy through biophysical and chemical processes and fluctuating temporally and spatially regardless of its space and time scale [Mihailovic and Balaž, 2007]. This broad definition allows us to apply a numerous different approaches when dealing with environmental interfaces, depending on problems that are in focus of our attention. For example, great number of particular problems can be solved using formalisms taken from nonlinear dynamics [see: Doelman and Van Harten, 1995] or environmental fluid mechanics [see for example contributions from Gualtieri and Mihailovic, 2008]. However, dealing with biological interfaces, we are entering a peculiar field, which, in order to be fully grasped, demands development of new conceptual and formal tools.

As it is already implied, one of the pillars of establishing and maintaining functionality in living systems is successful formation of operative interface with external environment. It should provide robust and prompt translation of vast diversity of external physical and/or chemical changes into a set of signals that are "understandable" for an organism [Amils et al., 2007; May and McLean, 2007]. Through organism's physiological processes, such signals are then processed resulting in activation of adaptive reactions to observed changes. However, process of creation of understandable signals from raw environmental material is far from simple and unambiguous. There have been several attempts to analyse it, either from the most wider perspective by developing general theory of signs by C.S.Peirce [for

an overview see: Atkin, 2006], or from more particular perspectives, taking human language [de Saussure, 1965] or animal communication [Hoffmeyer, 1997] as focal points. Although they provide us with some very useful concepts and make a solid path for further investigations, they are primarily developed as philosophical considerations and as such have to be further operationalised for our present purpose. Therefore, in this paper we will take most general concepts developed in aforementioned literature:

- each individual is considered as an observer which resides within some environment;
- observation is an act of assimilating a segment of external changes with appropriate set of internal operations in order to produce reaction;

and extend them further with more specific postulates, derived from general patterns of functioning of living systems:

- operational environment is defined from the perspective of that individual so that external changes are functionally existent for individual only if they can be observed;
- by observing, the environment is divided (schematized) into operative segments, spatial and temporal;
- schematisation models are the same endogenous factors that determine internal functioning of the individual;

The whole process, described in these five postulates will be called “operative assimilation”.

Since all organisms are result of slow and long process of evolution, their interfaces (realized through operative assimilation) are well suited to recognize environmental changes, which can occur in their niches. However, environment is not a fixed system with fully predictable outcomes, especially on a long-term scale. Accordingly, interfaces, as well as complete organization of living systems, is organized in such a way to eventually enable later incorporation of external changes not existent or not accessible at a current moment. Such property is called evolvability. More precisely, it has been defined as the capacity to generate heritable, selectable phenotypic variation [Kirschner and Gerhart, 1998]. Since that definition is formulated having real living organisms in mind, it presuppose a great number of conditions which, in the field of modelling, should be one by one explicitly stated and generated. However, full analysis will lead us to far more famous question of how living systems are organized and how their organization can be grasped by formal language, which is out of scope of this paper. Therefore, we will concentrate only on a subset of posed problem, i.e. what kind of computational architecture can be able to eventually incorporate occurrence of arbitrary complex objects in time t_{n+1}, which are not existent or not accessible at time t_n, behaving in accordance with outlined approach termed as operative assimilation. In achieving this goal, we employed individual-based modelling strategy.

Individual-based modeling is somewhat imprecise umbrella term for a loosely connected group of computational models that relies on population-based techniques. Foundations for them lies in machine learning technique called Learning Classifier Systems [Holland, 1992], which is a rule-based system where rules are in form of IF/THEN conditions. Such systems are composed of (i) input interface, which receive allowable information from the environment, (ii) base of rules consisting of N condition-action rules (called classifiers), (iii) internal mechanism for evolution usually based on application of Genetic Algorithm and Reinforcement learning techniques, and (iv) of output interface which define corresponding action of an individual, after receiving input information. Like in natural systems, not all external changes are observable by an individual, but only those for which it have appropriate receptors. These receptors are able to recognize strings composed from defined ternary alphabet $T = \{0,1,\#\}$, where # is a wildcard, which can take a role of either 0 or 1. After recognizing external information, list of classifiers is scanned and all rules that match the external information are selected to form the current “match set”. Each rule has assigned fitness value to indicate its usefulness in receiving external reward. From the formed match set, appropriate rule is selected, according to its highest fitness value, and transferred to the output interface to become the individual’s external action. Such

architecture is very general and applicable to a variety of disciplines like artificial intelligence, ecological modeling, robotics, adaptive control [Bull, 2004].

In this paper we represent architecture of a new individual-based model of evolvable biological interfaces based on the notion of operative assimilation, which is evolutionary able for autonomous recognition of potentially unlimited number of new substances and changeability of performed transformations on them. The model has been written using Python language, Firebird database and kinterbasdb API.

2. DESCRIPTION OF THE MODEL

2.1 Introduction

It is clear that, in order to obtain non-trivial, real-life reflecting evolvability of the formal system, it should be placed in environment with which it have some kind of functional interaction, defined in this paper through process of "operative assimilation". Otherwise, system can produce variations, and they can be selected according to some rule(s) but within a framework of modeling ecological process of evolution, it is a dead-end procedure which can be useful for solving some specific problem, but will lack general flexibility. Another necessary condition is existence of common pattern of creating building blocks of both: elements of the system itself and elements of external environment. Here, external environment is composed of other systems, as well as of extra-systemic influences. It should be emphasized that, no matter how variable are extra-systemic influences, they all share a common property of being a milieu in which particular building blocks of systems are created. In other words, all of them greatly affect formation of outcomes of systemic processes and, even more important, these processes evolved to have particular organization and dynamics exactly because existence of typical set of external influences. It means that there cannot exist rule that is valid for external environment, and non-valid for the system itself (valid does not means the same).

2.2. Mathematical background

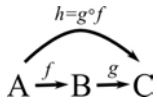
Stated more precisely, and following Mesarovic's mathematical Theory of General Systems [Mesarovic, 1972], if we observe interactions of agents with surrounding environment, such system can be defined as a set of interacting objects $S \subseteq O_1 \times O_2 \times O_3 \times ... \times O_n$. If we denote population of agents under consideration by $P = \{p_1, p_2, p_3, ..., p_n\}$ and a set of external influences by $E = \{e_1, e_2, e_3, ..., e_n\}$ (to recall: these influences can be either other agents, or extra-systemic influences), then state of such formed system at any particular moment in time can be defined as a Cartesian product $S \subseteq P \times E$. Since our system is not a static graph, but a dynamical network of interactions where at each moment hierarchical status of network elements can be significantly changed, we have to define state of the population P as a mapping $\omega : e \to p \,|\, e \in E, p \in P$, where both e and p are defined as temporal sequences of events, such that $E = \{e : T \to I\}$ and $P = \{p : T \to R\}$, where T is a set of time points t, I is a set of external stimuli on a particular agent such that at each time point system receive stimulus $i(t)$ and R is a set of appropriate responses, $r(t)$. Further, both P and E are formal systems that stands for themselves. Therefore, occurrence of p and occurrence of e at some particular time point t are governed not only by mapping ω but also by internal rules of these systems, which are partially (see further text) independent. Structure of these systems applied here in creating computing architecture will be explained later.

From above stated, it is obvious that changes in environment induce appropriate responses in agents through the model of coupled input/output pairs. In real systems, reverse situation is also possible, such that some external changes can be influenced by the activity of organisms but for the sake of simplicity, and since it is not directly connected with the main topic of this paper, here we will not consider that problem. It is clear that critical point in

building an evolvable model as described above, is in choosing appropriate structure of the mapping $I \to R$. When dealing with usual models developed as a prediction tools, it is enough to take an attitude of analyzing a "black box". In that case we can propose a function which should sum up available experimental data, and obtain a set of more or less accurate predictions for various initial conditions. However, in that case we will neglect a real meaning of the nature of mappings within E and P. Taking a slightly closer look at these relations, and having in mind a notion of "operative assimilations", we can see that somewhat hidden problem is how I is generated from the wholeness of external changes, and what is a connection between generating I, with constitution of corresponding R. Although this connection can be efficiently represented using a segment of applied lattice theory called Formal Concept Analysis [Crvenkovic et al., 2007], its evolvability demands more advanced formal treatment in order to be fully comprehended.

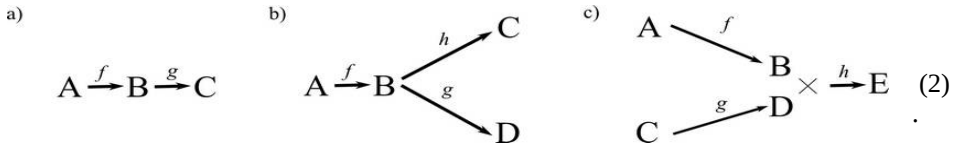
Firstly, we should more strictly define what could be considered as external environment. Let us define set $A = \{a_1, a_2, a_3, ..., a_n\}$ as a set of all external changes. In order to functionally treat them, agent needs to associate each observable external change with some attribute. Hence, we define set of attributes $M = \{m_1, m_2, m_3 ... m_n\}$. From the Formal Concept Analysis [Wille, 1982] we could apply the rule which states that objects and attributes are mutually defined, so we automatically obtain set of external influences $E \subseteq A : \{\forall g \in E \,|\, (g, m) \in \mathrm{O}\}$, where O is a binary relation between M and E. In other words, every object (attribute) which does not form a binary relation with corresponding attribute (object), simply does not exist from the perspective of particular agent. In that way, the concept of "whole" environment changes its epistemological status and is reduced to an observable environment, i.e. to a set of treatable objects. Also, since E is a derivate of p, mapping $I \to R$ can now be realized only as an indirect derivate of each p.

Secondly, if we put additional attention on mapping ω, we can see that from the perspective of both formal systems E and P, it is unentailed. It does not belong to neither formalism, and accordingly cannot be derived from anything within these formalisms. We have in hand two possible solutions to that problems. On the one side, we can leave it as it is. In that case, appropriate function for the focal problem can be applied and the model will be valid until field conditions change out of defined boundaries. If that happens, we have to create a new formal system by choosing another function or/and by modifying organization of the whole system. However, on the other side, we can incorporate ω into another formalization that deals with comparison of formalisms [Rosen, 1991]. This discipline is called category theory and is developed by McLane [1971]. Within such formalism, if domain of one mapping intersects the range of another, they can be composed into:



(1)

where by capital letters are denoted objects within formal systems, while by small letters are denoted corresponding mappings between objects. For aforementioned case of two systems E and P, following cases are valid:



(2)

Accordingly, if we properly transfer them into domain of the category theory, situations denoted by b and c in (2) can be collapsed into $\varphi : A \to C \times D$, where $\varphi(a) = \big(f \circ g(a), f \circ h(a)\big)$, and $\psi : A \times C \to E$, where $\psi(a,c) = \big(h(f \times g)(a,c)\big)$

respectively [Rosen, 1991]. In other words, regardless of complicity of interacting pathways, they can be reduced into corresponding simpler mappings. In a similar manner, we can perform mappings from one category to another that preserves the categorical structure (being an object, being a morphism, the composition, and the identities). Such mapping is performed by functors. Therefore, if we have categories ✌ and ✍, then the functor F from ✌ to ✍ is a mapping that sends objects from ✌ to objects of ✍, and mappings from ✌ to mappings of ✍ preserving identities and composition. Going back to the context of evolvable biological interfaces, it means that within proposed formal framework we finally came to the stated precondition in order to construct open evolvability; namely: there cannot exist rule that is valid for external environment, and non-valid for the system itself. In other words, structure of mapping ω is construed so that no matter how I is changed as a result of dynamics within the E, there can be constructed response R given only a set of rules within the P.

Since further development of rigorous mathematical treatment of stated problem, and more generally, development of mathematical theory of evolvable biological systems is out of scope of this paper, here we are leaving strict mathematical derivations, and continue with presenting a new computing system, based on outlined approach, and using individual-based modeling strategy, which should demonstrate practical possibilities of outlined conceptualization.

2.3 Structure of individuals

Following above described strategy, we introduce two major differences to usual Learning Classifier System (LCS) architecture. Firstly, instead separating input, action-rules, output parts of an individual, we merged them into one unified interface which structure is organized so that recognition of external signal automatically means performing an action defined by the structure of the individual itself. In that way, mimicking most elementary level of functioning of living systems (molecular level of proteins), we fully applied notion of operative assimilation. By taking this strategy, we didn't imply pre-existence of any developed regulatory or operative mechanisms, as it is a case in LCS architecture. Also, in our model it is not possible to arbitrarily include, change or exclude operative rules, without any consequences for (and any inherent connection with) input/output interfaces. Instead, our model is organized so that in order to be functional on a higher level (not only on a level of local actions) it should self-organize after a period of adaptive evolution. It could look like a disadvantage, but again, such opinion is a matter of perspective and chosen framework of consideration.

Secondly, in order to incorporate such strategy, instead of unambiguous symbols in above-mentioned ternary alphabet including one wildcard, we introduced alphabet where each symbol is characterized by a set of characteristics.

As it is already mentioned, this model is based on reflecting level of proteins and pattern of their structure and mode of functioning. Since each protein is characterized by primary, secondary and tertiary structure, we also introduced a gap between primary structure of an agent and its final, operative structure. Primary structure is determined as a string composed of elements from the basic alphabet Ω (Table 1). We construed Ω following empirical data of properties of proteins' basic building blocks. These data can be further abstracted, but we left them in original form in order to make simulation results more obvious for possible applications in a domain of environmental modeling. Each element of alphabet is represented as ordered tuple characterized by its identification (ID) symbol and six properties:

$$\Lambda = (\lambda[+,-,/], \lambda[+,-,/], \aleph[0,1,2,3], \aleph[0,1,...,5], \aleph[0,1,...,14], \lambda[Am, Ca, Al, Ke, Th, /]) \qquad (3)$$

where Λ denotes any ID symbol, $\aleph$ denotes any real number in a given range indicated by square brackets, while λ denotes any symbol from the corresponding sub-alphabet indicated within square brackets. Properties of all elements of Ω are stored in a database. During all processes, desired property of Λ has been called by indicating appropriate place in the tuple, such that $\Lambda[0]$ indicate first place $\Lambda[1]$ second, and so on.

Table 1. Composition of the basic alphabet Ω. One letter ID symbols, charge, polarity, influence radius and existence of functional group are taken in accordance to Whitford [2005] so that for indicating charge and polarity we preserve original nomenclature, while for influence radius we applied Van der Waal index and normalized it for a real number scale in a range [0-3]. For functional groups we take first two letters of their names. Hydrophobicity index and pKa values are taken from Copeland [2000]. Hydrophobicity index is normalized for interval [0-5], while for pKa we preserved original values, rounded to nearest real number.

ID	Charge	Polarity	Hydrophobicity index	Influence radius	pKa	Functional group
R	+	/	**0**	**3**	**12**	***Am***
H	+	/	**0**	**2**	**6**	***Am***
K	+	/	**0**	**3**	**11**	***Am***
D	-	/	**0**	**2**	**4**	***Ca***
E	-	/	**0**	**2**	**4**	***Ca***
S	/	**δ-**	**0**	**2**	/	***Al***
T	/	**δ-**	**0**	**2**	/	***Al***
N	/	**δ+**	**0**	**2**	/	***Ke***
Q	/	**δ+**	**0**	**2**	/	***Ke***
Y	/	**δ-**	**0**	**3**	**10**	***Al***
G	/	/	**0**	**1**	/	/
C	/	/	**3**	**2**	**8**	***Th***
P	/	/	**0**	**2**	/	/
A	/	/	**2**	**1**	/	/
I	/	/	**5**	**3**	/	/
L	/	/	**4**	**3**	/	/
M	/	/	**2**	**3**	/	/
F	/	/	**3**	**3**	/	/
W	/	/	**0**	**3**	/	/
V	/	/	**4**	**2**	/	/

Initially, each individual is obtained as a randomly generated string Ψ, with variable length between 10-20 ID symbols from Ω. Such primary structure is then evaluated by decentralized rule-based algorithm with nested levels of evaluation, in order to determine its functional, secondary structure. Outline is given by following pseudocode:

FOR Λ IN Ψ:

IF $\Lambda_n[x] == \Lambda_{n+1}[x]$ THEN $(\Lambda_n = \Lambda_n^{t_x}) \wedge (\Lambda_{n+1} = \Lambda_{n+1}^{t_x})$

FOR Λ^{t_x} IN Ψ:

IF $(\Lambda_{n+1}^{t_x} == \Lambda) \wedge (\Lambda_{n+2}^{t_x} == \Lambda^{t_x})$ THEN $(\Lambda^{t_x} == \Lambda^{t_x^2}) \wedge (\Lambda == \Lambda^{t_x^2})$

FOR $\Lambda^{t_x^2}$:

IF $x == 2 \wedge (\Lambda^{t_x^2}[0] \vee \Lambda^{t_x^2}[1] == \neg/)$ THEN $\Lambda^{t_x^2} = \Lambda^{t_x^3}$

Here, n denotes sequence number of Λ in Ψ, x denotes place in the tuple for Λ in a range [0-2], t_x is a first-level grouping tag, t_x^2 and t_x^3 are second- and third-level grouping tags, respectively. As it can be seen, here we applied stepwise top-down approach in evaluating primary structure. First step is formation of "groups" in accordance to selected properties of Λ. For a given property, group is formed if two or more Λ with that property are in a sequence. In a second step, "groups" are evaluated and if they are separated only by

a single Λ, they are grouped together into the second-level groups. Finally, third-level of groups are formed if some "charge/polar" Λ occurred within "hydrophobic" group. All remaining second-level groups where number of $\Lambda \geq 3$ are marked as "active places".

After establishing secondary structure for an individual, it obtains a new ID tag Θ and its new properties are stored into a database. Finally, it has been released in artificial environment characterized by following structure:

$$\Theta = (\aleph[0-x], \aleph[0-y], \aleph[0-z], \aleph[0-5], \aleph[0,1]) \quad (4)$$

where $\aleph$ denotes any real number in a given range indicated by square brackets. First three places in the tuple indicate spatial dimensions, and *x*,*y*,*z* are determined by the size of the environment, fourth place indicate scale of instability (see section 2.3.1), while current state (active/inactive) is determined by the last place in the tuple. At each time step, one of three spatial dimensions have been randomly changed by 1, while active/inactive fluctuations are performed in accordance to the level instability indicated by $\Theta[3]$. If value is 5, probability of being active is up to 1/10, while for the value of 0 in the scale, probability of being active is defined in the range above 9/10 of total time spent in the environment.

2.3.1 Stability of individuals

Since in empirical world, proteins can become unstable if their primary structure is altered, resulting in partial or full loss of functionality, we introduced that property by making third level groups. During simulation running, total number of third-level groups for a given Θ has been calculated and normalized for a real number "instability" scale in a range [0-5]. Larger the value in the scale, Θ will spend more time as inactive. In that way, on a long run and in competitive environment, some regions of primary structure will become conserved, thus alleviating preserving of evolutionary achieved level of functionality.

2.3.2 Operative assimilation performed by individuals

All steps considered under the term operative assimilation, are defined to be performable only by an individual's active place(s). Within active place each Λ is characterized by "index of substrate affinity" which is determined in accordance to $\Lambda[3]$ property. Two Λ in boundary regions (up to three places from start/end of active place) with greatest affinity values have been marked as potential sites for binding. When encountering virtual substance, its structure (see section 2.4) is compared with properties of binding sites ($\Lambda[0,1,2]$) and if achieved complementarity, appropriate operation on that substance is launched. Such operation is not defined as a separate list of available actions, but is determined by the composition of a region between active sites. For each Λ in that region values for $\Lambda[4]$ and $\Lambda[5]$ are compared with structurally corresponding indexes in the substance and if complementary matching is achieved, properties of substance's elements are evaluated calling rules from separate database (see section 2.4).

2.3.3 Evolution of individuals

For presented simulation we introduced only single point mutations, deletions and insertions. More advanced possibilities for changing primary structures, like crossing-over are not included for the sake of simplicity, but overall architecture is such that they can be easily incorporated. Given relatively short strings Ψ, and in order to alleviate evolutionary process, we set mutation frequency at $1/10\,\Lambda$/generation. At each generation time, Ψ have been duplicated so that type of mutation is randomly chosen and in accordance to defined frequency an element of Ω have been replaced, inserted to or deleted from Ψ. It should be strongly emphasized that in a given architecture, single point mutation can have no consequences at all or can lead to complete loss/change of functionality of an individual, depending on a relative position in the primary structure. This is a completely new property in the domain of individual-based models.

2.4 Artificial environment

Each virtual substance is described in accordance to real chemical species and similarly as Λ, they possess various characteristics: charge, polarity, hydrophobicity, pKa and functional groups. Same as with constructing individuals their structure is determined as a string composed of elements from the basic alphabet $\mathrm{H} = \{H, C, N, O, Na, Mg, P, S, Cl, Ca, K\}$. As it is obvious, for elements of H we utilized most common chemical elements observable in living systems. Each element here have its set of properties (charge, number of bonds which it can make and electronegativity) which are stored in separate database. Based on that we also construed small generator (not described here) of substances which have been introduced into the virtual environment, based on given atomic properties. Resulting substance is characterized by:

$$\Phi = (\lambda[+,-], \aleph[0-5], \aleph[1-4], \lambda[Am, Ca, /]) \quad (5)$$

where Φ denotes assigned ID symbol, $\aleph$ denotes any real number in a given range indicated by square brackets, while λ denotes any symbol from the corresponding sub-alphabet indicated within square brackets. Substances do not have secondary structure, but they can be decomposed or merged together under appropriate conditions. For presented simulated model, on the side of virtual substances we introduced only two functional groups, but overall architecture is organized so that arbitrary number of functional groups can be included, each with its own reaction patterns, as long as they satisfy condition for mapping ω. Of course it will demand more computational power but having in mind that in nature exists only around 10 life-significant chemical functional groups, great complexity can be achieved even utilizing relatively small number of functional groups. What is major difference from elements of Ω is the lack of individual properties when composed in unaltered substance. If some operation by individual have been performed on them (decomposition, transfer of functional groups…) only then individual properties will be visible on Φ by changing corresponding values in the tuple.

Common space for individuals and substances is composed as a 3D grid denoted by U, containing a number of grid units $u_{x,y,z} \in U$ where x, y, z are space coordinates. Each individual Θ and each substance Φ can occupy only one grid unit at some particular time point. Their movement is defined by random-walk algorithm. If encounter each other, and if individual is in active state, recognition/action process is started following procedure described in 2.3.2.

3. SIMULATION RESULTS

In order to demonstrate general pattern of behavior of the model, we run two kind of simulation. First, in a common space with dimensions 100x100x100 we injected 20 randomly generated individuals and 10 different, randomly generated substances, each in 5 copies, and run a model for 1000 time steps. Results can be seen in the Table 2.

Table 2. Overall results for 1000 steps run in the common space including 20 individuals and 50 substances.

Number of individuals	Number of substances	Number of individual-substance encounters	Number of tried assimilations	Number of successful operative assimilations
20	50	3842	136	0

As it can be seen, very small number of individuals have been in active state during encounters (indicated by the number of tried assimilations) which reveal their very unstable secondary structure. As it was expected, none of external substances was recognized. After that, we take these 20 individuals and run them through evolutionary process for 100 generations. After each generation, we repeated first experiment. Results for five of these individuals, regarding number of successful operative assimilations can be seen in Figure 1.

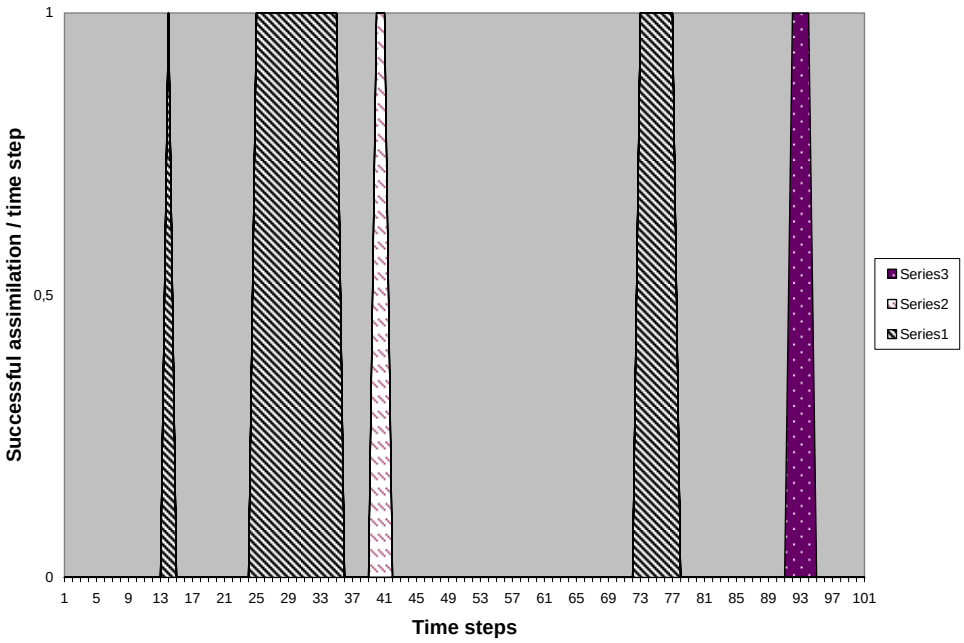


Figure 1. Number of successful operative assimilations of three chosen individuals during 100 time steps

In showed case three of 20 individuals, during evolutionary process successfully assimilate some of external substances. Since competitions are not included at this stage, there were no favouritism of successful individuals, and hence, due to further mutations, successful assimilations were only observed transiently. Nevertheless, it is a clear example of possibilities of proposed computational architecture.

4. CONCLUSIONS

We have presented a novel approach to modeling of evolvable biological interfaces. It is based on the notion of operative assimilations inspired by mechanisms which living systems use in interaction with their environment. Practical implementation relied on paradigm of learning classifier systems but with two key modifications: merging of input-action-output system into one unified interdependent whole where each element's role is relative to its surrounding, and instead of unambiguous symbols we introduced alphabet where each symbol is characterized by a set of characteristics. As a result, we obtained system which is able to eventually, through evolution, recognize and assimilate unlimited number of new substances and to perform on them non-fixed set of transformations.

ACKNOWLEDGEMENTS

The research work described here has been funded by the Serbian Ministry of Science and Technology under the project "Modelling and numerical simulations of complex physical systems", No. ON141035 for 2006-2010.

REFERENCES

Amils, R., Ellis-Evans, C., and Hinghofer-Szalkay, H., *Life in Extreme Environments*, Springer, New York, pp. 450, 2007.

Atkin, A., *Peirce's Theory of Signs*, In The Stanford Encyclopedia of Philosophy (ed. Zalta, E.N.), URL = http://plato.stanford.edu/entries/peirce-semiotics/, 2006

Bull, L., *Applications of Learning Classifier Systems.* Springer, New York, pp. 315, 2004

Copeland, R.A., Enzymes-A Practical Introduction to Structure, Mechanism and Data Analysis, Wiley-VCH, New York, pp. 397, 2000.

Crvenkovic, S., Mihailovic, D.T., and Balaz, I., Use of Formal Concept Analysis for Construction of Subjective Interface Between Biological Systems and Their Environment, In Environmental, Health and Humanity Issues in Down Danubian

Region: Multidisciplinary approach (eds. D.T., Mihailovic and M. Miloradov), World Scientific, London, Singapore (in press)

de Saussure, F., *Course in General Linguistics*, McGraw-Hill, New York, pp. 240, 1965

Doelman, A., Van Harten, A., *Nonlinear Dynamics and Pattern Formation in the Natural Environment*. Chapman & Hall/CRC, London, 344 pp. 1995.

Gualtieri, C., Mihailovic, D.T. (eds.), *Fluid mechanics of environmental interfaces.* Taylor & Francis Group, London, 339 pp. 2008.

Hoffmeyer, J., Biosemiotics: Towards a New Synthesis in Biology, European Journal for Semiotic Studies, 9(2), 355-376, 1997.

Holland, J.H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, artificial intelligence*, The MIT Press, pp. 228, 1992

Kirschner, M., and Gerhart J., Evolvability, *Proc. Natl. Acad. Sci. USA*, 95, 8420-8427, 1998

May, R.M., and McLean, A.R., *Theoretical Ecology: Principles and Applications*, Oxford University Press Inc., New York, pp. 257, 2007.

Mclane, S., *Categories for the Working Mathematician*, Springer Werlag, pp. 272, 1971

Mihailovic, D.T, and Balaž, I., An essay about modelling problems of complex systems in environmental fluid mechanics, *Idojaras*, 111(2-3), 209-220, 2007.

Mesarovic, M.D., *A Mathematical Theory of General Systems*. In Trends in General Systems Theory (ed. Klir, G.J.), Willey-Interscience, New York, pp. 462, 1972.

Rosen, R., *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*, Columbia University Press, New York, pp. 285, 1991.

Whitford, D., *Proteins – Structure and function*, Willey & Sons, West Sussex, pp. 543, 2005

Wille, R., *"Restructuring lattice theory: an approach based on hierarchies of concepts"*, In Ordered Sets (ed. Rival, I.), Dordrecht-Boston, Reidel, pp. 445-470, 1982.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

The Impact of Geometrical Effects on Calculated Value of Albedo on the Heterogeneous Environmental Interfaces

D.V. Kapor[a], **A. M. Cirisan**[b] **and D.T. Mihailovic**[c]

[a] *Department of Physics, Faculty of Sciences, University of Novi Sad, Dositej Obradovic Sq. 4, 21000 Novi Sad, Serbia (dvk@if.ns.ac.yu)*

[b] *Institute of Physics, Pregrevica 118, 11080 Zemun, Serbia (anacir@polj.ns.ac.yu)*

[c] *Faculty of Agriculture, University of Novi Sad, Dositej Obradovic Sq. 8, 21000 Novi Sad, Serbia (guto@polj.ns.ac.yu)*

Abstract: The albedo of the interface is an important parameter for the calculation of the radiation fluxes in environmental studies. The problem arises when the surface is a heterogenous one. There have been developed various manners to calculate the aggregated albedo. Our previous research has indicated, however, existence of a geometrical effect related to the fact that different parts of the interface may have different heights. In these studies we have offered the general approach to the calculation of the flux that is lost due to the absorption on the vertical lateral boundaries. We derived analytically the expressions for this loss coefficient which for some ideal urban geometries coincides with Oke's sky-view factor. The aim of this paper is to elaborate the effect for more complex geometry. The number of geometries allowing easy analytic solutions is rather restricted, so it was necessary to develop an efficient numerical procedure. In this paper we expose so called ray-trace MC approach, then we use it to test the known analytical solutions and combine it with a parameterisation scheme (LAPS). An example of central geometry is considered in this study. It is assumed that the region consists of two-patch grid-cell with a square geometrical distribution and different heights of its parts. Simulations were done for several patch areas, with different horizontal dimensions and heights ("propagating building") as well as for different surface types. The multiple scattering effect and the dependence of the albedo on the zenithal angle of the incident radiation were neglected. The derived expression for the albedo is compared with the conventional approach. Changes in albedo lead to a significant change in partitioning of the energy at environmental interface. The most remarkable changes are in values of sensible and latent heat fluxes, and surface temperature. Their values are calculated using LAPS parameterisation scheme and then compared to the values obtained with a conventional parameterisation of the albedo.

Keywords: Albedo; Environmental interface; Parameter aggregation; Urban modelling.

1. INTRODUCTION

Surface albedo is a fundamentally important meteorological parameter. Through its influence on the energy balance of the Earth, and thus indirectly on many other factors (fluxes of momentum, energy, carbon-dioxide, moisture etc.), albedo became a key variable in the parameterization of the land-surface radiative transfer over the grid cell in numerical modelling.

Nowadays, one of the basic tasks in order to obtain more precise parameterization of physical processes is determination of albedo over heterogeneous areas. There have been developed various manners to calculate the aggregated albedo [Dickinson, 1983, Delage et al., 1999, Mihailovic et al., 2001]. The most common approach usually used is a simple arithmetical averaging to determine the albedo as the grid cell-average albedo. According to physics based analysis, a significant deviation of the albedo is observed from that calculated by simple averaging.

In our previous work shown in Kapor et al. [2002] and Mihailovic et al. [2004] we have proposed a new method for aggregating the albedo over a very heterogeneous surface in land surface schemes for use in environmental models. The basic assumption of our research is that geometrical factor plays an important role in albedo calculation. We have demonstrated that even the same material at different heights might produce interesting effects.

After a brief description of our previous work, in this paper we will present a simulation aimed to illustrate our further work, concentrated on elaboration of this effect using more complex geometries.

2. SHORT REVIEW OF PHYSICAL BASIS OF AGGREGATION

2. 1 Introduction (basic assumptions)

In our aforementioned studies we have proposed a new approach for aggregating the albedo over a very heterogeneous surface. For the simplicity we analysed two-patch grid-cell with a simple geometrical distribution and different heights of its components. We supposed that the basic constituent of the albedo, coming from the grid-cell, describes the diffuse, homogeneous single scattering of incoming radiation from a given surface. The multiple scattering effect and the dependence of the albedo on the zenithal angle of the incident radiation were also neglected.

Our main idea was relied upon the fact that a part of radiation reflected from the lower surface is completely absorbed by the lateral sides of the surface lying on a higher level. We introduced a “loss coefficient”, flux that is completely lost due to the absorption on the vertical lateral boundaries. Its definition is conceptually analogous with Oke’s sky-view factor [Oke, 1987], especially for some ideal urban geometries. The derived expression for the albedo of the particularly designed grid-cell is compared with the conventional approach, using a common parameterization of albedo over the same grid-cell as used by Delage et al. [1999].

2.2 Analytical treatment

The average albedo over the grid-cell divided into two subregions with different albedos, according to conventional approach is given as

$$\overline{\alpha}_c = \alpha_1\sigma_1 + \alpha_2\sigma_2 \tag{1}$$

while our analytic treatment for calculating the average albedo over this grid is

$$\overline{\alpha}_n = (1-k)\alpha_1\sigma_1 + \alpha_2\sigma_2 \tag{2}$$

α_i is corresponding albedo of particular patch. σ_i is fractional cover, calculated as a ratio of patch’s area S_i and total grid cell area $S\left(\sigma_i = S_i / S, i = 1,2\right)$. k is before mentioned loss coefficient, which can be derived from the following relation

$$k = \frac{\left(\frac{dE}{dt}\right)_l}{\left(\frac{dE}{dt}\right)_h} \tag{3}$$

The denominator shows the amount of flux emitted from the lower surface into the upper space, while the numerator is a part of total energy coming from lower horizontal surface towards lateral surface. The expression to calculate the radiant energy flux dE/dt was taken following Liou [2002].

The amount of emitted flux reaching the vertical lateral boundary is defined as a sum of all infinitesimal amounts of radiant flux emitted from the infinitesimal surface element $dxdy$ (centred around the point with position vector $\vec{r}$) confined in the solid angle $d\Omega$ under which the element $dxdy$ „sees" the lateral surface, shown shaded in Figures 1 and 2. Since the flux of radiation that reaches the vertical boundary surface is completely lost, according to our basic assumptions, the contribution of the radiation reflected from the lateral surface is not taken into account of the total reflected flux of radiation.

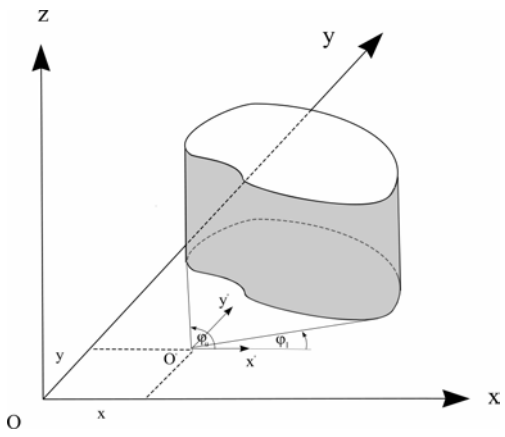


Figure 1. Definition of the boundaries for the integration over the local azimuthal angle

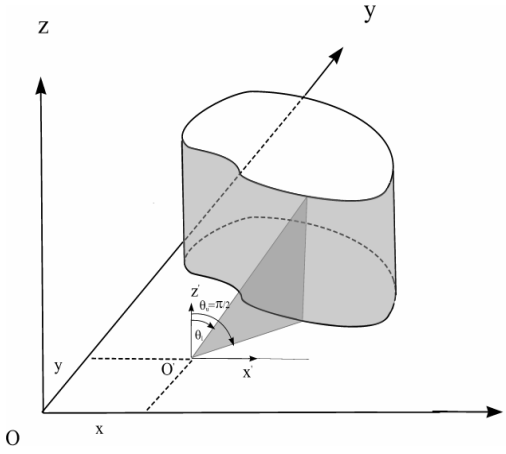


Figure 2. Definition of the boundaries for the integration over the local zenithal angle

The boundaries of the integration for a given point are determined over the azimuthal (φ_l, φ_u) and zenithal (θ_l, θ_u) angles in terms of the x, y coordinates, as it is shown in Figures 1 and 2. The subscripts l and u denote lower and upper boundary, respectively. Accordingly, the equation that calculates a part of total energy coming from lower horizontal surface towards lateral surface is given in the following form

$$\left(\frac{dE}{dt}\right)_l = I\iint_S dxdy\int_{\varphi_l(\vec{r})}^{\varphi_u(\vec{r})} d\varphi\int_{\theta_l(\vec{r},\varphi)}^{\theta_u(\vec{r},\varphi)} \cos\theta\sin\theta d\theta \qquad (4)$$

I is the total intensity of radiation obtained from the monochromatic intensity by integrating it in the range of the whole spectrum. $\cos\theta$ describes the direction of the radiation stream, while $\sin\theta d\theta d\varphi$ is the element of solid angle within which our differential amount of energy is confined to. The amount of flux emitted from the lower horizontal surface into the upper space is calculated as $(dE/dt)_h = IS_l\pi$. The loss coefficient k we need for calculating the average albedo can be evaluated combining Equations (3) and (4).

2.3 Numerical approach

Beside analytical treatment, we expose a version of Monte Carlo approach. Since the number of simple geometries which allow easy analytic solutions is rather restricted, it was necessary to develop an efficient numerical procedure for calculating the loss coefficient. We made a direct system simulation using a well known „ray tracing" MC method.

The essence of MC „ray tracing" method is to retrace a ray destiny. The light transport equation is solved using MC methods to simulate the propagation of light in a scenery. Monte Carlo simulation is a method for iteratively evaluating a deterministic model using

sets of random numbers as inputs. By using random inputs, the deterministic model is turned into a stochastic one. Result is obtained in the limits of large number of numerical experiments.

As we mentioned, the main idea is to follow the appropriately chosen ray of light, after it had undergone diffuse, homogeneous single scattering from the lower surface of the grid-cell. Averaging the observed behaviour over a large number of followed light paths, we determine the value of "loss-coefficient" k within a given grid-cell geometry.

More precisely, we guide our numerical experiment as follows. We generate first two random numbers r_1 and r_2 uniformly distributed in the interval (0,1) in order to randomly sample the point $A(x, y)$ which belongs to the lower surface and represents a point of intercept of this surface and the incoming beam. Then we choose a random direction in upper half-space (θ, φ) [with $\theta \in (0, \pi/2)$; $\varphi \in (0, 2\pi)$] to simulate trace of scattered beam. The case was positive for absorption if diffusively scattered beam would reach the vertical boundary. This procedure was repeated $N = 10^6$ times and the "loss-coefficient" was estimated as

$$k = N_a / N \tag{5}$$

where N_a is number of cases which were positive for absorption and N is number of conducted numerical experiments.

The first test of the method was to reproduce the analytical results for the simplest geometry ("step-like"patch) Kapor et al. [2002], Mihailovic et al. [2004]. After successful reproduction for various parameters of the patches, we concluded that the MC ray-tracing method application is justified in this case.

3. NUMERICAL TEST

In this study we have considered an example of central geometry. For the simplicity, it is assumed that the region consists of two-patch grid-cell, a grid-cell of size L and „propagating building" with the edge l and height h, taking a different values, simulating the propagation of that area over the grid-cell. The basis of building is a square centred at the center of cell.

In our study the dimension of a grid-cell is $100 \times 100\, m$. The fractional cover of the central area is $\sigma = (l/L)^2$, while for the rest of the cell is $1 - (l/L)^2$. Calculations of albedo were done using reduced dimensionless quantities $\hat{l}$, defined as l/L ratio and $\hat{h}$, defined as h/L ratio. Depending on the variations of the reduced length $\hat{l}$, we have calculated values of average albedo according to both, conventional and our approach. These values were utilized then as average values of albedo in prognostic equation for evolution of the surface temperature over heterogeneous grid-cell. The mentioned prognostic equation is

$$C_g \frac{\partial T}{\partial t} = R_{net} - H - \lambda E - G \tag{6}$$

where T is surface temperature [K], t is time [T], C_g is effective heat capacity (heat capacity per unit area) [M/KT^2], R_{net} is net radiation [M/T^3], H is sensible heat flux [M/T^3], λE is latent heat flux[M/T^3], and G is soil heat flux [M/T^3]. (Here, standard meteorological terminology is used. Actually, using the correct terminology, all these are not fluxes, but flux densities). Net radiation, as a sum of incoming and outgoing radiation, could be determined through the following terms

$$R_{net} = (1 - \alpha) R_{short} + R_{long} \tag{7}$$

where R_{short} is a shortwave radiation [M/T^3] and R_{long} is a total longwave radiation [M/T^3]. The surface fluxes aggregation approach was applied to obtain the mean surface fluxes over heterogeneous grid-cell, while fluxes over particular area were obtained from

LAPS (Land Air Parameterization Scheme). This land surface scheme was designed to be run either as a standalone model or as the part of an atmospheric model, Mihailovic et al. [2001]. The analyzed period in our case was one day, with time step of 10 min.

Simulations were done for several patch areas, with different surface types. Accordingly, each of the subregions had different and corresponding albedos. Three situations were analyzed. In all of them, the central area was concrete or rocky land, having the same albedo value 0.30. The other patch of grid-cell, in these three simulations, was covered with grass, orchard and concrete, having albedo values 0.20, 0.15 and 0.30, respectively.

The albedo was calculated taking the following values of the reduced lengths 0.50, 0.70 and 0.870. We wanted to analyze daily temperature profile differences obtained by both approaches in a case of grid-cell covered with each type of surface with 25%, 50% or 75%. Altitude of a central patch was case sensible. In the grass-concrete situation, concrete was 2m high, while grass was 0.5m. Results are shown in Figure 3. In the case of orchard-concrete grid-cell, orchard is taken to be 2.5m high, while concrete 10m. The same height of a central area concrete is applied when both patches were concrete covered. Figures 4 and 5 show the obtained results, respectively.

As we can see from the exposed graphics, aggregated approach in every situation gives a higher daily maximum of surface temperature. The significant difference between temperature profiles obtained by aggregated and conventional approach is noticed in a case of orchard-concrete situation and especially in concrete-concrete case. Temperature depends on albedo value, which has the largest deviation when central area occupies 75% of the grid, since in that case the "loss-coefficient" has the largest value. Figure 6 shows how "loss-coefficient" varies with different values of reduced length and height. Although this geometry was not that simple, we made an analytical calculation, and as we can see in Figures 3-5, the Monte Carlo simulation reproduced the obtained results quite well, which gave a reliable base for further research in a case of more complicated geometry.

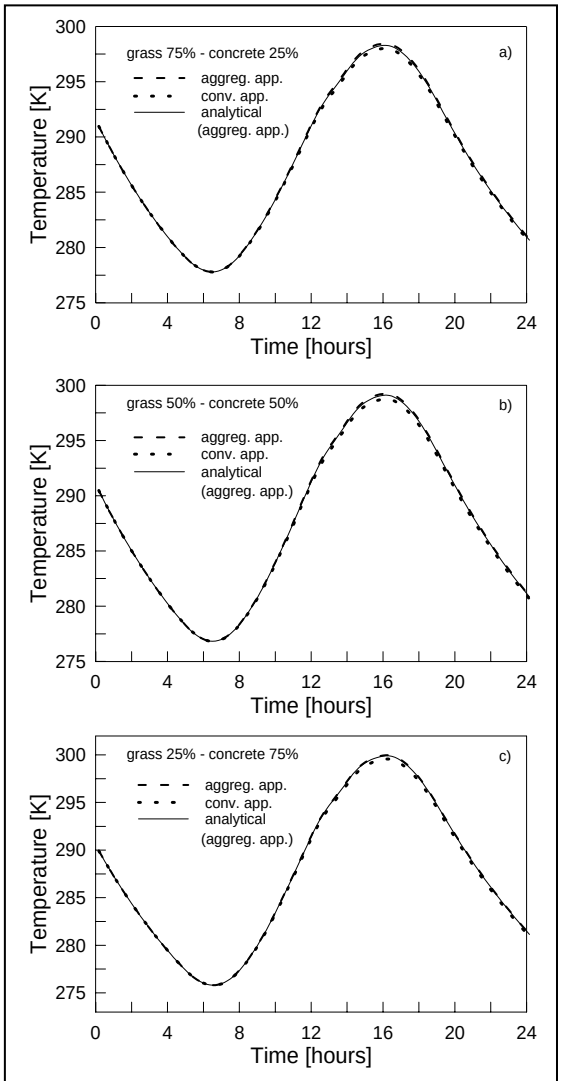


Figure 3. Daily profile of surface temperature in a case of grass-concrete covered grid-cell, with a) grass 75%, b) concrete 50%**,** c) concrete 75% using both approaches

In order to define more precisely the difference between temperature profiles obtained by aggregated and conventional approach, the root mean square error (RMSE) is found. From results depicted in Figure 7 we can see that maximum increase in the temperature and decrease in albedo is for 50-50 area coverage when both patch areas are covered with concrete. In orchard-concrete situation we also obtained a significant RMSE, especially for case when central area occupies 25% of cell.

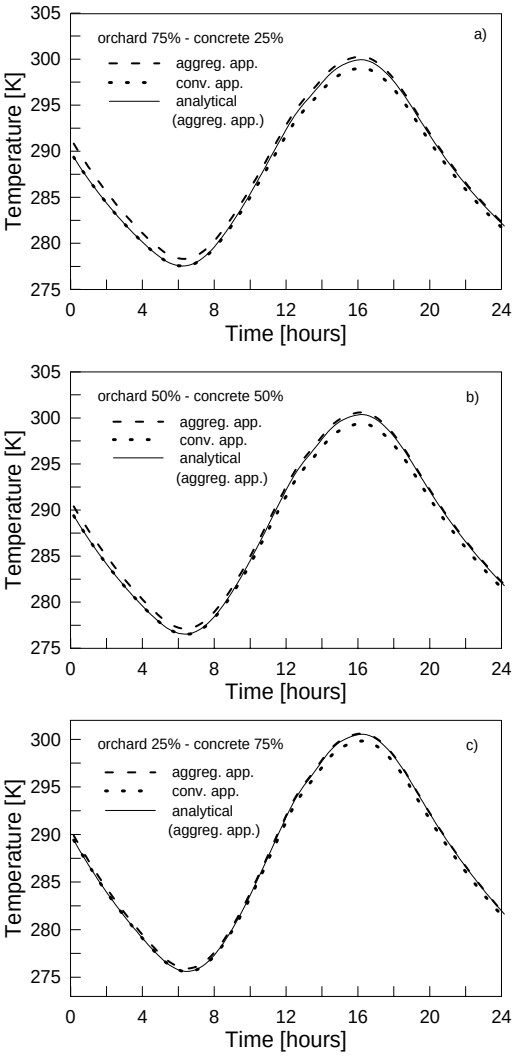


Figure 4. Daily profile of surface temperature in a case of orchard-concrete covered grid-cell, with a) orchard 75%, b) concrete 50%**,** c) concrete 75% using both approaches

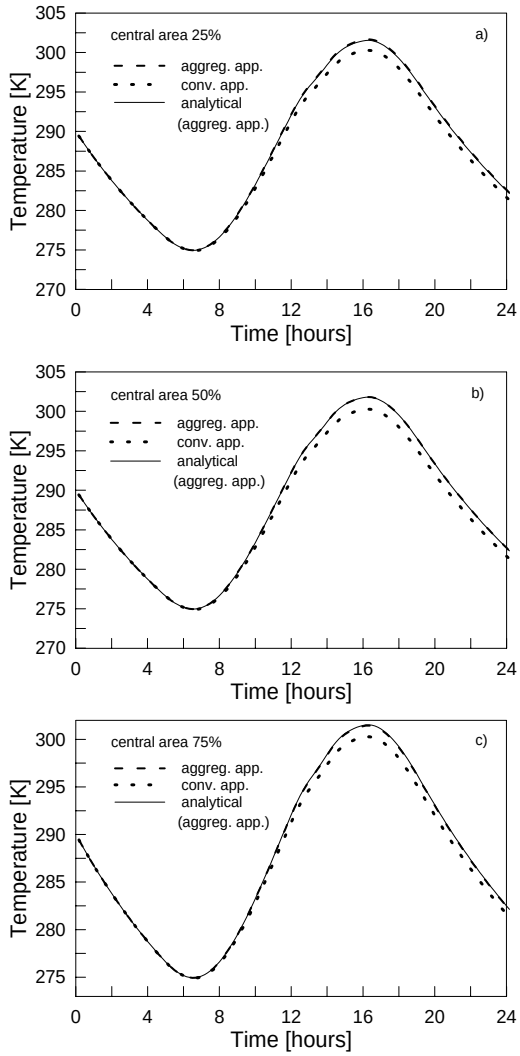


Figure 5. Daily profile of surface temperature in a case of concrete-concrete covered grid-cell, with a) central area 25%, b) central area 50%**,** c) central area 75% using both

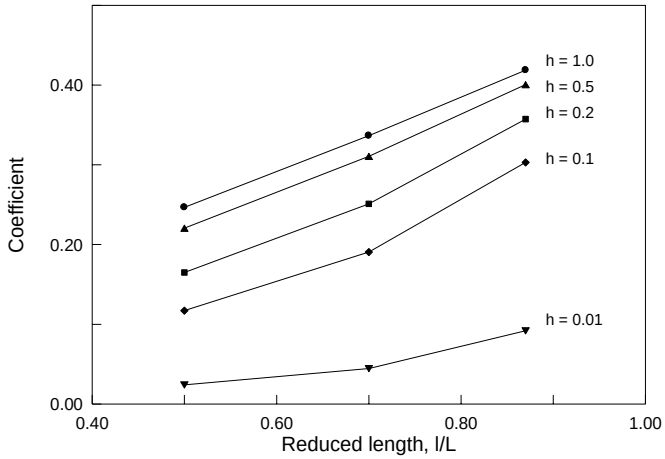


Figure 6. Dependence of "loss coefficient" on reduced length l/L and height h/L

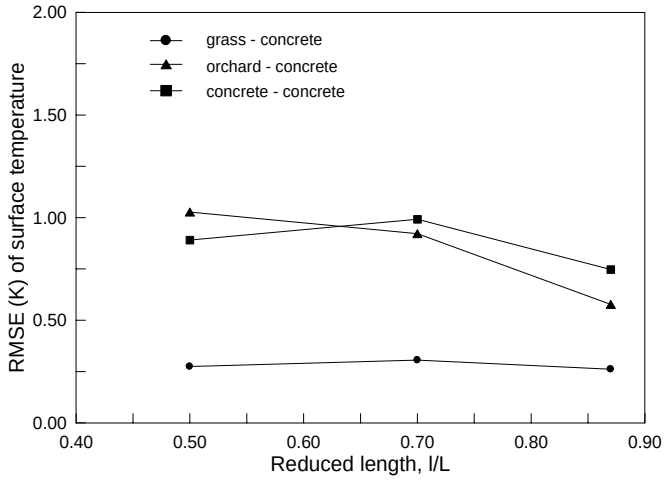


Figure 7. Dependence of RMSE of surface temperature on the reduced length l/L

4. CONCLUSIONS

The aim of this study was to show further achievements related to the new approach for the aggregation of albedo over the heterogeneous grid-cell. We have elaborated the effect for a central geometry for the most general case using only two patches distinguished by relative height of the central one. Simulations were done for different surface types, with different horizontal dimensions and heights.

After a short review of the general approach to the calculation of the flux that is lost due to the absorption on the vertical lateral boundaries, a theory was applied to a simple model of square patches. Because of a rather cumbersome analytical solution, we exposed an efficient numerical procedure, using Monte Carlo ray tracing method, for albedo calculations. Results were compared with the results of the conventional method, but also with the analytical ones, and were applied in a calculation of surface temperature over the heterogeneous area. Sensitivity tests were performed using land surface scheme LAPS to compute the fluxes needed for surface temperature calculations.

Some of the key findings we obtained through this study is that depending on the relative size of the patches and their heights, and especially depending on the surface type, the decrease of albedo can be significant. Accordingly, it also leads to differences (increase) in

the surface temperature calculations. This indicates that geometrical effect plays an important role in albedo estimation. Our further work will be converted to a more complex and more demanding geometries.

ACKNOWLEDGEMENTS

The research we described in this paper has been funded by the Serbian Ministry of Science and Environmental Protection under the project "Modelling and numerical simulations of complex physical systems", No. ON141035 for 2006-2010.

REFERENCES

Delage, Y., Wen, L., and Belanger, J.M., Aggregation of parameters of the land surface model CLASS, *Atmosphere-Ocean*, 37, 157-178, 1999.

Dickinson, R.E., Land surface processes and climate-surface albedos and energy balance. *Advances in Geophysics*, Vol. 25, Academic Press, 305-353, 1983.

Kapor, D., Mihailovic, D.T., Tosic, T., Rao, S.T., and Hogrefe, C., An approach for the aggregation of albedo in calculating the radiative fluxes over heterogeneous surfaces in atmospheric models. In: A. Rizzoli and A. Jakeman (eds.), *Proceedings of International Environmental Modeling and Software Society (iEMSs)*, Vol.2, pp.389-394, Lugano, Switzerland, June 24-27. 2002.

Liou, K.N., An introduction to atmospheric radiation. *Second Edition. Academic Press*, Inc., NY, 2002.

Mihailovic, D.T., Rao, S.T., Alapaty, K., and Ku., J.Y., A study of the effects of land-use types on the boundary layer evolution. *Abstracts of Scientific Conference on Environmental Monitoring, Evaluation, and Protection in New York: Linking Science and Policy*, pp.39, 24-25 September 2001, Albany, New York, USA, 2001.

Mihailovic, D.T., Kapor, D., Hogrefe, C., Lazic, J., and Tosic, T., Parameterization of albedo over heterogeneous surfaces in coupled land-atmosphere schemes for environmental modeling. Part I: Theoretical background. *Enivron. Fluid Mech.*, Vol.4, pp.57-77(21), 2004.

Oke T.R, Boundary Layer Climates, 2nd edn., Methuen, London, New York, 435 pp., 1987.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Integration of Environmental Models Based on Fluid Mechanics in GIS

L. Matejicek[a], **Z. Janour**[b] **and L. Benes**[c]

[a] *Institute for Environmental Studies, Faculty of Science, Charles University in Prague, Benatska 2, Prague 2, 12801, Czech Republic (lmatejic@natur.cuni.cz)*

[b] *Institute of Thermomechanics, Academy of Sciences, Dolejskova 5, Prague 8, 18200, Czech Republic*

[c] *Department of Technical Mathematics, Czech Technical University in Prague, Karlovo Namesti 13, 12135, Czech Republic*

Abstract: An exploration of dust dispersion over the surface of coal mines requires more complex modelling tools in order to estimate dust emission and dust transport. Thus, spatial data and data originating from a numerical simulation are integrated into a GIS environment. The Reynolds averaged Navier-Stokes equations for incompressible flows with turbulent algebraic extensions are used for dispersion modelling of the wind flows and dust transport. In addition to spatial data management, the GIS is used for the creation of a digital terrain model originating from surface laser scanning, GPS measurements, a geodetic survey and existing data sources. This more precise mapping of the surface is required in order to estimate the wind flows over a temporary coal storage site and adjacent slopes. The flow patterns are displayed together with aerial images in the GIS for assessment of potential erosion fields over the stockpiles, and for an evaluation of environmental impacts. Integration of modelling in the GIS is carried out for testing of various environmental scenarios, including man-made barriers as well as different wind flow conditions.

Keywords: Dispersion modelling; Dust emission; Surface laser scanning; GIS.

1. INTRODUCTION

Environmental modelling covers a wide range of principles describing the motion of matter or energy in dependence on spatial scales, time scales and other attributes [Parker et al., 2002]. Nowadays, various software tools have been developed to solve environmental models based on spatial modelling or dynamic modelling, but they rely on different data structures and functionality. Spatial modelling, traditionally represented by geographic information systems (GISs), is focused on spatial relationships [Zeiler, 1999]. Dynamic modelling, based on the numerical solution of algebraic or differential equations [Jørgensen et al., 2001], is represented by models implemented in a number of simulation packages or standalone applications. This creates barriers in the linking of spatial models and dynamic models. In general, there are two ways in which to overcome these difficulties in environmental modelling.

First, from the spatial modelling point of view, existing GISs can be extended by data structures and numerical methods that can assist in solving dynamic models. The strategies can range from simple pre- and postprocessor linkage through shared data files to building dynamic models as fully functional extensions by using programming development tools in the GIS environment [Maidment, 1996].

Second, in the case of dynamic modelling, there are case oriented extensions for the display of simulation outputs. However, the data structures are dedicated for the calculation of numerical models [Littlefield, 2004]. In spite of the fact that the graphic user interface allows the models to be created as a work flow by stringing individual processes together, and running each model via its dialog box or the command line [Dabney, 2003], the interconnection with the original spatial environment is quite limited.

In order to demonstrate the integration of environmental models based on dispersion modelling, processes of dust emission and dust transport over a selected surface mining area are solved in the framework of the GIS project. Spatial data are used for a more accurate numerical simulation of wind flows over the temporary coal storage site and surrounding neighbourhoods. After simulation, the results are displayed together with map layers.

2. SPATIAL INFORMATION SYSTEM

The purpose of the selected spatial information system, namely ESRI's ArcGIS, is to provide a spatial framework for dispersion modelling. In addition to spatial data management, advanced spatial analysis is used to create the digital terrain model (DTM), to provide pre-processing and post-processing for numerical simulation, to interpolate output variables in the frame of thematic map layers, and to create virtual scenes for visualization. For these reasons, a new object-oriented data model is developed in the framework of the ESRI geodatabase [Maguire, 2005] that can characterize features and their relationship in a more natural way. Features represented by vectors, rasters, TINs or other user defined structures are stored and centrally managed in one database together with the defined spatial relationship. The main spatial data sources originate from the Landsat 7, aerial images, surface laser scanning, GPS measurements, a geodetic survey and existing digital thematic map layers.

2.1 Spatial data from passive remote sensors

The satellite scene from the Landsat 7 Enhanced Thematic Mapper (ETM+) is used for giving an overview of the whole mining area, and for the mapping of land degradation features related to soil erosion [Metternicht, 1998]. A true colour composite and pseudo colour composites combining the bands 1-7 at a resolution of 30 meters support display of the coal mine and its surroundings. The aerial images at a resolution of 0.5 meters are draped on the DTM to display the surface of the local areas of interest, the temporary storage site and the nearby residential zone.

2.2 Surface laser scanning and GPS mapping

In order to provide more accurate simulation of wind flows over the potential emission sources, surface laser scanning was chosen for acquisition of the DTM in the area of the temporary coal storage site and its neighbouring slopes. This represents the most important parts that influence dust transport by wind flows drifting from the mining areas to the residential zones. Multiple scans from different locations are provided to obtain complete data sources from shielded parts of the terrain. Finally, the clouds of points are geo-referenced in the framework of the GIS project together with satellite images and aerial images. In spite of the fact that airborne laser scanning is becoming the standard technique for acquisition of digital elevation models [Vosselman, 2005], surface scanning can capture spatial points from shielded areas. But, more data processing is needed to filter redundant points or incorrect points caused by vegetation interference and reflection of the mining equipment. A set of spatial points and lines is collected by GPS for mapping slopes and surface objects (excavators, conveyors, sorting places, man-made barriers, pylons). The accuracy of the GPS measurements is improved during the post-processing phase by data from the nearest reference stations in the territory of the Czech Republic. Thus, more

precise mapping is provided for the creation of the DTM that represents an input into the numerical modelling. Finally, the DTM together with thematic map layers complement visualization of the dust dispersion based on numerical simulation.

3. NUMERICAL SIMULATION

A number of software packages focused on air dispersion modelling can be used for integration into the spatial information system. For example, the ISC-AERMOD with its U.S. EPA models (ISCST3, ISC-PRIME, AERMOD and AERMOD-PRIME) can assess pollution concentration and deposition from a wide variety of emission sources. In order to run the simulation, a DTM complemented by surface objects (buildings and mining equipment) and meteorological data must be imported and pre-processed by tools included in the package. Though the package includes display tools, export of the data into the GIS is needed to provide more complex spatial analysis and visualization. In the case of dust dispersion over the surface mine, pre- and postprocessor linkage through shared data files is needed to run the simulation and to import data into the GIS [Matejicek, 2008].

Another approach is represented by a standalone software application that shares data by simple pre- and postprocessor linkage through shared data files. This approach can be made more comfortable by providing direct implementation in the GIS environment. Thus, the pre- and post processing procedures can be reduced through the sharing of data in the framework of a common database. But, in cases of time consuming numerical operations, actual implementation by scripts or macros is not computationally efficient.

In the case of the presented numerical simulation focused on dust dispersion over the surface coal mine, the ESRI geodatabase is used for spatial data management and support of advanced spatial analysis. The numerical simulation of wind flow and dust transport is based on Reynolds averaged Navier-Stokes (RANS) equations for incompressible flows with turbulent closure of the model by an algebraic turbulence model [Kozel, 2007]. The numerical simulation is solved by a semi-implicit finite-difference scheme. The DTM and data from local meteorological stations are used for setting internal parameters.

4. RESULTS

This study is focused on dust emissions from the selected dominant source, the temporary coal storage site. The procedures provided in the framework of the GIS project are illustrated in Figure 1.

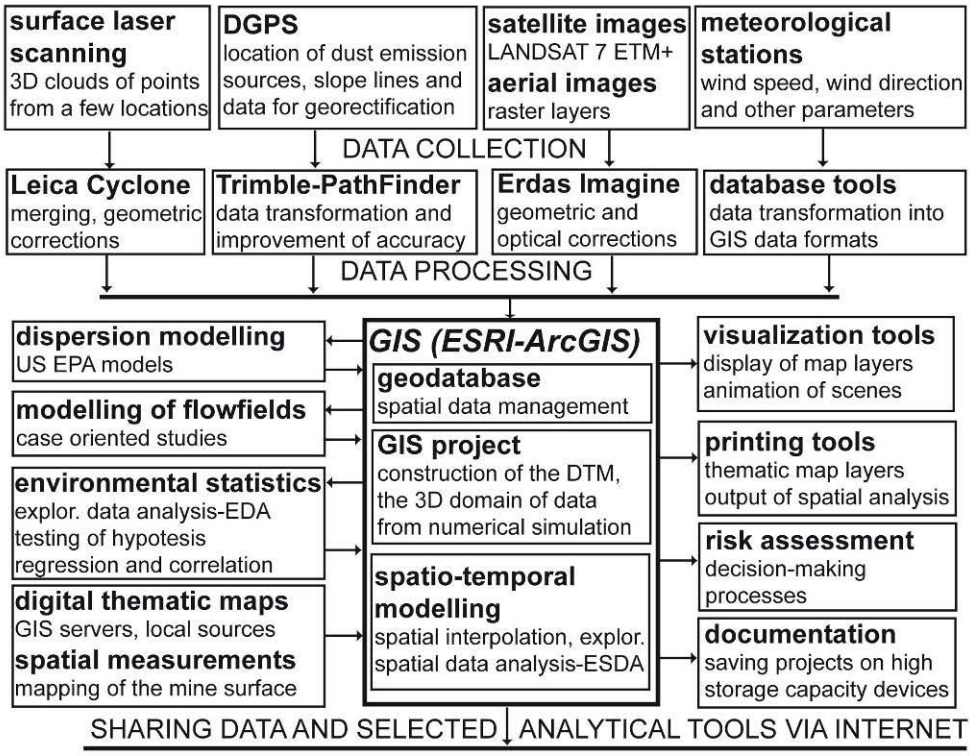


Figure 1. Data processing in the framework of the GIS project.

The upper part of the scheme in Figure 1 illustrates the main spatial data sources focused on the creation of the DTM (surface laser scanning, GPS and satellite/aerial images). The local meteorological stations assist in estimating the meteorological conditions (wind velocity, wind direction, temperature and atmospheric pressure) for numerical simulation of the wind fields and dust transport. After pre-processing, all the data are imported into the ArcGIS geodatabase. The final DTM complemented by the thematic layers focused on the dust concentration is shown in Figure 2. The aerial images in the background are draped on the DTM in order to provide a more realistic view of the dust transport over the surface mine and the neighbouring residential zone. The bottom layer illustrates the dust transport with the estimated predominant wind direction blowing from the west. A wind rose and histogram based on local meteorological measurements are shown in Figure 3. Thus, the predominant dust transport is in the direction of the residential zone. The grey fields indicate dust concentration levels greater than 1 $\mu g/m^3$.

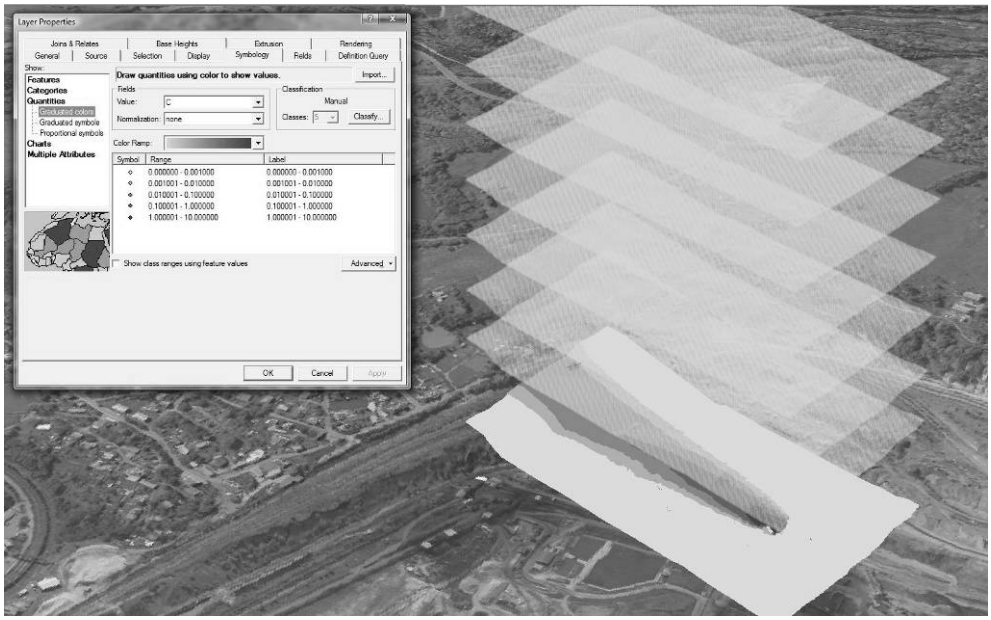


Figure 2. Scene with aerial images draped on the DTM, and the horizontal layers containing the dust concentrations (the bottom layer illustrates the dust dispersion from the emission source over the slope and the nearby residential zone; some of the upper layers are switched off or transparent in order to illustrate the dust concentration in the bottom layer).

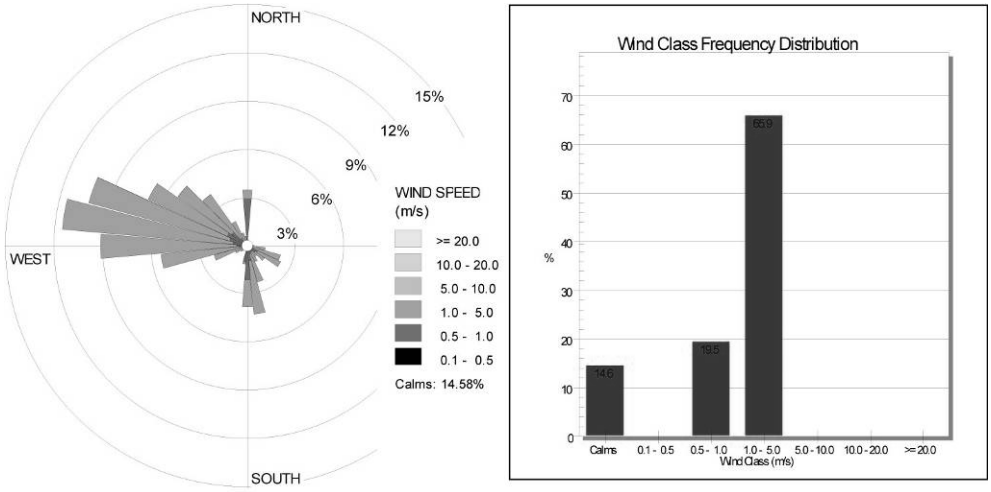


Figure 3. A wind rose and histogram indicating the wind class frequency distribution (the predominant wind flows are primarily blowing from the west).

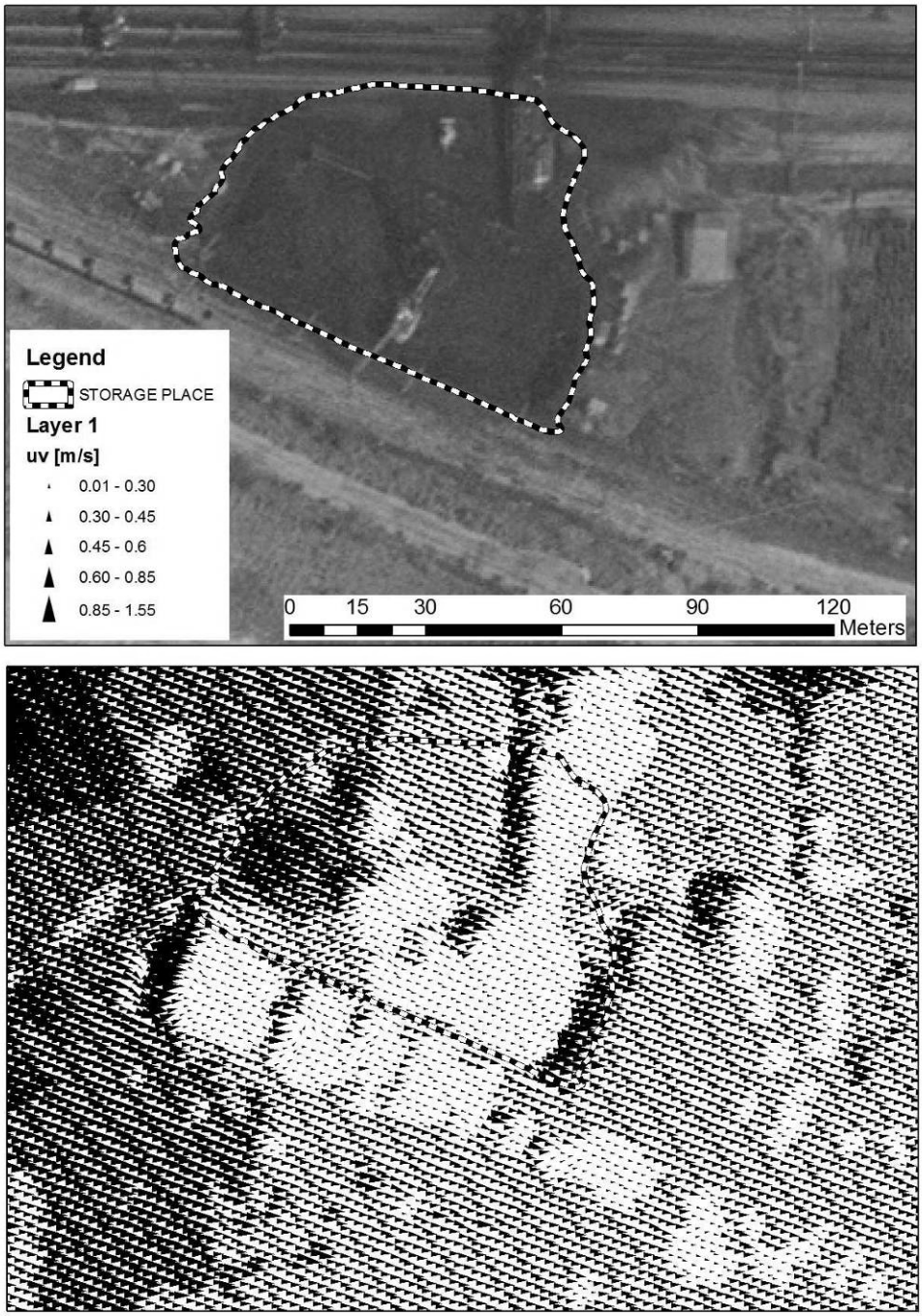


Figure 4. The wind flows over the surface of the temporary coal storage site (the arrows indicate wind speed and wind direction in the horizontal plane, approximately 1 meter above the surface).

5. CONCLUSIONS

This case study focused on the integration of spatial data and the outputs/inputs of dispersion modelling demonstrates more complex tools for the risk assessment of dust transport over a surface coal mining area. Compared with some previously published dust-transport studies [Chakraborty, 2002; Lu, 2001], sharing of the spatial data based on surface laser scanning together with the wind flow fields can result in a more detailed view of the processes of wind erosion. The processes depend strongly on local wind flow characteristics that are affected by terrain geometry. In the case of this studied area and the

temporary coal storage site in particular, the prediction of the wind flows over stockpiles represents a key stage in the assessment of potential erosion that causes significant environmental impacts. Generally, use of the three-dimensional numerical simulations based on RANS equations with turbulence extensions has been tested and validated in wind tunnels for simplified surface objects [Badr, 2005]. But, the numerical simulation based on the precise DTM allows more realistic predictions of dust emission and dust transport. Integration of the dispersion modelling in the GIS environment allows more complex insights into the studied processes. Thus, other similar studies will be carried out for the testing of both man-made barriers as well as different flow conditions, in order to minimize the dust emission and deposition in the neighbouring residential zones.

ACKNOWLEDGEMENTS

This paper was carried out in the framework of a project supported by the Academy of Sciences in the Czech Republic (AVCR 1ET400760405). The GIS project was developed in the GIS Laboratory at the Faculty of Natural Science, Charles University in Prague.

REFERENCES

Badr, T., and J.T. Harion, Numerical modeling of flow over stockpiles: Implication on dust emissions, *Atmospheric Environment*, 39, 5576-5584, 2005.

Chakraborty, M.K., M. Ahmad, R.S. Singh, D. Pal, C. Bandopadhyay, and S.K. Chaulya, Determination of the emission rate from various opencast mining operations, *Environmental Modelling & Software*, 17, 467-480, 2002.

Dabney, J.B., and T.L. Harman, Mastering Simulink, Prentice-Hall, New York, 2003.

Ghose, M.K., and S.R.Majee, Sources of air pollution to coal mining and their impacts in Jharia coalfield, *Environmental International*, 26, 81-85, 2000.

Jørgensen, S.E., and G. Bendoricchio, Fundamentals of ecological modelling, 3rd edition, Elsevier, Amsterdam, pp. 19-92, 2001.

Kozel, K., T. Bodnar, E. Gulikova, and V. Pisa, Numerical simulation of flow and pollution dispersion over obstacles in complex terrain, Colloquium Fluid Dynamics 2007, 43-44, Institute of Thermomechanics AS CR, Prague, 2007.

Littlefield, B.L., and D.C. Hanselman, Mastering Matlab 7, Prentice-Hall, New York, 2004.

Lu, H., and Y. Shao, Toward quantitative prediction of dust storms: an integrated wind erosion modelling system and its application, *Environmental Modelling & Software*, 16, 233-249, 2001.

Maidment, D.R., Environmental modeling within GIS. In M.F. Goodchild, L.T. Steyaert, B.O. Parks (eds), GIS and environmental modeling: Progress and research issues, GIS World Books, Fort Collins CO, pp. 315-323, 1996.

Maguire, D.J., M. Batty, and M.F. Goodchild, GIS, Spatial Analysis, and Modeling, ESRI Press, Redlands, CA, 2005.

Matejicek, L., Spatio-temporal modeling of the dust emissions from an opencast coal mining area. In A.N. Dubois (ed), Soil Contamination: New Research, NOVA Publishers, New York, in press.

Metternicht, G.I., and A. Fermont, Estimating erosion surface features by linear mixture modeling, *Remote Sensing of Environment* 64 (3), 254-265, 1998.

Parker, P., R. Letcher, and A. Jakeman, et al., Progress in integrated assessment and modelling, *Environmental Modelling & Software*, 17, 209-217, 2002.

Vosselman, G., P. Kessels, and B. Gorte, The utilisation of airborne laser scanning for mapping, *International Journal of Applied Earth Observation and Geoinformation*, 6 (3-4), 177-186, 2005.

Zeiler, M., Modeling our world: The ESRI guide to geodatabase design, ESRI Press, Redlands, CA, pp. 1-22, 1999.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Locating a Source of Air Pollution Using Inverse Modelling and Pre-computed Scenarios

M. Vujadinovic[a], B. Rajkovic[a] and Z. Grsic[b]

[a] *Institute of Meteorology, Faculty of Physics, Belgrade University, Dobracina 16, 11000 Belgrade, Serbia (bora@ff.bg.ac.yu)*
[b] *Institute of Nuclear Sciences Vinca, P.O. Box 522, 11001 Belgrade, Serbia (grsa@vin.bg.ac.yu)*

Abstract: An extension of the transition of a possible source of air pollution, as a combination of measurements and inverse modelling, based on Bayesian statistics, has been proposed. The method consists of two steps. The first one is creating a library of possible scenarios, where each scenario includes meteorological parameters, parameters related to the emission and calculated concentrations at measurement points. Once the library is formed, we calculate marginal probability function for the parameter that we would like to estimate. In the most cases parameters under consideration would be those related to the emission its position, intensity or duration. The parameter that we did estimation for was the position of the possible source, creating simulated measurements that had incorporated possible measuring errors. This, the second part of calculation, is extremely efficient and fast. The simplicity of the approach and its numerical efficiency qualifies it for the problem, especially in the operational mode. Members of the library and simulated measurements were generated using a puff model.

***Keywords**:* library of scenarios; point source; puff model; Bayesian statistics; probability density function.

1. INTRODUCTION

At the first sight, we can infer parameters of the pollution field from measurements obtained from a network of measuring stations, capable to detect presence of pollutant. If we wanted to make quantitative assessment we would need measurements of the amount of pollutant i.e. its concentrations in several (the more the better) stations. Further examinations for the case of localized sources and pollutant released in the atmosphere is not really possible, except in some special cases. Two, the most important, reasons are turbulent nature of the motions in the Planetary Boundary Layer (PBL) and very large gradients that are characteristic for the concentration fields. The large gradients are the primary reason why it is difficult to synthesize (analyze in meteorology) concentration filed only from measurements and therefore it is impossible to localize source(s) of specific pollution field. For more detailed formulation, please see the accompanied paper by Rajkovic et al., in this conference.

Still inverse modeling, search for the source, position or strength of a passive substance, must have, as its starting point, at least detection of the presence of a pollutant. Measurements of the amount of a pollutant are even better. So, we may say that one aspect of inverse modeling could be design of an optimal network of measuring points, stations, relative to the envisioned sources of pollution.
Problem of the source detection can be approached using transport equation in its Eulerian or Lagrangian form. In both cases, we have 3D domain with lateral boundaries, bottom and top. Initialization and formation of boundary conditions may be quite complicated problem. On the other end of modeling spectrum are Gaussian plum models as the “bulkiest” ones.

Wind is treated as a known constant, while thermodynamics is heavily parameterized through stability classes [Pasquill and Smith 1983, Melli and Runca, 1979, Mocioaca and Stefan 2003]. The thermodynamics can be improved using Monin-Obukhov theory or even second-order closure theories [Sykes and Gabruk, 1997]. The most problematic part in Gaussian plum approach is constant wind assumption. The next generation of "bulk" models avoids this simplification introducing series of puffs, whose superposition gives the concentration field. But, even using a puff model, as relatively simple and efficient, to perform direct integrations, covering all possible variations of the atmosphere state and emission properties, is not a practical approach.

The most efficient and complete (accurate) approach is a combination of direct modeling and inverse techniques with the Bayesian probabilistic approach. The main point is to construct probability density function (PDF) of the problem and then, from its marginal distribution, calculate the value of considered parameter. The Bayesian method [West and Harrison 1997, Bernardo and Smith 1994, Gelman et al. 2004] combines observed data and possible prior knowledge, in order to calculate value of unknown model parameters. Such knowledge is inherently probabilistic and therefore more complete and closer to the real world, since, in principal, it can take into account all the uncertainties in the measurements and modeling. The main problem is an error distribution's shapes. Now, the most of them are assumed to be Gaussian.

Calculating PDF can be done in several ways, like Markov Chain Monte Carlo method, or better suited sequential Monte Carlo approach [Gliks et al. 1996, Liu 2001, Andrieuet et al. 2003, Doucet et al. 2001, Arulampalam et al. 2002]. We propose a much simpler version, where we create a library of possible scenarios. Each scenario includes meteorological parameters, parameters related to the emission and calculated concentrations at measuring points. Once the library has been created and stored, it is very easy to generate probability density function of the problem and estimate any desired parameter from the corresponding marginal distribution.

2. MODEL, RESULTS AND ITS ANALYSIS

2.1 The Probabilistic Approach

Conjunction of probability is the most complete description of a physical system. It has, as its building blocks, probability functions defined on three sets: the parameter space, simulated measurements space and actual measurements space. For more detailed formulation see accompanying paper Rajkovic et al.

2.2 The Model and Source Localization Procedure

The model that we used to construct the library was a puff model [Grsic and Milutinovic 2000, Rajkovic et al. 2008]. The grid had 301x301 points with spatial distance of 150 meters. The time interval between two consecutive puffs was one minute. For further details about the model, we refer to the accompanying paper Rajkovic et al.

The methodology has two phases. The first phase is generating a set of many scenarios (realizations, in the probabilistic terminology) which we will call the library. These scenarios are runs of the puff model with variations of the problem parameters. There are two groups of parameters. The first one is consisting of parameters that define the source of the pollution, source's strength, its position and the duration of the release. In the second group are meteorological parameters: wind intensity, its direction and stability characterization of the PBL through Pasquill-Gifford categories. Using one combination of parameter's values, model produces concentrations in all measuring points. In our case, there were 13 measuring and 3 possible source points. The ranges of all parameters should be wide enough to cover all reasonable possibilities in the problem.

The computation of PDF, that characterizes the problem, is often done by some variation of the Monte Carlo method. But, if we really want to cover substantial part of possible ranges of all parameters using MC method, the computational effort will be extremely big. Instead, we have "digitized" values of various parameters into classes, appropriate for each parameter. For the wind intensity, we propose logarithmic scale, with winds of 0.1, 0.2, 0.5, 1, 2, 5, 10, 20 and 50 m/s, so that we certainly cover all possible situations. The stability classes are anyway quantized having values from 1 to 6. During testing of our method, we have found a very large sensitivity to the wind direction. Considering this, we propose wind direction interval of 1 degree, that is quite different from either 1/8 or even 1/16 of the full circle, as is standard way of characterizing wind direction. For the source locations we take positions of all possible sources in the particular configuration. For the source strength we take the exponential increment from 1 to 10^6 of the appropriate units. Finally, we have assumed that accident's duration is between 1 minute and 3 hours, which should be enough to cover realistic accidental release that can be treated using either Gaussian plum or puff models. Duration also depends on the size of the domain and relative positions of the sensors and possible pollution sources. At the end, we note that all the above values should be reexamined in each concrete case. The above choices are viewed as typical rather then all possible. This, first, phase can be quite computing expensive but it is done only once for a particular application. Please note that if we change the model or any of its components or parameterizations, the re-computing of the library is necessary.

The second phase can be conditionally called "operational". To carry it out, we must have meteorological observations of wind, its intensity and direction, decision which stability category is relevant for that moment (we do not want to discuss how this should be done, we only note that it can be done in many ways, in a simple manner or with some sub-model of small or quite high sophistication) and measured concentrations at measuring points. Starting from the implicit assumption, that deviation of the "true" values of all parameters are Gaussian, we form PDF as

$$\rho = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[\frac{(Qobs - Qlib)^2}{\sigma^2}\right], \tag{1}$$

for all members of the library whose meteorological parameters have been observed. The standard deviation (σ) measures the spread between observations (Q_{obs}) and values from the library (Q_{lib}). In general, it takes into account uncertainty in our a priori knowledge, uncertainties in measurements (σ_D) and modelling (σ_T). In this case, there is no a priori knowledge, so total σ is the sum of σ_D and σ_T. We have assumed that measurements errors were in the range of 5%, which together with values of simulated measurements gives an estimate for σ_D of 10^{-4}. On the other hand, we took total σ equal to $5 \cdot 10^{-2}$, which is two orders of magnitude larger than σ_D. This should take into account all modelling errors. Since wind intensity does not vary significantly, we could consider its average value during the accident. But, due to very high sensitivity of the model to wind direction, we should take the whole range of the observed directions. Once we have computed PDFs, its summation over all but desired parameter give us marginal probability for that parameter and its maximum gives the best estimate of it.

2.3 Results

Relative positions of the assumed source and measurement points are presented in the Figure 1.. Black dots are measurement points (13), the stars are sources (3) while the diamond is the source which we will look for. Instead of real measurements, we created simulated measurements using realistic (measured) winds, its intensity and direction, and temperatures at two meters height. These concentrations were then randomly perturbed by 5% mimicking the measurement errors. In most cases concentrations were "observed" at only two or three measuring points. The reason is that we have point source whose plume is quite narrow.

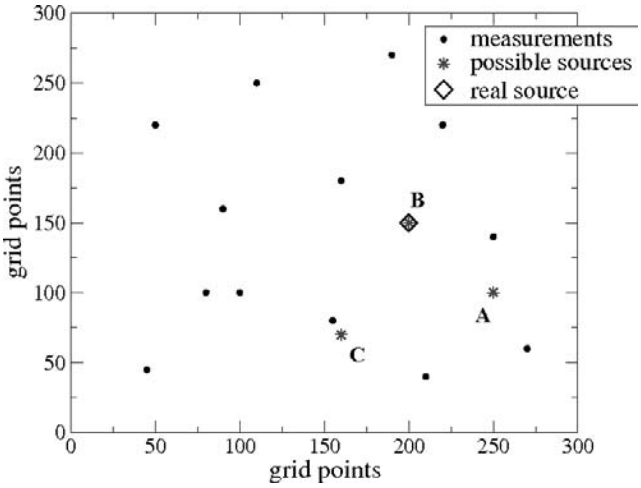


Figure 1. The position of measuring points (dots), possible sources A, B, C (stars) and real source (diamond).

As we said, the first step is library creation. From the meteorological data we knew that the mean wind intensity was about 6 m/s and stability had the value of 5 for the Pasquill-Gifford stability class, while the wind direction was in the range from 90 to 120 degrees. Based on these values, we were able to select the appropriate members of the library. We have also varied the duration of the accident in seven time intervals from 1 minute to 181 minutes, every 30 minutes. The source parameters were as explained in the previous section. In all three cases of possible sources we have found the correct one.

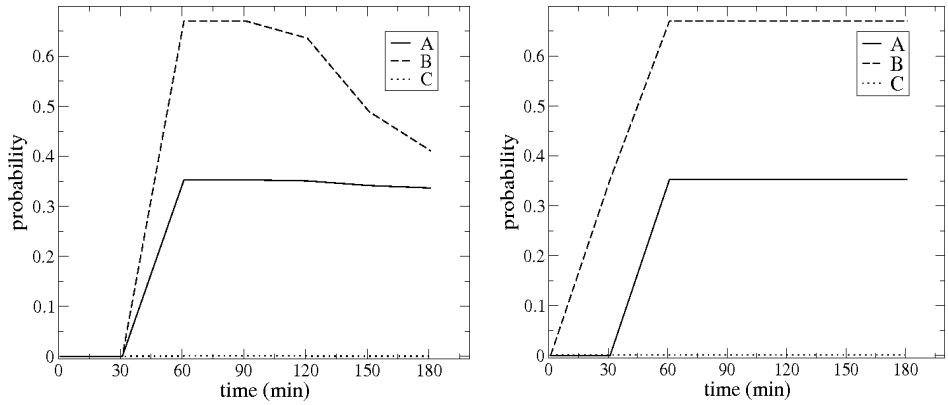


Figure 2. Marginal probability function for each of three sources (A, B, C) in 30 minutes intervals. On the left panel is its change in the case of measurements simulated using observed, while the right panel is for the case of measurements simulated using constant wind direction.

It is interesting to see how the probability changes with time for each of the source (Figure 2., left panel). Each source has different line style and the correct source (in this case it is source B) has the largest value, during entire integration. At the beginning, probability was very small, then reaches its maximum, thus indicating the duration time as well as the position, and then decreasing. The reason for this is that wind in the simulated measurements was realistic, changing its direction during the simulation, while the library records have constant wind direction. To check that, we redid our determination, but this time we kept the wind direction constant. Indeed, as shown in Figure 2. on the right panel, once the probability reaches its maximum, it stays constant.

3. CONCLUSIONS

Combination of pre-computed scenarios and simulated measurements, based on Bayesian statistical approach, results in an accurate and yet efficient method of detecting point like source of pollution. Time evolution of the problem indicates to the duration of the accident or how long ago, relative to the measurements, the accident occurred.

This methodology is independent of the choice of the air dispersion model, its complexity, parameterizations, etc.

ACKNOWLEDGMENTS

This work has been partially funded by Serbian Ministry of Science, Technology and Environment, grant no. 1197 and Italian Ministry of Environmental and Territories through its two projects, SINTA and ADRICOSM-STAR. The third author is partially funded through the project of the decommission of the nuclear reactor in Vinca.

REFERENCES

Andrieue, C., da Freitas, N., Doucet, A. and M.I. Jordan, Introduction to MCMC for machine learning, *Machine Learning*, 50, 5-43, 2003.

Arulampalam, M.S., Maskell, S., Gordon, N. and T. Clapp, A tutorial on practice filters for on-line non-linear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing*, 50(2), 174-188, 2002.

Bernardo, J.M. and A.F.M. Smith, *Bayesian theory*, Wiley, Chichester, 1994.

Doucet, A., de Freitas, N. and N.J. Gordon, *Sequential Monte Carlo methods in practice*, Springer-Verlag, Berlin, 2001.

Gelman, A., Carlin, J.B., Stern, H.S. and D.B. Rubin, *Bayesian data analysis*, Chapman and Hall, 668 pp.,London, 2004.

Gliks, W.R., Richards, S. and D.J.E. Spieglhalter, *Markov Chain Monte Carlo in practice*, Chapman and Hall, London, 1996.

Grsic, Z., and P. Milutinovic, Automated meteorological station and the appropriate software for air pollution distribution assessment, *Air Pollution Modelling and Its Application XIII*, 2000.

Liu, J.S., *Monte Carlo strategies in scientific computing*, Springer-Verlag, 343 pp., Berlin, 2001.

Melli, P. and E. Runca, Gaussian plum model parameters for ground-level and elevated sources derived from the atmospheric diffusion equation in a neutral case, *Journal of Applied Meteorology*, 18(9), 1216-1221,1979.

Mocioaca G. and S. Stefan, Different parameterization of a Gaussian scheme: intercomparison study, *International Journal of Environment and Pollution*, 19(1), 32-45, 2003.

Pasquill, F. and F.B. Smith, *Atmospheric diffusion*, John Wiley, 437 pp., New York, 1983.

Rajkovic, B., Arsenic, I. and Z. Grsic, *Fluid Mechanics of Environmental Interfaces*, Taylor and Francis, 348 pp., London, 2008.

Sykes, R.I. and R.S. Gabruk, A second-order closure model for the effect of averaging time on turbulent plum dispersion, *Journal of Applied Meteorology*, 36(8), 1038-1045, 1997.

West, M. and J. Harrison, *Bayesian forecasting and dynamic models*, Springer, 680 pp., New York, 1997.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Two Interacting Environmental Interfaces: Bifurcation in Coupled Asymmetric Logistic Maps

Dragutin T. Mihailovic
Faculty of Agriculture, University of Novi Sad, Dositej Obradovic Sq. 8,
21000 Novi Sad, Serbia
(guto@uns.ns.ac.yu)

Abstract: The field of environmental sciences is abundant with various interfaces and it is the right place for application of new fundamental approaches leading towards better understanding of environmental phenomena. For example, following definition of environmental interface by Mihailovic and Balaž [2007], such interface can be placed between natural or artificially built surfaces and atmosphere. In this case, visible radiation provides almost all of the energy received on the environmental interface. Since all the energy transfer processes occurs in the finite time interval the energy balance equation for this environmental interface can be written in terms of finite differences of ground and air temperatures and then, under some conditions, further transformed into the logistic equation [Mihailovic et al., 2001]. In the case of the two interacting environmental interfaces, we have a higher-dimensional complex system. The chaos in a higher dimensional system is one of the focal subject of physics today. One way, in studying this complex system, naturally leads to the model of coupled logistic maps with different strength parameters. In this paper, we report the results of numerical investigation on the system of two logistic maps, representing energy exchange on environmental interfaces, with different strength parameters such that the one map lies in a period one stable attractor or a bifurcation point and the other in chaotic region when decoupled.

Keywords: Chaos; Environmental Interface; Logistic Equation; Energy Balance Equation; Coupled Asymmetric Logistic Maps.

1. INTRODUCTION

The field of environmental sciences is abundant with various interfaces and it is the right place for application of new fundamental approaches leading towards better understanding of environmental phenomena. We defined the environmental interface as an interface between the two either abiotic or biotic environments which are in a relative motion exchanging energy through biophysical and chemical processes and fluctuating temporally and spatially regardless of its space and time scale [Mihailovic and Balaž, 2007]. This definition broadly covers the unavoidable multidisciplinary approach in environmental sciences and also includes the traditional approaches in sciences that are dealing with the environmental space. The environmental interface as a complex system is a suitable area for occurrence of the irregularities in temporal variation of some physical or biological quantities describing their interaction [Rosen, 1991; Selvam and Fadnavis, 1998; Zupanski and Navon, 2007]. For example, such interface can be placed between: human or animal bodies and surrounding air, aquatic species and water and air around them, natural or artificially built surfaces (vegetation, ice, snow, barren soil, water, urban communities) and atmosphere. The environmental interface of different media was recently considered for different purposes [Glazier and Graner, 1993; Nicolov et al., 1995; Martins et al., 2000; Niyogi and Raman, 2001]. In these systems visible radiation provides almost all of the energy received on the environmental interface. Some of the radiant energy is reflected back to space. The interface also radiates some of the energy received from the sun. The quantity of the radiant energy remaining on the environmental interface is the net radiation, which drives certain physical processes important to us. Since all the energy transfer processes occurs in the finite time interval the energy balance equation at any

environmental interface can be written in terms of finite differences of ground and air temperatures and then, under some condition, further transformed into the logistic equation [Mihailovic et al., 2001].

The chaos in higher dimensional system is one of the focal subjects of physics today. Along with the approach starting from modeling physical system with many degrees of freedom, there emerged a new approach, developed by Kaneko [1983] to couple many one-dimensional maps to study the behavior of the system as a whole. However, this model can only be applied to study the dynamics of a single medium such as the pattern formation in a fluid. What happens if two media border on each other like environmental interface? One may naturally leads to the model of coupled logistic maps with different strength parameters. Thus it is appropriate to inquire whether loosening the condition of strict identity might bring any new feature while keeping both maps to be logistic to hold the redundant controlling parameters minimal. Even two logistic maps coupled to each other may serve as the dynamical model of driven coupled oscillators [Midorikawa et al., 1995]. It has been found that two coupled identical maps possess several characteristic features which are typical of higher dimensional chaos. In this paper, we report the results of numerical investigation on the system of two logistic maps, representing energy exchange on environmental interfaces, with different strength parameters.

2. ENERGY BALANCE EQUATION FOR ENVIRONMENTAL INTERFACE

We see the outside world, i.e. the world of phenomena (*ambience*) from the observer's perspective (its inner world). In the ambience, there are *systems* of different levels of complexity and their environments. *System* in the ambience is a collection of precepts while whatever lies outside, like the component of a set, constitutes the environment [Rosen, 1991; Sivertsen, 2005]. The fate of science lies in the fact that it is focused on the system [Rosen, 1991]. Furthermore, to anticipate anything, the system is described by states (determined by observations) while the environment is characterised through its effects on the system.

Our basic equation is the energy balance equation. Since all the energy transfer processes occur in the finite time interval we shall immediately write this equation in terms of finite differences i.e. in the form of difference equation

$$\mathbf{D}T_i = \mathbf{F}_n \tag{1}$$

where $\mathbf{D}$ is the finite difference operator defined as $\mathbf{D}T_i = (T_{i,n+1} - T_{i,n} / \mathbf{D}t$, T_i is the environmental interface temperature, n is the time level, $\mathbf{D}t$ is the time step, $\mathbf{F}_n = (R_n - H_n - E_n - S_n)/c_i$ is defined at the nth time level, R is the net radiation, H, and E are the sensible and the latent heat, respectively, transferred by convection, and S the heat transferred by conduction into deeper layers of underlying matter while c_i is the environmental interface soil heat capacity per unit area,. Eq. (1) is also written in the finite difference form from an additional reason. It can be explain if we follow comprehensive consideration by van der Vaart [1973] about replacing given differential equations by appropriate difference equations in modelling of phenomena in physical and biological world. According to its many mathematical models for physical and biological problems have been and will be built in the form of differential equations or systems of such equations. With the advent of digital computers one has been able to find (approximate) solutions for equations that used to be intractable. Many of the mathematical techniques used in this area amount to replacing the given differential equations by appropriate difference equations, so that extensive research has been done into how to choose appropriate difference equations whose solutions are "good" approximations to the solutions of the given differential equations.

In Eq. (1) the sensible heat is calculated as $C_H(T_i - T_a)$ where C_H is the sensible heat transfer coefficient and $T_a(t)$ is the gas temperature given as the upper boundary condition. The heat transferred into underlying soil matter is calculated as $C_D(T_i - T_d)$ where C_D is the heat conduction coefficient while $T_d(t)$ is the temperature of

deeper layer of underlying matter that is given as the lower boundary condition. Following Bhumralkar [1975] and Holtslag and van Ulden [1975] the net radiation term in can be represented as $C_R(T_i - T_a)$ where C_R is the radiation coefficient. According to Mihailovic et al. [2001], for small differences of T_a and T_d, the expression for the latent heat can be written in a form $C_L f(T_a)[b(T_i - T_a) + b^2(T_i - T_a)^2]/2$. Here C_L is the latent heat transfer coefficient, $f(T_a)$ is the gas vapor pressure at saturation and b is a constant characteristic for a particular gas. Calculation of time dependent coefficients C_R, C_H, C_L and C_D can be found in Monteith and Unsworth. [1990]. After collecting all terms in Eq. (1) we get

$$\mathbf{D}T_i = A_1(T_{i,n} - T_{a,n}) - A_2(T_{i,n} - T_{a,n})^2 - A_3(T_{i,n} - T_{d,n}) \tag{2}$$

where $A_1 = [C_R - C_H - C_L b f(T_a)]/c_i$, $A_2 = b^2 f(T_a)/(2c_i)$ and $A_3 = C_D/c_i$ are coefficients also depending on $\mathbf{D}t$.With $\mathbf{D}t_p = 1/(A_1 - C_D/c_i)$ we indicate the scaling time range of energy exchange at the environmental interface including coefficients, that express all kind of energy reaching and departing the environmental interface. For any chosen time interval, for solving Eq. (2), there always exists $\mathbf{D}t_{p,l} = \min[\mathbf{D}t_p(c_i, C_R, C_H, C_L)]$ when energy at the environmental interface is exchanged in the fastest way by radiation, convection and conduction. If we define dimensionless time $\tau = \mathbf{D}t/\mathbf{D}t_{p,l}$ and use lower boundary condition, $T_{d,n} = T_{a,n} - (c_i/C_D)\mathbf{D}T_a$, then Eq. (2) after some transformations takes the form of the logistic equation, i.e.

$$\Gamma_{a+1} = \beta\Gamma_n(1 - \Gamma_n) \tag{3}$$

where the symbols introduced have the following meaning: Γ is the dimensionless quantity [Mihailovic et al., 2001] while $\beta = 1 + \tau$ taking values in the interval $1 < \beta < 4$.

3. TWO INTERACTING ENVIRONMENTAL INTERFACES

Under the aforementioned conditions Eq. (3) represents energy exchange at the uniform environmental interface. However, in the nature usually we encounter mixture of two environmental interfaces, for example, like surface covered by the sparse vegetation. In this case there are two different environmental interfaces. One is barren soil while the second one is the vegetative part. In this case of two interacting surfaces the energy exchange is more complex because it has to be described with two equations having the form of Eq. (3).

The system we study is two linearly coupled maps

$$x_{n+1} = (1 - \varepsilon) f(\mu, x_n) + \varepsilon f(\nu, y_n) \tag{4}$$

$$y_{n+1} = \varepsilon f(\nu, x_n) + (1 - \varepsilon) f(\nu, y_n) \tag{5}$$

where the map $f(\mu, x) = \mu x(1 - x)$ is taken to be the logistic map with strength parameters μ while ε is the coupling parameter. In case of $\mu = \nu$, two maps soon become synchronized no matter what initial conditions may be, i.e., coupled maps are identical with a single logistic map. The interesting is the case of $\mu \neq \nu$. In the following we fix the strength parameters above and below critical value $\mu_* = 3.56994$ for of μ, and ν respectively. We choose the strength parameters μ and ν and regard the coupling parameter ε as the controlling parameter**.** In Figs. 1 and 2, the attractors of the coupled-map are displayed as functions of coupling ε. Fig. 1 shows the result of $\mu = 4$ and $\nu = 3$, and Fig. 2 that of $\mu = 4$ and $\nu = 2$. In both cases for each value of ε, we used the final value of the previous ε and 1500 iterations are plotted. They are two typical examples of various values of μ and ν. One immediately notices several interesting features. The fact that there are two chaotic regions in both ε = 0 and ε = 1 ends seems odd at first sight,

but after some reflection, one realizes that very weak ε means very strong $(1-\varepsilon)$ which brings chaos first to the variable x, then to y however weak the coupling term may be. The most salient feature is the appearance of stable period four cycle right after the period one around ε = 0.77 in Fig. 1. Another, found in both (Fig. 1) and (Fig. 2) cases, is the sudden filling of the x and y space around ε = 0.85 and above. The broad window-like region with period four around ε = 0.9 in the case of (Fig. 2) is also noteworthy.

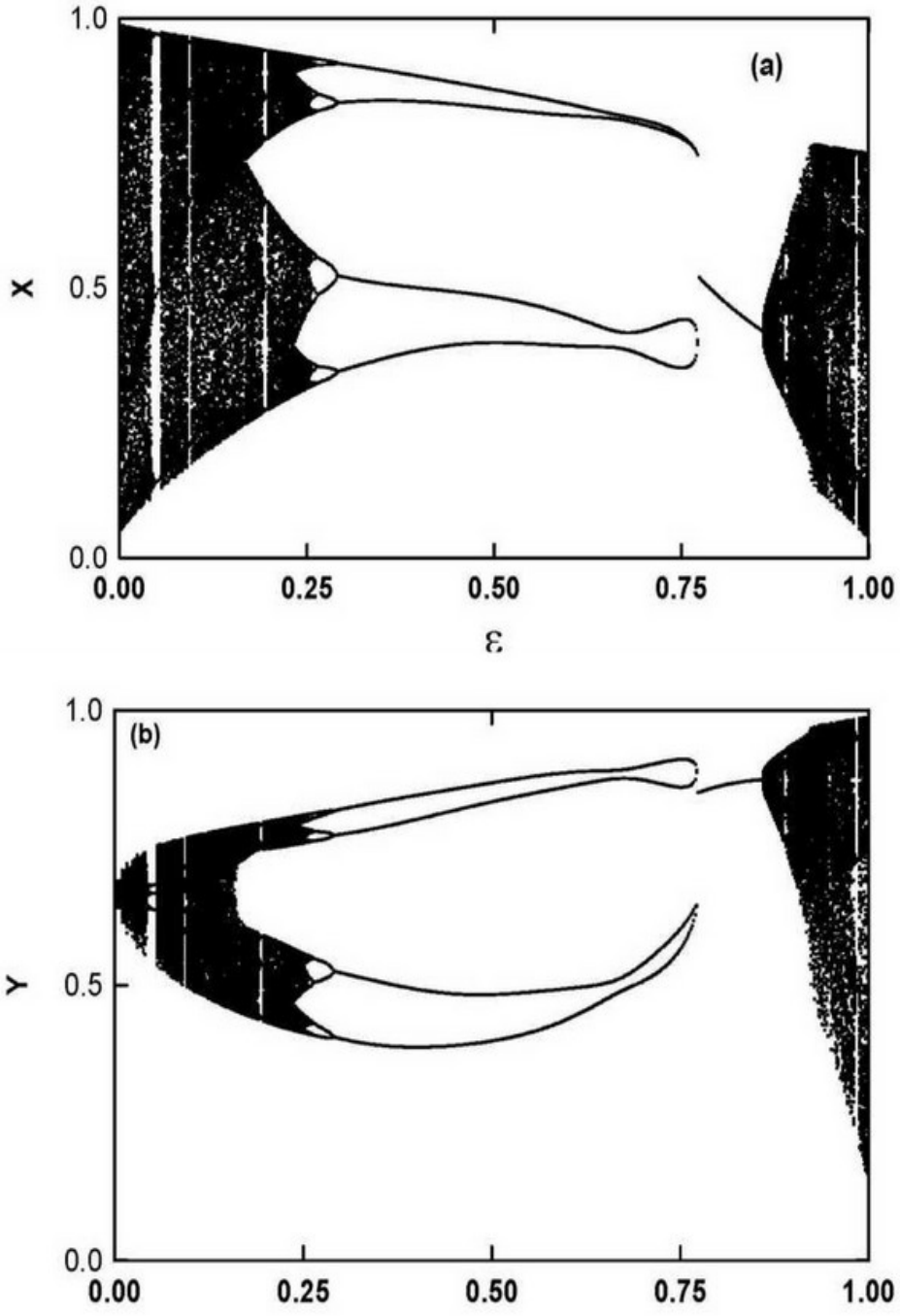


Figure 1. The phase diagram of the coupled maps Eqs. (4)-(5) for parameter values $\mu = 4$, $\nu = 3$; (a) is for x and (b) is for y.

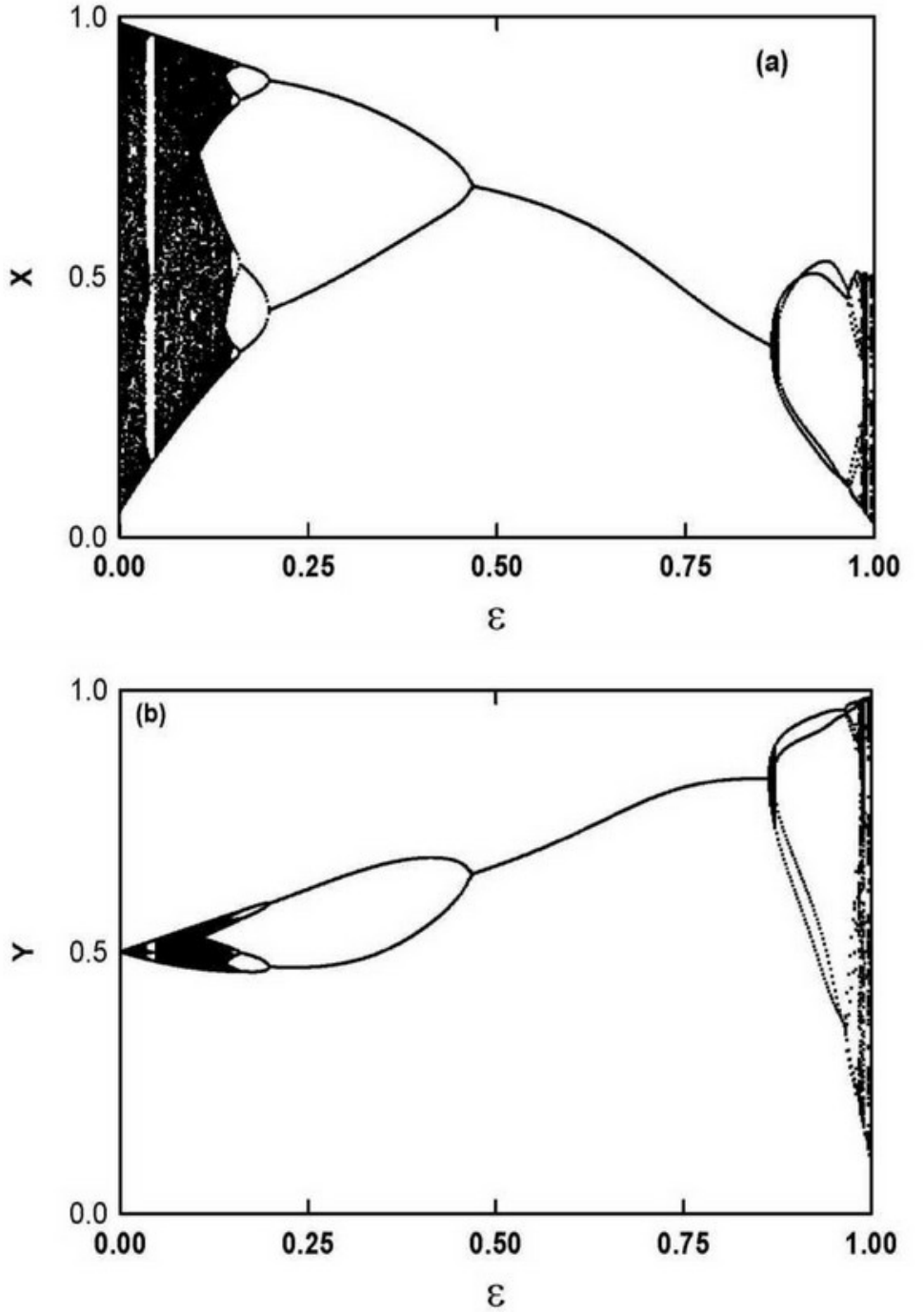


Figure 2. The phase diagram of the coupled maps Eqs. (4)-(5) for parameter values $\mu = 4$, $\nu = 2$; (a) is for x and (b) is for y.

The asymptotic behaviour of a series of iterates of the map can be characterised by the largest Lyapunov exponent which, any initial point $\vec{x}_0$ in an attracting region, is defined by

$$\lambda = \lim_{N\to\infty} \left\{ \ln[\left\| \underline{\mathbf{D}}^{(N)}(\vec{x}_0)/N \right\|] \right\} \qquad (6)$$

where $\left\|\underline{\mathbf{D}}^{(N)}\right\|$ is the norm of the derivative matrix, $\vec{x}_n$ comes from general vector mapping while N is the number of the last iterate. We calculated the Lyapunov exponent λ to see the behaviour of the coupled maps given by Eqs. (4)-(5) depending on different values of coupling parameter ε. Figure 3 shows Lyapunov exponent for the coupled maps as a function of ε ranging from 0 to 1. Each point was obtained by iterating many times from the initial condition to eliminate transient behaviour and then averaging over another 50 000 iterations starting from initial condition $x = 0.2, y = 0.25$ with 500ε values.

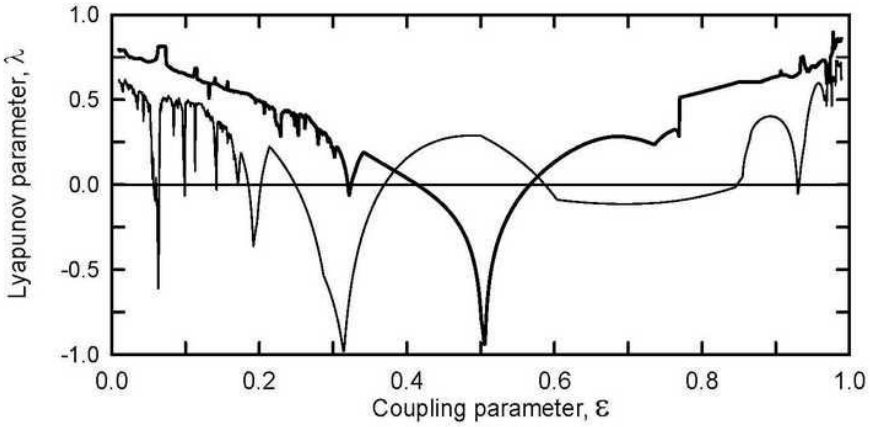


Figure 3. Lyapunov exponent of the coupled maps as a function of coupling parameter ε ranging from 0 to 1 for parameter values $\mu = 4$, $\nu = 3$ (thick line) and $\mu = 4$, $\nu = 2$ (thin line).

4. CONCLUSION

We reported the results of numerical investigation on the system of two coupled logistic maps, representing energy exchange of two interacting environmental interfaces. It has been done by calculating the phase diagrams of the coupled maps for different values of the strength parameters as well as calculation of the Lyapunov exponent as a function of the coupling parameter. It seems that further analysis of this system will be useful for understanding the process of energy exchange between two interacting environmental surfaces.

ACKNOWLEDGEMENT

The research work described in this paper has been funded by the Serbian Ministry of Science and Technology under the project "Modelling and numerical simulations of complex physical systems ", No. ON141035 for 2006-2010.

REFERENCES

Bhumralkar, C.M.,Numerical experiments on the computation of ground surface temperature in an atmospheric general circulation model. *J. Appl. Meteorol. 14*, 1246-1258, 1975.

Glazier, J., and Graner, F., Simulation of the differential adhesion driven rearrangement of biological cells, *Phys. Rev. E*, 47, 2128 – 2154, 1993.

Holtslag, A.A., and A.P. van Ulden, 1975: A simple scheme for daytime estimates of the surface fluxes from routine weather data. *J. Appl. Meteor.*, 22, 517-529.

Kaneko, K., Transition from Torus to Chaos Accompanied by Frequency Lockings with Symmetry Breaking— In Connection with the Coupled-Logistic Map, *Prog. Theor. Phys.* 69(5), 1427-1442, 1983.

Martins, M.L., Ceotto, G., Alves, G., Bufon, C.C.B, Silva, J.M., and Laranjeira, F.F., Cellular automata model for citrus variegated chlorosis, *Phys. Rev. E*, 62, 7024 – 7030, 2000.

Midorikawa, S., Takayuki, K., and Taksu C., Folded bifurcation in coupled asymmetric logistic maps, *Prog. Theor. Phys.*, 94, 571-575, 1995.

Mihailovic, D.T, and Balaž, I., An essay about modelling problems of complex systems in environmental fluid mechanics, *Idojaras*, 111(2-3), 209-220, 2007.

Mihailovic, D.T., Kapor, D.V., Lalic, B., and Arsenic, I., The chaotic time fluctuations of ground surface temperature resulting from energy balance equation for the soil-surface system, *Abstracts of the 26th General Assembly of European Geophysical Society*, March 20-25. Nice, France, 2001.

Monteith, J.L., Unsworth, M., Principles of Environmental Physics, Second Edition, Elsevir, p. 304, 1990.

Nikolov, N., Massman, W., Schoettle, A., Coupling biochemical and biophysical processes at the leaf level: an equilibrium photosynthesis model for leaves of C3 plants, *Ecol. Modelling*, 80, 205-235, 1995.

Niyogi, D.S., and Raman, S., *Numerical modeling of gas deposition and bidirectional surface–atmosphere exchanges in mesoscale air pollution systems*. In: Mesoscale Atmospheric Dispersion, Ed. Z. Boybeyi, WIT Publications, Southampton, UK, Advances in Air Pollution, Vol. 9, 1-51, 2001.

Rosen, R., *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press. New York, pp. 285, 1991.

Selvam, A.M., and Fadnavis, S., Signatures of a universal spectrum for atmospheric interannual variability in some disparate climatic regimes, *Meteor. Atmos. Phys.*, 66(1-2), 87-112, 1998.

Sivertsen, T.H. Discussing the scientific method and a documentation system of meteorological and biological parameters. *Phys. Chem. Earth.* 30:35-43, 2005.

van der Vaart, H.R., A comparative investigation of certain difference equations and related differaential equations: Implications for model building. *Bull. of Math. Biol.,* 35, 195-211, 1973.

Zupanski, M., and Navon, I.M. Predictability, observations, and uncertainties in the geosciences. *Bull. Amer. Meteor. Soc.*, 88, 1431–1433, 2007.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Validation of the Hydrodynamic Module of a 3D Ecological Model for the Ria de Aveiro Lagoon (Portugal)

C.I. Silva[a] and J.F. Lopes[b]

[a] *Departamento de Física, Universidade de Aveiro, Portugal (isabelvs@aeiou.pt)*

[b] *CESAM, Departamento de Física, Universidade de Aveiro, Portugal (jflopes@ua.pt)*

Abstract: The purpose of this study is to validate a hydrodynamic module of a *3D* ecological model. During the calibration, the parameters were adjusted so that the outputs match field data. The model is forced by considering tidal surface elevation at the open sea boundary and the river. The results show good agreement with the field data, for both the calibration and the validation experiments.

Keywords: Ria de Aveiro, lagoon, hydrodynamic flow model, validation

1. INTRODUCTION

Validated mathematical models for lagoonar coastal areas can be seen as tools for reconstruction real phenomena occurring in these systems. It is, therefore, required that numerical models for estuarine hydrodynamics should be calibrated and validated before used in practical applications. However, the procedures to perform these tasks are not widely accepted (Cheng et al., 1991). Calibration and validation methods appear in several forms, depending on data availability, water body characteristics and researchers opinion (Hsu et al., 1999). In the case of hydrodynamic modelling, the calibration is frequently made comparing short time series of predicted and measured data for the same location and period of time (Cheng et al., 1993). Another method consists in comparing the harmonic constituents generated from predicted and observed data (Smith, 1977).

2D models have been extensively applied in the study of the Ria de Aveiro hydrodynamics (Dias and Lopes, 2006; Lopes and Dias, 2006; 2007) and water quality (Lopes et al., 2006; 2008). The purpose of this work is to implement the hydrodynamic module of a fully *3D* ecological model of the Ria de Aveiro lagoon by calibration and validation. The model calibration has been performed, as first approach, through a qualitative comparison of the simulated temporal evolution of SSE (sea surface elevation) with data for 1987/1988 on several locations of the lagoon. The validation procedure was performed using an independent data set of *SSE* values (1997). In the overall the model was evaluated using the minimum error involved in the calibration and the validation process.

Having the hydrodynamic model fully implemented, it will be used for future studies concerning the lagoon *3D* hydrodynamics structure, namely the transport of salt, heat as well as non conservative biogeochemical particles.

2. Ria de Aveiro

Ria de Aveiro (Fig. 1) is a shallow lagoon situated in the Northwest Atlantic coast of Portugal (40º38'N, 8º45'W) which corresponds to a *UTM-22* zone. The lagoon is 45 km long and 10 km wide and covers an area of 83 km^2 at high tide (spring tide), which is reduced to 66 km^2 at low tide. It is supplied with freshwater by two main rivers, the Antuã river (5 m^3/s average flow) and the Vouga river (50 m^3/s) (Dias et al., 1999), the last contributing the highest fresh-water input to the lagoon. Five other small rivers, with

average flow of $1m^3/s$, have small contributions to the overall freshwater budget of the lagoon.

Tides, which are semi-diurnal, are the main forcing of the circulation in the Ria de Aveiro lagoon. The estimated lagoon tidal prism is $136.7*10^6$ m^3 for maximum spring tide and $34.9*10^6$ m^3 for minimum neap tide (Dias, 2001). The bathymetry is probably the most important factor that affects the flow in shallow systems like Ria de Aveiro. The bathymetry controls the spatial variability of current magnitude and direction.

Numerical studies of the general circulation of the Ria de Aveiro show, that tides are strongly deformed in their propagation through the Ria de Aveiro, from the mouth to the far end of each channel, due to channels geometry and bathymetry (Dias et al., 2000). The general characteristics of the tidal wave in the lagoon are those of a damped progressive wave (Dias, 2001).

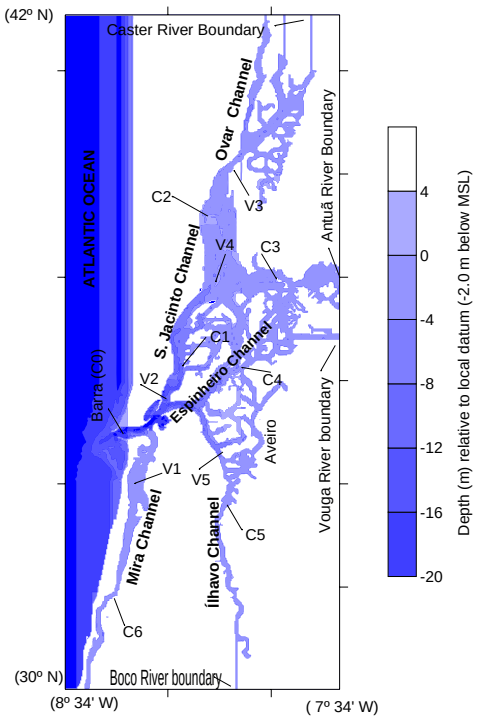


Figure 1. Ria de Aveiro lagoon, with locations of the stations used in the calibration and validation of the numerical model and the open boundaries.

3. NUMERICAL SIMULATION OF THE RIA DE AVEIRO

3.1 Model Description

The three dimensional model presented here is based on finite volume techniques. This model was developed at the *DHI Water & Environment*. It is composed by a hydrodynamic module (*HD*) which is based on the numerical solution of the three dimensional continuity, momentum, temperature, salinity and density equations (*MIKE3-Flow Model*, 2005). The momentum equations are used in their incompressible *Reynolds* averaged version for the *Navier-Stokes* equations, involving the *Boussinesq* assumption and the hypothesis of the hydrostatic pressure in the vertical. The turbulence closure is affected using the *Smagorinsky* formulation for diffusion in the horizontal and the *k-ε* model in the vertical direction. In the horizontal plane both *Cartesian* and spherical coordinates can be used. The free surface is taken into account using a sigma coordinate transformation approach.

The physical model can be represented by the set of the basic equations referent to the hydrodynamic *3D* mode:

$$\frac{\partial u}{\partial t}+\frac{\partial u^2}{\partial x}+\frac{\partial vu}{\partial y}+\frac{\partial wu}{\partial z}=fv-g\frac{\partial \eta}{\partial x}-\frac{1}{\rho_0}\frac{\partial p_a}{\partial x}-\frac{g}{\rho_0}\int_z^\eta \frac{\partial \rho}{\partial x}dz+F_u+\frac{\partial}{\partial z}\left(\upsilon_t\frac{\partial u}{\partial z}\right)+u_sS \qquad (1)$$

$$\frac{\partial v}{\partial t}+\frac{\partial v^2}{\partial x}+\frac{\partial uv}{\partial y}+\frac{\partial wv}{\partial z}=-fu-g\frac{\partial \eta}{\partial y}-\frac{1}{\rho_0}\frac{\partial p_a}{\partial y}-\frac{g}{\rho_0}\int_z^\eta \frac{\partial \rho}{\partial y}dz+F_v+\frac{\partial}{\partial z}\left(\upsilon_t\frac{\partial v}{\partial z}\right)+v_sS \qquad (2)$$

$$\frac{\partial p}{\partial z}+\rho g=0 \qquad (3)$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \qquad (4)$$

where *t* is the time; *x*, *y* and *z* are the *Cartesian* co-ordinates; $h = \eta + d$, η is the surface elevation; d is the still water depth and h is the total water depth; u, v and w are the velocity components in the *x*, *y* and *z* direction respectively; $f = 2\Omega \sin\phi$ is the *Coriolis* parameter (Ω is the angular rate of the Earth and ϕ the geographic latitude); g is the gravitational acceleration; ρ is the density; ν_t is the vertical turbulent or eddy viscosity; p_a is the atmospheric pressure; ρ_0 is the reference density. *SD* is the magnitude of discharge due to point sources and (u_s, v_s) is the velocity by which water is discharged into the ambient water.

The vertical velocity is the velocity across a level of constant σ and is solved using a vertical sigma transformation and can be written as:

$$\omega = \frac{1}{h}\left[w + u\frac{\partial d}{\partial x'} + v\frac{\partial d}{\partial y'} - \sigma\left(\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x'} + v\frac{\partial h}{\partial y'} \right) \right] \tag{5}$$

At the bottom, the friction shear stress is imposed assuming a logarithmic velocity profile:

$$\frac{\vec{\tau}_b}{\rho_0} = C_d \vec{u}_b |\vec{u}_b| \tag{6}$$

Where C_d is the drag coefficient defined as:

$$C_d = \left[\frac{k\left(1 - \frac{\kappa_s}{30Z_b}\right)}{\log\left(\frac{30Z_b}{k_s}\right) - \left(1 - \frac{\kappa_s}{30Z_b}\right)} \right] \tag{7}$$

where *k* is the turbulent kinetic energy calculated farther (11), Z_b in the vertical extent of the bottom grid cell, $\vec{u}_b$ =(u_b,v_b) is the flow velocity above the bottom and κ_s is the bed roughness length scale. This expression has been obtained by assuming a logarithmic velocity profile at the bottom grid cell and depth averaging over the extent of the grid cell. At the free surface, the flux of momentum is also imposed in the form of a shear stress.

The transports of temperature, *T*, and salinity, *s*, follow the general transport–diffusion equations as:

$$\frac{\partial T}{\partial t} + \frac{\partial uT}{\partial x} + \frac{\partial vT}{\partial y} + \frac{\partial wT}{\partial z} = F_T + \frac{\partial}{\partial z}\left(D_\upsilon \frac{\partial T}{\partial z} \right) + \hat{H} + T_s S \tag{8}$$

$$\frac{\partial S}{\partial t} + \frac{\partial uS}{\partial x} + \frac{\partial vS}{\partial y} + \frac{\partial wS}{\partial z} = F_s + \frac{\partial}{\partial z}\left(D_\upsilon \frac{\partial S}{\partial z} \right) + S_s s \tag{9}$$

where T_s and s_s are the temperature and the salinity of the source, *SD* is the magnitude of discharge due to point sources, D_v is the vertical turbulent (eddy) diffusion coefficient, $\hat{H}$ is a source term due to heat exchange with the atmosphere. *F* corresponds to the horizontal diffusion terms defined as:

$$(F_T, F_S) = \left[\frac{\partial}{\partial x}\left(D_h \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y}\left(D_h \frac{\partial}{\partial y} \right) \right](T, s) \tag{10}$$

where D_h is the horizontal diffusion coefficient farther defined.

Turbulence is modelled using the eddy viscosity concept. The eddy viscosity is often described separately for the vertical and the horizontal transport. The hydrodynamic model has several turbulence models that we can apply: a constant viscosity, a vertically parabolic viscosity and a standard *k*-ε model.

Parameters and their units	Values
Time step (s)	6
Flood and drying depth (m)	0.2
Flooding depth (m)	0.4
Initial constant value for Surface elevation (m)	2.6
Bed roughness (m)	
κ_s	0.2
Turbulence (empirical constants)	
($c_{1\varepsilon}$, $c_{2\varepsilon}$, $c_{3\varepsilon}$)	(1.44, 1.92, 0)
Diffusion parameters	
(σ_k, σ_ε, σ_T)	(1, 1.3, 0.9)
k and ε lower limits (m^2/s^2, m^2/s^3)	
	($1.0*10^{-7}$, $5.0*10^{-10}$)
Eddy viscosity limits (m^2/s)	
1st range	($1.8*10^{-6}$ - 12000)
2nd range	($1.8*10^{-6}$ – 12000)
3rd range	($1.8*10^{-6}$ – 30)
Background salinity (Psu)	32
Background temperature (ºC)	13

Table 1. Parameters values used in the hydrodynamic model calibration and validation.

In the *k-ε* model the eddy-viscosity is derived from turbulent variables, *k* and *ε*, as $\upsilon_t = c_\mu k^2/\varepsilon$ (Hartmut and Baumert, 2007; Burchard and Beckers, 2004), where *k* is the turbulence kinetic energy per unit mass (*TKE*), *ε* is the dissipation of the *TKE*. The horizontal and the vertical diffusion coefficients can be related to the eddy viscosity by $D_h = \upsilon_t/\sigma_k$ and $D_v = \upsilon_t/\sigma_\varepsilon$ respectively, where σ_k and σ_ε are *Prandtl-Schmidt* numbers and c_μ is an empirical constant defined in Tab. 1.

k and *ε*, are obtained from the following transport equations:

$$\frac{\partial k}{\partial t}+\frac{\partial uk}{\partial x}+\frac{\partial vk}{\partial y}+\frac{\partial wk}{\partial z}=F_k+\frac{\partial}{\partial z}\left(D_v\frac{\partial k}{\partial z}\right)+P+B-\varepsilon \tag{11}$$

$$\frac{\partial \varepsilon}{\partial t}+\frac{\partial u\varepsilon}{\partial x}+\frac{\partial v\varepsilon}{\partial y}+\frac{\partial w\varepsilon}{\partial z}=F_\varepsilon+\frac{\partial}{\partial z}\left(D_v\frac{\partial \varepsilon}{\partial z}\right)+\frac{\varepsilon}{k}\left(C_{1\varepsilon}P+C_{3\varepsilon}B-C_{2\varepsilon}\varepsilon\right) \tag{12}$$

where the shear production, *P*, and the buoyancy production, *B*, are given as:

$$P=\frac{\tau_{xz}}{\rho_0}\frac{\partial u}{\partial z}+\frac{\tau_{yz}}{\rho_0}\frac{\partial v}{\partial z}\approx \upsilon_t\left(\left(\frac{\partial u}{\partial z}\right)^2+\left(\frac{\partial v}{\partial z}\right)^2\right) \tag{13}$$

$$B=-\frac{\upsilon_T}{\sigma_T}N^2 \qquad \text{and} \quad N^2=-\frac{g}{\rho_0}\frac{\partial \rho}{\partial z} \tag{14}$$

$c_{1\varepsilon}$, $c_{2\varepsilon}$ and $c_{3\varepsilon}$ are empirical constants defined in Tab. 1.

The horizontal diffusion terms for turbulence are defined as:

$$(F_k, F_\varepsilon) = \left[\frac{\partial}{\partial x}\left(D_h \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y}\left(D_h \frac{\partial}{\partial y} \right) \right](k, \varepsilon) \quad (15)$$

3.2 Discretisation

The numerical bathymetry used in this study was developed by the *Hydrographic Institute of Portuguese Navy (IH)*. The numerical grid must be sufficiently refined to resolve all variations of depth and geometry of the lagoon. This horizontal grid is composed by 266 X 654 cells, with dimensions of 60 X 60 meters. A time step as small as 6 s is required to satisfy the *Courant-Friedrichs-Levy* condition (*CFL*) and to avoid instabilities.
The spatial discretisation of the primitive equations is performed using a cell-centred finite volume method. In the horizontal plane an unstructured grid is used while in the vertical direction the discretisation is structured. The elements can be prisms or bricks whose horizontal faces are triangles or quadrilateral elements, respectively. A Rieman solver is used for computation of the convective fluxes, which makes it possible to handle discontinuous solutions. For the time integration a semi-implicit approach is used where the horizontal terms are treated implicitly.

3.3 Boundary conditions

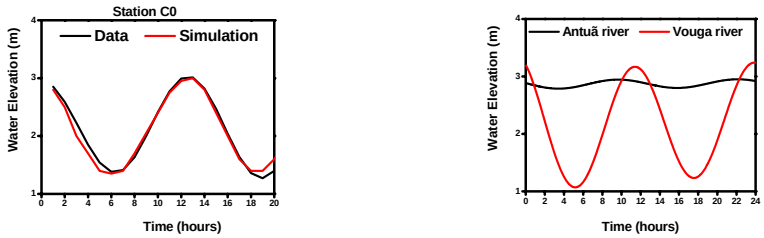


Figure 2. Water elevations at the open boundary (Barra) and at the Antuã and the Vouga rivers.

Several types of boundaries were used in this application: ocean and rivers open boundaries.

At the ocean open boundary the free surface elevation is specified whereas at the river boundaries the water flow or elevation can be imposed (Fig. 2). In this study the water elevation was specified at the river boundaries. No mass fluxes at the surface and bottom were considered.

The Ria de Aveiro lagoon system is affected by various natural and human stresses. The major natural stresses are tidal forces and variable recharge due to climatic variations. On the other hand it is also a place of discharge of domestic and industrial wastes, which flows do not contribute to the lagoon hydrodynamic (Dias et al, 1999).

3.4 Calibration and validation of the hydrodynamic model and sensitive analysis

The hydrodynamic model calibration was carried out with sea surface elevation data for 1987/1988.

The results show that friction dissipates energy as the tidal wave propagates from the lagoon mouth towards the end of channels (Dias et al., 2006). This behaviour is a common feature in estuarine environments (Hsu et al., 1999) and was analyzed and quantified in Ria de Aveiro. Considering the lagoon geometry and the tidal range at the mouth, the

magnitude of the bottom friction coefficient determines the tidal range variation along the lagoon channels (Dias et al., 2006).

In many studies the procedure adopted to calibrate the model consisted in spatially varying the bottom roughness coefficient, by assuming *Manning*'s values to specific regions of the numerical domain. In this study the model is not very sensitive to the bottom roughness, but rather to the water depth and rivers flow. The bottom roughness has been kept constant and equal to 0.2 m in the simulation. It is known that the water depth strongly influences the bottom stress. This parameter is introduced into the calculations of shear stress.

Fig. 3 shows the comparison between data and computed water elevations time series for six stations used in the model calibration (this procedure involved 22 stations).

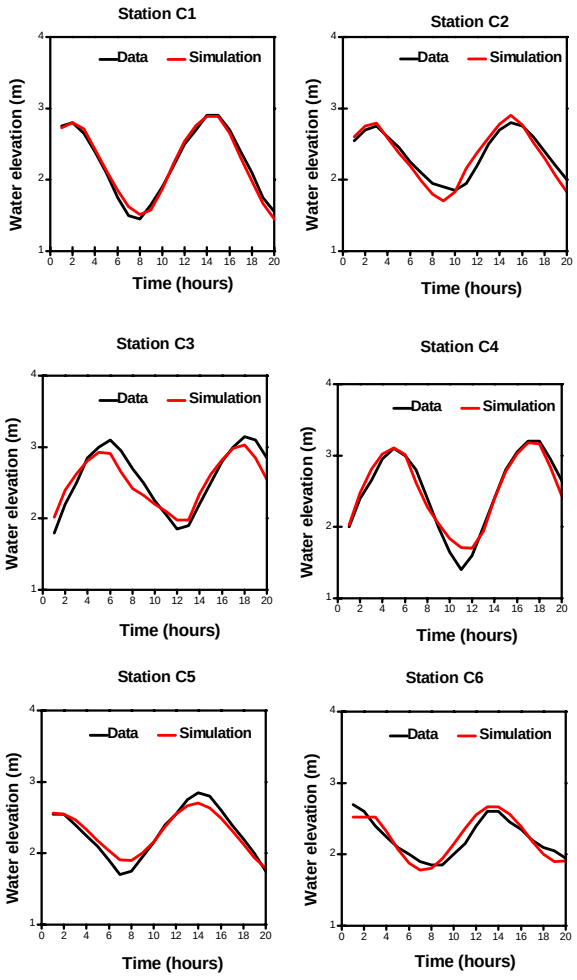


Figure 3. Comparison between water elevations for stations used in the model calibration.

The assessment of the relation between data measured and the model of SSE was performed with the help of *RMS* error defined as:

$$RMS = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\xi_0(t_i) - \xi_m(t_i) \right]^2 \right\}^{1/2} \qquad (16)$$

where $\zeta_0(t_i)$ and $\zeta_m(t_i)$ are the observed and modelled data of *SSE*, respectively, and *N* is the number of measurements in the time series.

Although the calibration involved 22 stations, only 6 stations are presented. Tab. 2 summarizes the *RMS* error between the model and data. In general there is a good agreement between the simulations and data, for all stations. The minimum *RMS* error was obtained for station *C1* and *C5* (0.099 m and 0.89 m respectively), whereas a maximum *RMS* error of 0.15 m was calculated for station *C3*. The higher *RMS* errors observed for *C3* as well as for *C6* may be explained by bathymetry effect, the shallowness of the area (see Fig. 1) and the uncertainties related to the river flows. The agreement between simulated and observed data from station *C0* (Barra, see Fig. 2) was not perfect, as *RMS* error of 0.1 m was found for this station. Indeed, this error may be explained by the fact that the boundary where the tide is imposed is located far western of *C0*.

Nevertheless, in overall, the 22 stations used in the calibration, show *RMS* values lower than 0.15 m, which corresponds to a relative error, for the water level, of the order of 6%. Considering the above results, it can be considered that the model is calibrated.

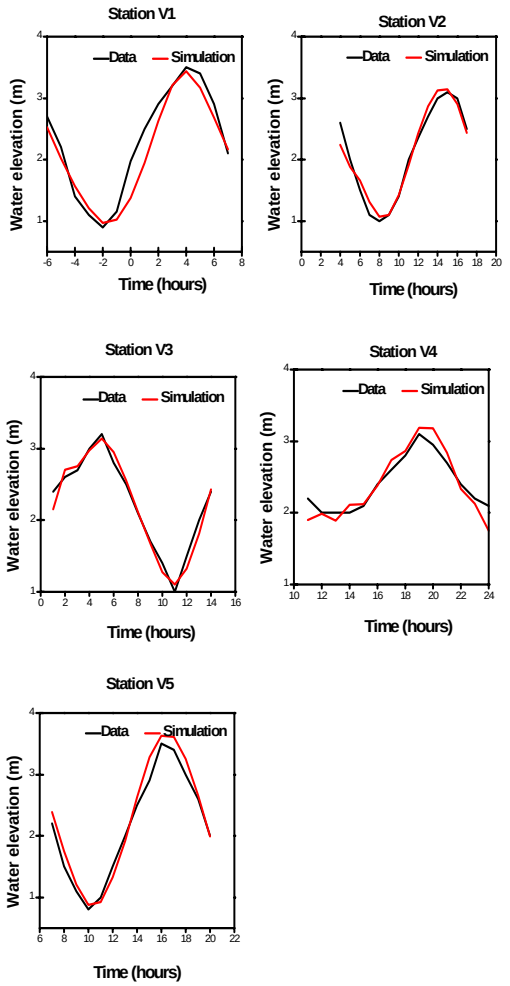


Figure 4. Comparison between water elevations for stations used in the model validation.

	Station C1	Station C2	Station C3	Station C4	Station C5	Station C6
RMS error	0.10 m	0.11 m	0.15 m	0.11 m	0.09 m	0.13 m

Table 2. Estimated *RMS* error for the model calibration.

The model validation is defined as a procedure consisting in comparing the model output with field or laboratory data to prove the model efficiency (Dias et al., 2006).

The measurements used in the validation have to be independent from the set used in the calibration.

The validation corresponds to a very wet period during June 1997 where the model performance was evaluated by comparing the *RMS* error between modelled and observed data of *SSE*. These conditions were used to study the model's response to the interaction of tidal forcing and varying river discharge.
Fig. 4 shows results concerning the model validation for water elevation for five stations. Tab. 3 summarizes the *RMS* error between simulation and data. A good agreement can be observed, as in general the *RMS* values are low.

The highest value was found at station *V1* with the *RMS* error around 0.15 m. From these results it may be concluded that the hydrodynamic model for the Ria de Aveiro has been successfully validated.

	Station V1	Station V2	Station V3	Station V4	Station V5
RMS error	0.146 m	0.050 m	0.041 m	0.055 m	0.062 m

Table 3. Estimated *RMS* error for the model validation.

The methodology of the sensitivity tests consisted in the assessment of the effect of the most sensitive parameters/processes involved in the calibration procedure, allowing variations of their values within a range of 10-30%.

During the calibration it was found that concerning the influence of the bottom stress, the model was more sensitive to the small variations of the water level than to the roughness length scale. It was also found that for stations situated near the far ends of the lagoon, the most important process was the river inputs.

Fig. 5 presents the sensitivity test concerning these two processes for the station *C4*. The case (a) compares the calibration to a situation for which the river flow was increased by a value of the order of 20%, whereas the case (b) compares the calibration to a situation for which the water depths at the area corresponding to station *C4* were increased by a value of the order of 30% relatively to the original bathymetry. In both cases it can be observed the deviations of the amplitude and phase lag relatively to the calibration/data. The inaccuracy in the bathymetry definition has, therefore, important repercussion in the tidal wave propagation inside the lagoon, namely at the shallow areas. The results show as well that the definition of an accurate river flow at the rivers boundaries is also important to correctly simulate the hydrodynamic of the lagoon.

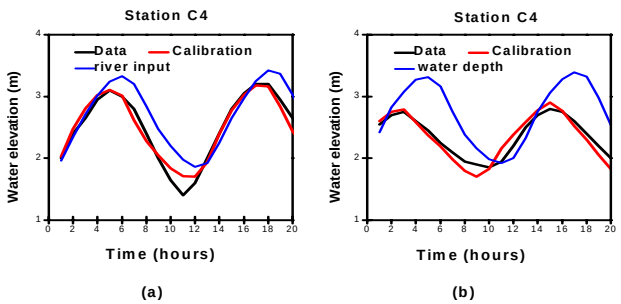


Figure. 5. Model sensitivity test for the station C4: (a) influence of the river input; (b) influence of the water depth.

4. Conclusions

The hydrodynamic model was successfully applied for the Ria de Aveiro lagoon. The results show that the calibration and validation of the hydrodynamic model was successfully carried out, revealing a good agreement between data and predictions. They show that inaccuracies in the bathymetry definition of the shallow areas, associated with the complex structure of the lagoon, composed by several very narrow channels, which are not well resolved by the horizontal grid, are crucial for the accurate simulation of the lagoon hydrodynamic.

ACKNOWLEDGEMENTS

This research was supported by the Portuguese Science Foundation (FCT) through Project Simclave POCI/MAR/56296/2004

REFERENCES

Burchard H., Beckers J.M., 2004. Non-uniform adaptive vertical grids in one-dimensional numerical ocean models. Ocean Modelling 6, 51–81

Cheng, R.T., Burau, J.R., Gartner, J.W., 1991. Interfacing data analysis and numerical modelling for tidal hydrodynamic phenomena. In: Parker, B.B. (Ed.), Tidal Hydrodynamics. John Wiley & Sons, New York, USA, pp. 201-219.

Cheng, R.T., Casulli, V., Gartner, J.W., 1993. Tidal, residual, intertidal mudflat (TRIM) model and its applications to San Francisco Bay, California. Estuarine, Coastal and Shelf Science 36, 235-280.

Dias J.M., Lopes, J.F., Dekeyser, I., 1999. Hydrological characterization of Ria de Aveiro, in early summer. Oceanologia Acta 22, 473-485.

Dias, J.M., Lopes,J.F., & Dekeyser, I., 2000. Tidal propagation in Ria de Aveiro lagoon, Portugal. Phys. Chem. Earth (B), 25, 369-374.

Dias, J.M., Lopes, J.F., 2006. Implementation and assessment of hydrodynamic, salt and heat transport models: The case of Ria de Aveiro Lagoon (Portugal). Environmental Modelling & Software 21, 1-15.

Hartmut P., Baumert H. Z., 2007. Validating a turbulence closure against estuarine microstructure measurements. Ocean Modelling 19, 183-203

Hsu, M.H., Kuo, A.Y., Kuo, J.T., Liu, W.C., 1999. Procedure to calibrate and verify numerical models of estuarine hydrodynamics. Journal of Hydraulic Engineering, 166-182.

Lopes, J.F., Dias, J. M., 2006. Numerical modelling of cohesive sediments transport in the Ria de Aveiro lagoon, Portugal. Journal of Hydrology, 319, 176-198.

Lopes, J.F., Dias, J. M., 2006. Residual circulation and sediment distribution in the Ria de Aveiro lagoon, Portugal. Journal of Marine Systems 68, 507-528.

Lopes, J.F., Silva C. I., 2006. Temporal and spatial distribution of dissolved oxygen in the Ria de Aveiro lagoon. Ecological Modelling 197, 67-88.

Lopes, J.F., Silva, C. I., Cardoso, A. C., 2008. Validation of a water quality model for the Ria de Aveiro lagoon, Portugal. Environmental Modelling & Software 23, 479-494.

Mike3 Flow Model (Ecolab Model), Release 2005. Hydrodynamic Module. D.H.I., Denmark.

Rodi W. (1984) Turbulence models and their application in hydraulics. International Association for Hydraulic Research, 2nd edition, Delft, Netherlands.

Smith, N. P., 1977. Metorological and tidal exchange between Corpus – Christi Bay, Texas, and the northwestern Gulf of Mexico. Estuarine Coastal Marine Scince 5, 511-520.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Numerical Simulation of Flow Patterns and Mass Exchange Processes in Dead Zones

C. Gualtieri

Department of Hydraulic, Geotechnical and Environmental Engineering (DIGA), University of Napoli Federico II, Napoli, Italy (carlo.gualtieri@unina.it)

Abstract: Contaminants and nutrients transport in streams and rivers is significantly affected by the presence of dead zones that are due to geometrical irregularities, that may have both natural and anthropic origins. Dead zones are characterized by mean flow velocity in the main stream direction approximately equal to zero and by exchange processes of solutes with the main stream. This paper presents the preliminary results of a numerical study undertaken to investigate the fundamental flow phenomena around and inside a simplified geometry, representative of flow in natural rivers with dead zones. Exchange rates between the dead zone and the channel were calculated resulting in a good agreement with previous literature experimental values for the same geometry.

Keywords: Environmental Fluid Mechanics; Numerical Simulation; Mass exchange; Dead zones; Verification.

1. INTRODUCTION

Solute transport in streams and rivers is strongly related to river characteristics, such as mean flow velocity, velocity distribution, secondary currents and turbulence features. These parameters are mainly determined by the river morphology and the discharge conditions. Most natural channels are characterized by relevant diversity of morphological conditions. In natural channels, changing river width, curvature, bed form, bed material and vegetation are the reason for this diversity, whereas in rivers which are regulated by man-made constructions, such as spur dikes, groins, stabilized bed and so on, the morphological diversity is often less pronounced and, thus, flow velocities are more homogeneous. In natural channels, some of these morphological irregularities, such as small cavities existing in sand or gravel beds, side arms and embayments, can produce recirculating flows which occur on different scales on both the riverbanks and the riverbed. These irregularities act as dead zones for the current flowing in the main stream direction. In regulated rivers, groyne fields are the most important sort of dead zones. Groyne fields can cover large parts of the river significantly affecting its flow field. Dead-water zones or *dead zones* can be defined as geometrical irregularities existing at the river periphery, within which the mean flow velocity in the main stream direction is approximately equal to zero [Weitbrecht 2004]. Dead zones significantly modify velocity profiles in the main channel as well affect dispersive mass transport within the river. Thus, in recent years exchange processes between the main stream and its dead zones were increasingly studied, mostly using experimental laboratory and field works [Kimura and Hosoda, 1997; Wallast et al., 1999; Uijttewaal, 1999; Muto et al., 2000a; Muto et al., 2000b; Weitbrecht and Jirka, 2001a; Weitbrecht and Jirka, 2001b; Uijttewaal et al., 2001; Kurzke et al., 2002; Engelhardt et al., 2004; Uijttewaal, 2005; Brevis et al., 2006; Le Coz et al., 2006; Jamieson and Gaskin, 2007; Weitbrecht et al., 2008]. Also, numerical studies were recently carried out to investigate flow hydrodynamics in channels with groyne fields [McCoy et al., 2008].

The objective of this paper is to present the preliminary results of a numerical study undertaken to investigate the fundamental flow phenomena around and inside a simplified

geometry, representative of flow in natural rivers with dead zones. Also, exchange processes of a solute between the main flow and the dead zones were studied with the evaluation of the exchange coefficient between the main flow and the dead zones. For the present study, two-dimensional steady-state numerical simulations were performed with Multiphysics 3.3™ in three different geometries representing dead zones, that is a single rectangular embayment and a groyne field with 10 and 5 groynes, respectively. These geometries were those previously experimentally studied by Muto et al. [2000a; 2000b], Wallast et al. [1999], and Weitbrecht and Jirka [2001b].

2. LITERATURE REVIEW

As already outlined, in streams and rivers, mass transport of solutes is significantly affected by the exchange processes between the main flow in the channel and the dead-water zones existing at its periphery [Valentine and Wood, 1977; Weitbrecht 2004]. In natural channels, dead zones can be formed by sand or gravel banks, side cavities, small pockets at the river bed or any other irregularity in the morphology both at the riverbanks and at the riverbed, whereas in regulated rivers, spur dikes and groin fields are typical dead zones. Due to momentum exchange across the interface with the main channel, flow patterns inside the dead zones are characterized by recirculating flows which occur on different scales and exhibit flow velocity in the main stream direction close to zero. Therefore, the most important effect of dead zones on mass transport in rivers is the storage of some amount of the contaminants or nutrients being transported by the main stream inside the dead zones, that is some mass of solutes are first trapped in the dead zones and only later, after an average storage time T_D, are released back into the main flow. The storage time depends on the strength of the exchange processes occurring between the main stream and the dead zones, which are due mostly to turbulent mixing in the lateral or in the vertical direction, if the dead zone is at the riverbanks or at the riverbed, respectively [Muto et al., 2000b; Weitbrecht, 2004]. As discussed below in details, the flow structure in a dead zone consists of a mixing layer, a primary eddy and a core region within this eddy. Depending on the dead zone aspect ratio a smaller corner eddy may exist. Turbulent mixing around a dead zone is mostly governed by two-dimensional coherent structures in the mixing layer between the dead zone and the main channel, even if other mechanisms and 3D structures are believed to be also involved [Brevis et al., 2006; Le Coz et al., 2006; Jamieson and Gaskin, 2007]. The effect of these exchange processes on the transport of solutes in a river with dead zones can be quantified considering the average transport velocity of solutes, defined as the velocity of the center of mass of the tracer cloud, and the strength of the longitudinal dispersion process, which describes the rate of stretching of the tracer cloud [Weitbrecht, 2004].

In a channel without dead zones, the average transport velocity of a tracer cloud that is completely mixed over the river cross-section equals the mean flow velocity. If dead zones are present, the part of the tracer cloud trapped in the dead zones is retarded compared to the part of tracer cloud travelling in the main stream with the mean flow velocity. In this case, the transport velocity of the tracer cloud is lower than the mean flow velocity.

Also, dead zones produce an increased stretching of a passing tracer cloud, which means that the longitudinal dispersion is enhanced. This results in a tail of the contaminants cloud longer than that predicted by the 1-D advection-diffusion equation [Valentine and Wood, 1979, Czernuszenko et al., 1998] and the length of the tail depends on the exchange between the main flow and the dead zones. This could be explained by two processes [Weitbrecht, 2004]. First, dead zones modify the transverse profile of flow velocity in the main channel and increase lateral turbulent mixing, which are the determining parameters for longitudinal dispersion (Fig.1). It is well known that transverse mixing is important in determining the rate of longitudinal mixing because it tends to control the exchange between regions of different longitudinal velocity. Particularly, transverse mixing and longitudinal mixing are inversely proportional. A strong transverse mixing tends to erase the effect of differential longitudinal advection and pollutants particles migrate across the velocity profile so fast that they essentially all move at the mean speed of the flow, causing

only a weak longitudinal spreading. On the other hand, a weak transverse mixing implies a long time for differential advection to take effect, so the pollutants patch is highly distorted while it diffuses moderately in the transverse direction and longitudinal mixing is large (Cushman-Roisin, 2007). Second, tracer cloud is stretched because solute parcels are trapped within the dead zones and only later released back into the main channel. Both processes result in an enhanced longitudinal dispersion, which leads to lower peak levels but to a longer period during which critical values could be exceeded. Finally, these processes are strongly related to the geomorphological conditions of the dead zones.

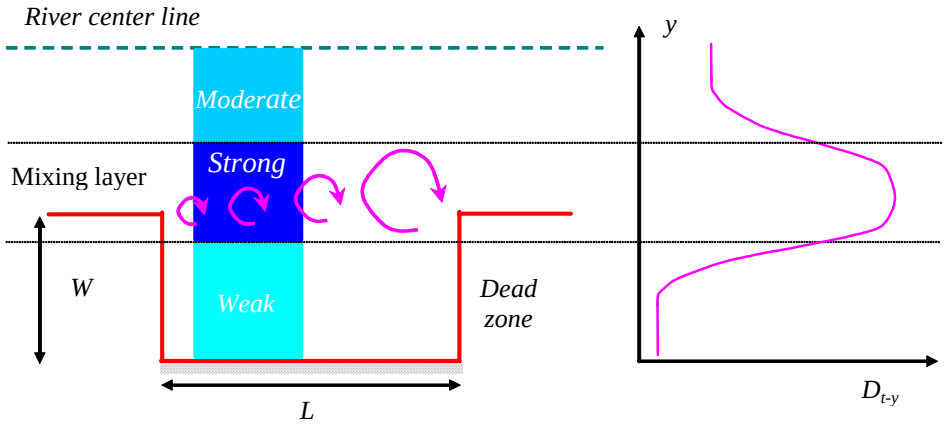


Figure 1. Transverse turbulent diffusivity profile in a river with dead zone

Since classical 1-D advection-diffusion equation cannot satisfactorily describe observed tracer concentrations from an instantaneous release in natural channels, where there is a relevant amount of dead zones, models accounting for the dead zones effects were proposed. The basic idea of the dead-zone model (DZM) is to distinct within the cross-section of a river, two zones, the main stream and the dead-zone. In the main stream the mass transport is governed by advection in the longitudinal direction, longitudinal shear due to the velocity distribution and transverse turbulent diffusion. Thus, in the main stream transport processes can be modeled under well mixed conditions using 1-D advection-diffusion equation. In the dead zone, since velocity in the main stream direction is close to zero, transverse turbulent diffusion across the interface between the dead zone and the main stream is the dominant mechanism, which leads to momentum and mass exchange processes. Assuming that in the dead zone the solute concentration is uniform, mass exchange between the dead zone and the main stream is proportional to the difference of the averaged concentration in the dead zone and in the main channel. To set up the DZM, conservation of mass for the main stream and the dead-zone has to be considered, that is [Czernuszenko et al., 1998; Weitbrecht, 2004; Weitbrecht et al., 2008]:

$$\frac{\partial C_S}{\partial t} = -U\frac{\partial C_S}{\partial x} + D_L\frac{\partial^2 C_S}{\partial x^2} + \frac{h_E\, E}{h_S\, B}\left(C_D - C_S\right) \tag{1a}$$

$$\frac{\partial C_D}{\partial t} = -\frac{h_E\, E}{h_D\, W}\left(C_D - C_S\right) \tag{2a}$$

where C_S and C_D are the solute concentration in the main stream and in the dead zone, respectively, U is mean velocity, D_L is longitudinal dispersion coefficient, B and W are channel width and dead zone width, respectively, h_S, h_D and h_E are the water depth in the main channel, in the dead zone and at their interface, respectively, and E is the exchange velocity across this interface (Fig.1). Usually, the ratios $h_E E/h_S B$ and $h_E E/h_D W$ which have the unit of $[T^{-1}]$, are called *exchange coefficients* K_S and K_D and the above equations yield:

$$\frac{\partial C_S}{\partial t} = -U\frac{\partial C_S}{\partial x} + D_L\frac{\partial^2 C_S}{\partial x^2} + K_S\left(C_D - C_S\right) \tag{1b}$$

$$\frac{\partial C_D}{\partial t} = -K_D \left(C_D - C_S\right) \tag{2b}$$

As previously noted, one important parameter of the dead zone is its *average storage time* T_D, which could be approximated by calculating the time needed to completely change the water volume in the dead zone. Thus, T_D could be obtained dividing the dead zone volume by the exchange velocity and the interface area, if L is the dead zone length, as [Weitbrecht, 2004]:

$$T_D = \frac{W\ L\ h_D}{E\ L\ h_E} = \frac{1}{K_D} \tag{3}$$

which demonstrates that T_D and K_D are reciprocal, that is higher exchange coefficients correspond to lower retention time inside the dead zone. Notably, if the concentration in the main stream is zero, Eq.(2b) corresponds to first-order decay law. Thus, for the case of a instantaneous and homogeneous mass release into the dead zone, the change of concentration inside the dead zone is described by:

$$C_D\left(t\right) = C_0\ exp\left(-K_D\ t\right) = C_0\ exp\left(-t/T_D\right) \tag{4}$$

where C_0 is the initial concentration in the dead zone. Eq.(4) confirms that the exchange coefficient K_D represents the slope of the decay process occurring inside the dead zone. Usually, K_D is normalized with the mean velocity in the mean stream and the width of the dead zone giving a dimensionless exchange coefficient [Valentine and Wood, 1979]:

$$k = \frac{K_D\ W}{U} \tag{5a}$$

Note that if the water depth in the main stream and in the dead zone are equal, that is h_E=h_D, the exchange coefficient K_D becomes:

$$K_D = \frac{E}{W} \tag{6}$$

and k is the ratio between the exchange velocity and the main stream velocity:

$$k = \frac{E}{U} \tag{5b}$$

Typical values of k parameter are in the range from 0.012 to 0.05 [Valentine and Wood, 1979; Wallast et al., 1999; Muto et al., 2000b; Weitbrecht and Jirka, 2001b; Weitbrecht, 2004].

Dimensional analysis leads to conclude that dimensionless exchange coefficient k has the form [Weitbrecht, 2004]:

$$k = f\left(Re, Fr, Pe, \frac{k_s}{h_S}, \frac{h_D}{h_S}, \frac{B}{h_S}, \frac{W}{L}, \alpha, S\right) \tag{7a}$$

where *Re*, *Fr* and *Pe* are Reynolds, Froude and Peclet number, respectively, k_s is bed roughness and α and S are the inclination to the main flow direction and the shape of the dead zone. However, it can be stated that under typical turbulent flow and low Froude number conditions, a major influence on the mass exchange is given by the aspect ratio of the dead zone *W/L* and the inclination angle leading to discard most of the parameters involved in eq.(7a), which yields:

$$k = f\left(\frac{W}{L}, \alpha\right) \tag{7b}$$

Previous experimental studies demonstrated that the dimensionless exchange coefficient k decreased with the increasing aspect ratio *W/L* of the dead zone [Muto et al., 2000a, 2000b; Uijttewaal et al., 2001; Weitbrecht and Jirka, 2001a; Weitbrecht et al., 2008].

Literature studies have already pointed out some typical patterns of flow structure around and inside a dead zone which are depending on the aspect ratio W/L [Booij, 1989; Kimura and Hosoda, 1997; Uijttewaal, 1999; Muto et al., 2000b; Uijttewaal et al., 2001; Brevis et al., 2006; Jamieson and Gaskin, 2007]. The flow inside a dead zone is a type of shear-driven cavity flow [Shen and Floryan, 1985; Shankar and Deshpande, 2000], where the flow separates at the upstream corner of the cavity and forms a wake, like behind a backward facing step, in the lee of the upstream end of the cavity. In turbulent flow, the wake may be bounded by a plane mixing layer (ML), which is the shear flow formed in the region between two co-flowing streams of different velocities [Raupach et al., 1996]. Due to the velocity difference, this region is characterized by a strong inflection in the mean velocity profile which gives rise to hydrodynamic instability processes resulting in the development of the mixing layer. Large scale turbulent structures with the vorticity aligned with the main flow vorticity are continuously fed from the main flow. These structures determine a turbulence length scale in the order of the width of the mixing layer, which develops in self-preserving way with a constant spreading rate, depending only on the relative velocity difference across the mixing layer [Booij and Tukker, 2001]. Depending on the dead zone aspect ratio, this wake may entirely fill the dead zone as a primary eddy, but for low aspect ratio, secondary or more eddies start to develop. All those eddies are mainly driven by the main stream flow. Particularly, if the aspect ratio *W/L* is in the range from 0.5 to 1.5 only the primary recirculating eddy is present inside the dead zone and it covers almost the complete area of the dead zone. If *W/L<0.5*, a second eddy develops in the upstream corner of the dead zone. This secondary gyre has no direct contact with the main stream so it is driven by momentum exchange with the primary eddy. Thus, flow velocity in this area is quite slow. With even smaller aspect ratio, that is *W/L<0.2*, the two eddies remain but the main flow starts to penetrate into the dead zone [Uijttewaal, 1999]. On the other hand, if *W/L>1.5*, a second eddy develops behind the primary in transverse direction to the main stream.

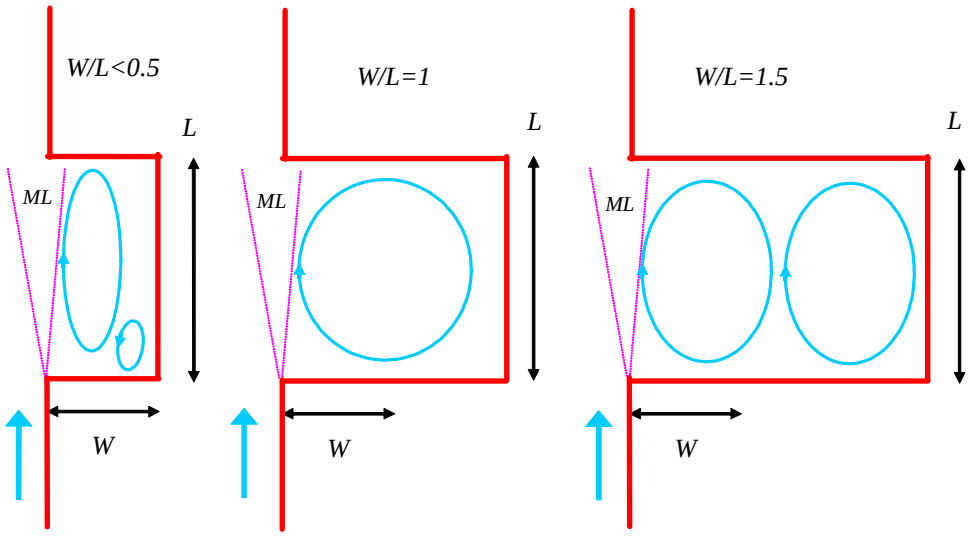


Figure 2. Influence of aspect ratio W/L in flow field in a dead zone

Experimental observations with Particle Image Velocimetry (PIV), Laser Induced Fluorescence (LIF) and dye methods of flow dynamics confirmed that large-scale coherent structures were generated at the head of the upstream wall of the dead zone and were growing as they were advected within the mixing layer governing mass exchange across the interface between the dead zone and the main stream [Uijttewaal et al., 2001; Weitbrecht and Jirka, 2001b; Jamieson and Gaskin, 2007]. Since a river can be considered to have shallow flow, that is horizontal length scale dominates on vertical length scale, these coherent structures are expected to be mainly two-dimensional. However, some experimental studies pointed out that the exchange process is not uniform over the water depth and the exchange rate increases towards the water surface [Weitbrecht and Jirka,

2001a; Brevis et al., 2006], whereas intermittent vertical structures were observed [Jamieson and Gaskin, 2007]. Recent experimental studies addressed the role in the exchange process of Kelvin-Helmholtz instabilities occurring at the dead zone-main stream interface [Brevis et al., 2006].

Coming back to the dimensionless exchange coefficient *k*, it was recognized in some laboratory experimental works that in dead zone with *W/L<0.5* the exchange between the dead zone and the main stream was a two stage process, where two different time scales could be observed [Booij, 1989; Uijttewaal et al., 2001; Engelhardt et al., 2004; Jamieson and Gaskin, 2007]. Some authors ascribed this two stage structure to an initial, fast exchange between the primary eddy and the main stream followed by an exchange between the primary eddy and the smaller, secondary eddy in the upstream corner of the dead zone [Booij, 1989; Uijttewaak et al., 2001]. Other authors suggested a model with a fast exchange for the mixing layer and a slower exchange between the primary eddy and its relatively standing core [Engelhardt et al., 2004; Jamieson and Gaskin, 2007].

3. NUMERICAL SIMULATIONS

As above outlined, the exchange process between a dead zone and the main stream is mainly governed by two-dimensional large-scale coherent structures and flow structure inside and around the dead zone may be considered as two-dimensional [Muto et al., 2000a; Uijttewaal et al., 2001; Weibrecht, 2004]. Thus, two-dimensional or depth-averaged models may be applied to describe hydrodynamics and mass-transfer processes. These computational fluid dynamics (CFD) models are based on the mass conservation equation and the Navier-Stokes equations of motion. Since the flow in the tank is turbulent, these equations must be averaged over a small time increment applying Reynolds decomposition, which results in the Reynolds-averaged Navier-Stokes equations (RANS). For a planar, incompressible flow these equations are:

$$\frac{\partial \overline{u}}{\partial x}+\frac{\partial \overline{v}}{\partial y}=0 \tag{8}$$

$$\begin{aligned}\overline{\rho}\left(\frac{\partial \overline{u}}{\partial t}+\overline{u}\frac{\partial \overline{u}}{\partial x}+\overline{v}\frac{\partial \overline{u}}{\partial y}\right)&=\overline{\rho}\,g_x-\frac{\partial \overline{p}}{\partial x}+\left(\mu+\rho\,\nu_t\right)\nabla^2\overline{u}\\ \overline{\rho}\left(\frac{\partial \overline{v}}{\partial t}+\overline{u}\frac{\partial \overline{v}}{\partial x}+\overline{v}\frac{\partial \overline{v}}{\partial y}\right)&=\overline{\rho}\,g_y-\frac{\partial \overline{p}}{\partial y}+\left(\mu+\rho\,\nu_t\right)\nabla^2\overline{v}\end{aligned} \tag{9}$$

where ρ and μ are fluid density and viscosity, p is fluid pressure and u, v are velocity components in the x and y directions, respectively. The overbar indicates time-averaged quantities. Notably, in eq. (9) there is the turbulent kinematic viscosity ν_t, that if isotropic turbulence assumption holds could be estimated following the k-ε model approach as:

$$\nu_t=\frac{C_\mu\,k'^2}{\varepsilon'} \tag{10}$$

where k' and ε' are turbulent kinetic energy per mass unit and its dissipation rate, respectively, and $C_\mu=0.09$. These parameters are estimated with the classical two equations of k-ε model:

$$\underbrace{\frac{\partial k'}{\partial t}}_{\text{Change in } k'}+\overline{\overline{V}}\cdot\nabla k'=\underbrace{\frac{\nu_t}{\sigma_k}\nabla k'}_{\text{Flux}}-\underbrace{\varepsilon'}_{\text{Dissipation}}+\underbrace{2\,\nu_t\,\overline{\overline{D}}\,\overline{\overline{D}}}_{\text{Production}} \tag{11}$$

$$\frac{\partial \varepsilon'}{\partial t}+\nabla\cdot\varepsilon'\overline{\overline{V}}=\frac{\nu_t}{\sigma_\varepsilon}\nabla\varepsilon'+2\,\nu_t\,C_{1\varepsilon}\frac{\varepsilon'}{k'}\overline{\overline{D}}\,\overline{\overline{D}}-C_{2\varepsilon}\frac{\varepsilon'^2}{k'} \tag{12}$$

where $\overline{\overline{D}}$ is deformation tensor, whereas C_μ, σ_k, σ_ε, $C_{1\varepsilon}$ and $C_{2\varepsilon}$ are constants, and their values are listed in Table 1.

Table 1. Values of the constants of *k-ε* model

C_μ	σ_k	σ_ε	$C_{1\varepsilon}$	$C_{2\varepsilon}$
0.09	1.00	1.60	0.1256	1.92

These equations were solved using Multiphysics 3.3™ modeling package, which is a commercial multiphysics modeling environment (Multiphysics, 2006). Multiphysics 3.3™ can solve for the same flow domain both motion equations and advection-diffusion equation. Particularly, both the *k-ε* model application mode and the advection-diffusion application mode were used. They solve eqs. from (8) to (12) for the pressure $\bar{p}$, the velocity vector components $\bar{u}$ and $\bar{v}$, and *k-ε* model parameters within the domain of the flow (Multiphysics, 2006). Multiphysics 3.3™ was applied to three different geometries representing dead zones, that is a single rectangular embayment and a groyne field with 10 and 5 groynes, respectively. These geometries were those previously experimentally studied by Muto et al. [2000a; 2000b], Wallast et al. [1999], and Weitbrecht and Jirka [2001b]. The studied geometries are presented in Fig.3.

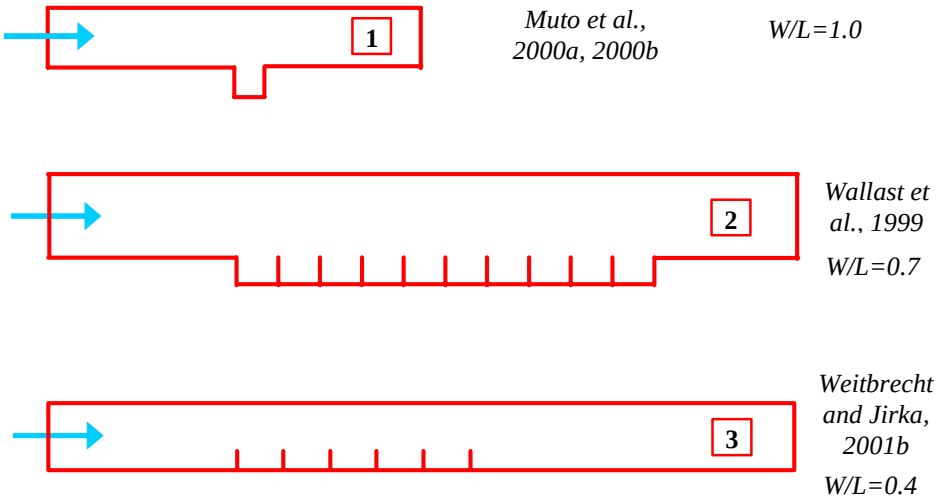


Figure 3. Dead zone geometries investigated in this study

For the simulations, *inflow* boundary condition was applied at the inlet, with uniform velocity profile and a fixed value for *k'* and *ε'* parameters, whereas *neutral* boundary condition was applied at the outlet. Finally, *logarithmic law of the wall* boundary condition was applied to the remaining boundaries, that is at walls and dead zones baffles.
Table 2 lists the main input parameters of the performed simulations.

Table 2. Characteristics of the performed simulations

Geometry	W – m	L – m	W/L	U_{in} – m/s	Reference
1	0.16	0.16	1.00	0.37	Muto et al., 2000a, 2000b
2	0.75	1.12	0.67	0.35	Wallast et al., 1999
3	0.50	1.25	0.40	0.16	Weitbrecht and Jirka, 2001b

The *k-ε* model application mode uses Lagrange *p2-p1* elements to stabilize the pressure. Thus, second order Lagrange elements model the velocity components and *k'* and *ε'* parameters while linear elements model the pressure. The default element settings in this application mode always provide one order higher Lagrange elements for the velocity components than for the pressure.

For the simulations water with density ρ=1000 Kg/m³ and molecular viscosity $\mu=1.00\cdot10^{-3}$ Kg/m×s, was selected as fluid. The mesh generation process was made assuming, among the others, as maximum element size scaling factor, element growth rate and mesh

curvature factor 1, 1.3 and 0.3, respectively. Different values for the maximum element size were selected for the main stream and the dead zone. Finer values were selected in the dead zones and in the mixing layer, where errors tend to be large due to significant velocity gradients. For geometry n.1, maximum element sizes were 0.5 and 0.05 m in the main stream and in the dead zones, respectively. For geometry n.2, maximum element sizes were 0.5 and 0.075 m in the main stream and in the dead zones, respectively, whereas a maximum element size of 0.05 m was assigned at the dead zone-main channel interface. For geometry n.3, maximum element sizes were 0.02, 0.01, 0.005 and 0.002 m in the main stream, at the channel walls, in the dead zone and at the dead zone-main channel interface, respectively. Therefore, the mesh for geometry n.1, n.2 and n.3 has 9450, 10172 and 10434 elements, respectively, with a minimum element quality of 0.7040, 0.6802, and 0.7090, respectively. The number of degrees of freedom were 101695, 109320, and 112456, respectively. About the solver settings, for steady-state analysis, stationary non-linear solver with direct linear system solver was used, where the relative tolerance and the maximum number of iterations were set to $1.0 \cdot 10^{-6}$ and 45, respectively. Streamline diffusion was introduced too.

4. ANALYSIS OF RESULTS. DISCUSSION

Numerical simulations provided field velocity and pressure, kinematic viscosity ν_t, k' and ε' values throughout the flow domain. Fig.4 presents measured (left) and simulated (right) velocity values (surface colour) and velocity vectors (arrows) in the square cavity experimentally studied by Muto et al. (2000a, 2000b), where the flow in the main stream is from the left. Inside the cavity a recirculating primary eddy was observed and its core with very low velocities (blue) is located around the midpoint of the cavity. Higher velocities can be observed in the mixing layer developing at the interface between the cavity and the main channel. Analysis of flow field demonstrated that experimental flow patterns were fairly reproduced by numerical results.

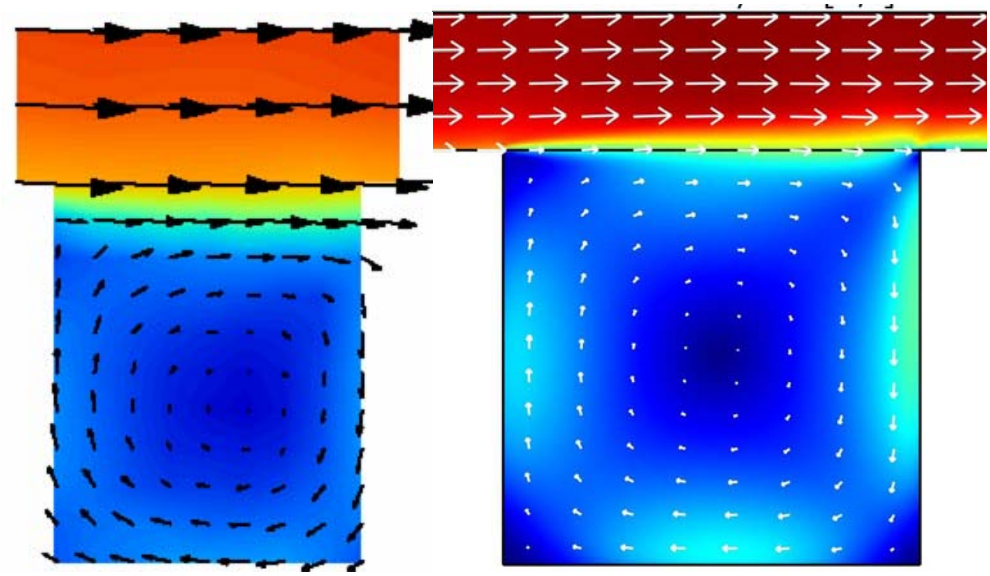


Figure 4. Measured (left) and simulated (right) velocity values and velocity vectors in the square cavity, that is geometry n.1

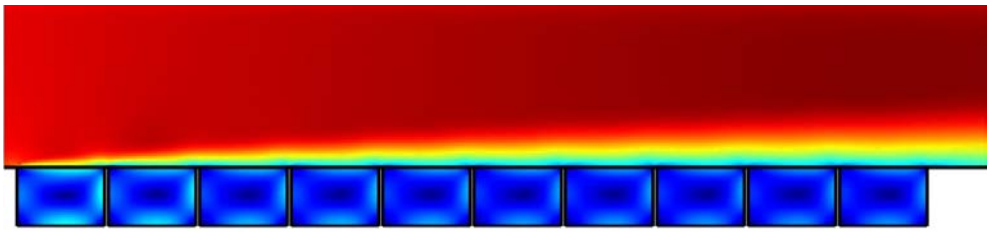


Figure 5. Simulated velocity values in geometry n.2

In the groyne field geometries, flow patterns are more complex. In the geometry studied by Wallast et al. (1999), with 10 groyne fields and a dead zone aspect ratio W/L equal to 0.7, the mixing layer develops from the upstream wall of the first groyne down to last groyne (Fig.5). Fig.6 presents the magnitude of the velocity field and velocity vectors in the main channel and in the groyne fields n.1 and n.2. A primary recirculating eddy is again observed but its relatively stagnant core is located in the downstream part of the dead zone.

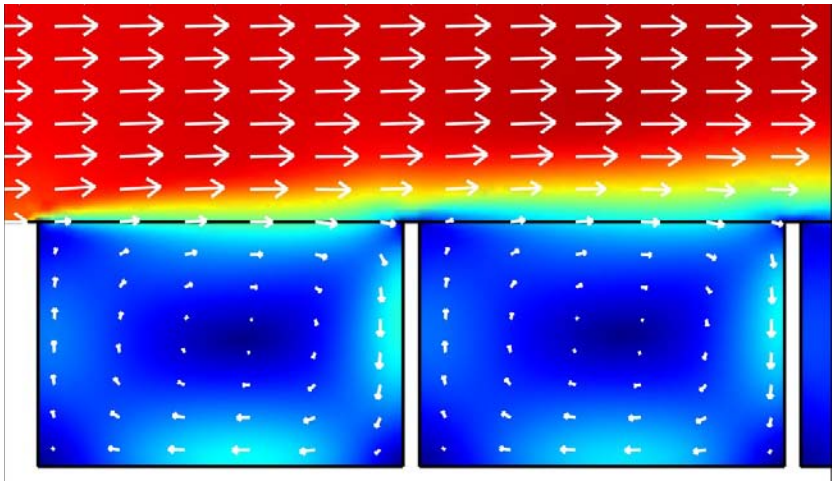


Figure 6. Simulated velocity values and velocity vectors for the groyne field n.1 and n.2

The development of the mixing layer was also observed in the geometry studied by Weitbrecht and Jirka (2001b), but in the first groyne field the primary eddy was partly located outside the dead zone due to the separation region produced by the upstream wall (Fig.7). The eddy core is located around dead zone-main channel interface. Also, the corner eddy was not observed in the numerical results.

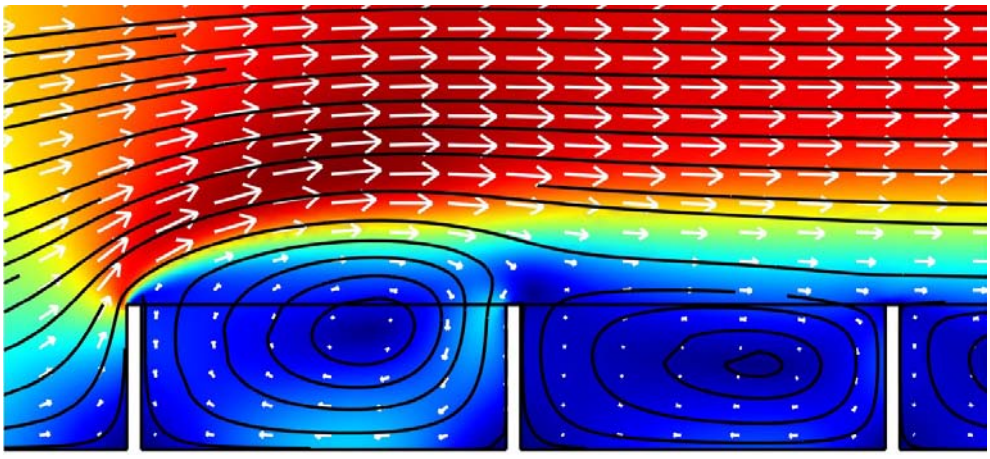


Figure 7. Simulated flow velocity values with streamlines and velocity vectors in geometry n.3

Fig.8 shows a comparison between transverse velocity along the dead zone-cross-main channel interface of the groyne field n.2 predicted by Multiphysics 3.3™ and the field data for the geometry n.3, where L is the length of the dead zone. Negative and positive values represented the exchange processes between the dead zone and the main stream highlighting a reasonable agreement between experimental data and numerical results.

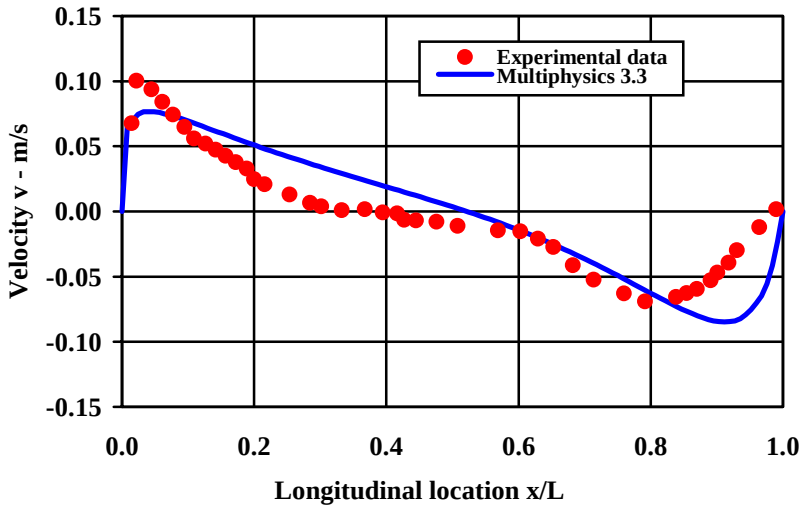


Figure 8. Transverse velocities along the dead zone-main channel interface in geometry n.3

Further comparison between experimental observations and numerical simulations was carried out using the values of dimensionless exchange coefficient k, which as above stated represents the rate of momentum and mass exchange between the dead zone and the main stream. This parameter was determined using the transverse velocity data along the dead zone-main channel interface and the procedure proposed by Weitbrecht and Jirka (2001b). That is, the norm of transverse velocity v_i was averaged over the interface, resulting in the total specific volume flux through the exchange interface. Division by 2 yielded the total specific out flux into the main channel:

$$\overline{E} = \frac{1}{2n}\sum_{i=1}^{n}|v_i| \tag{13}$$

which provides the exchange coefficient K_D and the dimensionless exchange coefficient k through (6) and (5a), respectively. For geometries n.2 and n.3, these parameters were obtained averaging their values in the 10 and 5 dead zones, respectively. Simulated k values were in the range from 0.0365 and 0.0121, which agrees with the literature available experimental data. Table 3 presents a successful comparison between numerical results and experimental data for k parameter.

Table 2. Characteristics of the performed simulations

		Multiphysics	Experiments	Reference
Geometry	W/L	k	k	
1	1.00	0.0121	0.0116	Muto et al., 2000a, 2000b
2	0.67	0.0160	0.0150	Wallast et al., 1999
3	0.40	0.0365	0.0340	Weitbrecht and Jirka, 2001b

As above noted, literature experimental data demonstrated that dimensionless exchange coefficient decreased with the increasing aspect ratio W/L of the dead zone [Muto et al., 2000a, 2000b; Uijttewaal et al., 2001; Weitbrecht and Jirka, 2001a; Weitbrecht et al.,

2008]. Thus, numerical values were plotted against the aspect ratio *W/L* and compared with the corresponding experimental data (Fig.9). Overall, numerical results were in good agreement with the experimental values and confirmed the trend with *W/L* already observed in previous literature studies.

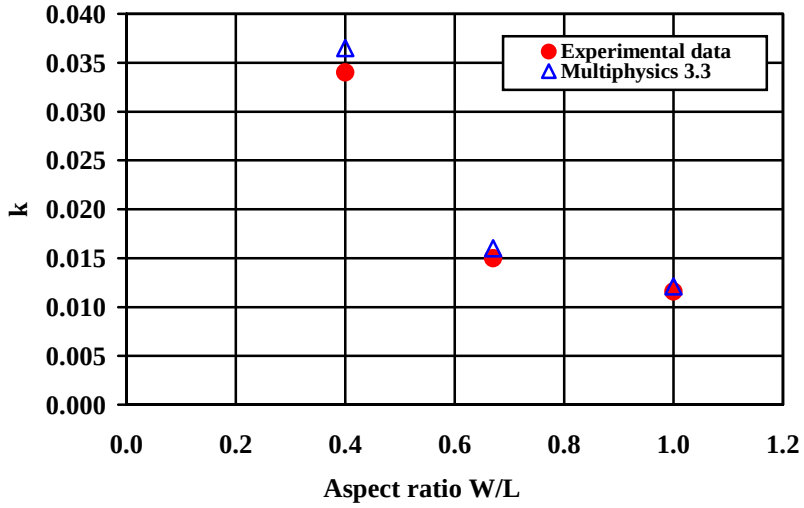


Figure 9. Dimensionless exchange coefficient *k* vs dead zone aspect ratio *W/L*

5. CONCLUSIVE REMARKS

The presence of dead zones in streams and rivers significantly affect the characteristics of mass transport. In a river, dead zones can be due to geometrical irregularities in the riverbanks and the riverbed and/or to spur dikes and groyne fields and they produce a difference between the concentration curves measured and modeled by the classical 1D advection-diffusion equation with sharper front and longer tails. In a dead zone, the mean flow velocity in the main stream direction is essentially zero and the main transport mechanism is transverse turbulent diffusion which controls the exchange processes of solutes with the main stream.

This paper presented some preliminary results of a numerical 2D study undertaken to investigate the fundamental flow phenomena around and inside a simplified geometry, representative of flow in natural rivers with dead zones. Flow patterns and exchange rates between the dead zone and the channel were simulated resulting in a good agreement with previous literature experimental values for the same geometry. The considered geometries were those experimentally studied by Muto et al. [2000a, 2000b], Wallast et al. [1999], and Weitbrecht and Jirka [2001b]. Future research will be addressed to simulate concentration field in a dead zone and to investigate the effect of the inclination of the dead zone to the main flow direction on the exchange processes.

REFERENCES

Booij, R., and Tukker. J., Integral model of shallow mixing layers, *Journal of Hydraulic Research, 39 (2)*, 169-179, 2001

Brevis, W., Niño, Y., and Vargas, J., Experimental characterization and visualization of mass exchange process in dead zones in rivers, *Proceedings of 3rd International Conference on Fluvial Hydraulics (RiverFlow 2006)*, 235-242, 2006

Cushman-Roisin, B. (2007). Environmental fate and transport, Lecture Notes, Thayer School of Engineering, Dartmouth College, NH, USA

Czernuszenko, W., Rowiński, P.M., and Sukhodolov, A., Experimental and numerical

validation of the dead-zone model for longitudinal dispersion in rivers, *Journal of Hydraulic Research, 36 (2)*, 269-280, 1998

Le Coz, J., Brevis, W., Niño, Y., Paquier, A., and Riviére, N., Open-channel side-cavities: A comparison of field and flume experiments, *Proceedings of 3rd International Conference on Fluvial Hydraulics (RiverFlow 2006)*, 235-242, 2006

Jamieson, E., and Gaskin, S.J., Laboratory study of 3 dimensional characteristics of recirculating flow in a river embayment, *Proceedings of the XXXII IAHR Congress*, 2007

Kurzke, M., Weitbrecht V., and Jirka G.H.J., Laboratory concentration measurements for determination of mass exchange between groin fields and main stream. *Proceedings of 1st International Conference on Fluvial Hydraulics River Flow 2002*, vol.1, 369-376, 2002

Multiphysics 3.3 *User's Guide*, ComSol AB. Sweden, 2007

Muto, Y., Imamoto, H., and Ishigaki, T., Velocity measurements in a straight open channel with a rectangular embayment. *Proceedings of the 12th APD-IAHR, Bangkok, Thailand*, 2000

Raupach. M.R., Finnigan, J.J., and Brunet, Y., Coherent eddies and turbulence in vegetation canopies: the mixing-layer analogy, *Boundary-Layer Meteorology, (78)*, 351-382, 1996

Shankar, P.N., and Deshpande, M.D., Fluid mechanics in the driven cavity, *Annual Review in Fluid Mechanics*, vol.32, 93-136, 2000

Shen, C., and Floryan, J.M., Low Reynolds number flow over cavities, *Physics of Fluids A*, Vol.28, n.11, November 1985, 3191-3202

Uijttewaal, W.S.J., Lehmann, D., and Mazijk, A. van, Exchange processes between a river and its groyne fields: model experiments, *Journal of Hydraulic Engineering, ASCE, 127 (11*), 928-936, 2001

Uijttewaal, W.S.J., Effects of groyne layout on the flow in groyne fields: laboratory experiments, *Journal of Hydraulic Engineering, ASCE, 131 (9)*, 782-791, 2005.

Valentine, E.M., and Wood, I.R., Longitudinal dispersion with dead zones, *Journal of Hydraulics Division, ASCE, 103 (9)*, 975-990, 1977

Valentine E.M., and Wood I.R., Experiments in longitudinal dispersion with dead zones, *Journal of Hydraulics Division, ASCE, 105 (8)*, 999-1016, 1979

Wallast, I., Uijttewaal, W., and Mazijk, A. van, Exchange processes between groyne field and main stream. *Proceedings of the XXVIII IAHR Congress*, 1999

Weitbrecht, V., and Jirka, G.H.J., Flow patterns and exchange processes in dead zones of rivers, *Proceedings of the XXIX IAHR Congress*, 2001a

Weitbrecht, V., and Jirka, G.H.J., Flow patterns in dead zones of rivers and their effect on exchange processes, *Proceedings of the 2001 International Symposium on Environmental Hydraulics*, 2001b

Weitbrecht, V., Influence of dead-water zones on the dispersive mass-transport in rivers, *Dissertation*, 2004

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Spatial Development of Viscous Flow Induced by Wave Propagation over Vortex Ripples

A. A. Dimas

Laboratory of Hydraulic Engineering, Department of Civil Engineering, University of Patras, Patras 26500, Greece (adimas@upatras.gr)

Abstract: In the present study, numerical simulations of the free-surface flow, developing by the propagation of nonlinear water waves over a rippled bottom, are performed assuming that the corresponding flow is two-dimensional, incompressible and viscous. The simulations are based on the numerical solution of the unsteady, two-dimensional, Navier-Stokes equations subject to the fully-nonlinear free-surface boundary conditions and appropriate bottom, inflow and outflow boundary conditions. The equations are properly transformed so that the computational domain becomes time-independent. For the spatial discretization, a hybrid scheme is used, in which the finite-difference method, in the horizontal direction, and a pseudo-spectral approximation method with Chebyshev polynomials, in the vertical direction, are applied. A fractional time-step scheme, which consists of three different stages, is used for the temporal discretization. A damping zone is placed at the outflow region in order to efficiently minimize reflection of waves by the outflow boundary. Over the rippled bed, the wave boundary layer thickness increases significantly, while flow separation at the ripple crests generates a circulation region. The amplitude of the wall shear stress over the ripples increases with increasing ripple height, while the corresponding friction force is insensitive to this increase. The amplitude of the form drag forces due to dynamic and hydrostatic pressures also increase with increasing ripple height, therefore, the percentage of friction in the total drag force decreases with increasing ripple height.

Keywords: Vortex Ripples; Wave Propagation; Viscous Flow; Free Surface; Numerical Simulation.

1. INTRODUCTION

The sediment transport in the coastal zone results into the formation of sand ripples along the beach bottom. The presence of a rippled bed modifies the propagation of water waves and the development of the wave boundary layer. Wiberg and Nelson [1992] studied experimentally the unidirectional flow over rippled bed and developed a flow model for the prediction of the mean velocity profiles and drag coefficients on each ripple. Ranasoma and Sleath [1994] studied experimentally the combined current and oscillatory flow over ripples, considering a rigid lid approximation for the free surface, and concluded that the velocity profiles do not collapse to a single curve. Fredsøe et al. [1999] studied experimentally and numerically the combined current and oscillatory flow over rippled bed and concluded that the velocity profiles comprise of two layers; one associated to the ripple height and the other to the wave motion. Huang and Dong [2002] studied numerically the wave propagation over rippled bed, utilizing the finite-analytic (FA) method, and concluded that the periodically averaged flow exhibits a current in the opposite direction of the wave propagation. Barr et al. [2004] studied numerically and experimentally the oscillatory flow over sand ripples, considering a rigid lid approximation for the free surface, and concluded that the wave boundary layer over rippled bed increases, while the average shear stress on the wall decreases, with ripple steepness. Chang and Scotti [2004]

studied numerically the unidirectional and the oscillatory flow over sand ripples, considering a rigid lid approximation for the free surface, and concluded that the large-eddy simulation (LES) performs better than the Reynolds-Averaged Navier-Stokes (RANS) k-ω model, which underestimates the Reynolds stress and overestimates the amplitude of the vertical velocity oscillations. Malarkey and Davies [2004] studied numerically the oscillatory flow over sand ripples, utilizing a discrete vortex method, and concluded that the flow in the lower part of the oscillating boundary layer is dominated by the process of vortex formation and shedding due to flow separation.

Typical sand ripples have parabolic shape, while their dimensions, length L_r and height h_r, depend on wave period, wave height and water depth according to laboratory and field data [Fredsøe and Deigaard, 1992]. In the present study, the case of wavelength to water depth ratio $\lambda / d_0 = 6$ and wave height to wavelength ratio $H_0 / \lambda = 0.05$ is considered which is consistent with $L_r / d_0 \approx 0.25$ and $0.02 \le h_r / d_0 \le 0.05$. The objective is to simulate the corresponding free-surface flow, considering a rigid rippled bed, and study the behaviour of shear-stress and pressure distributions along the bed, as well as the corresponding friction and form drag forces. In the following sections, the formulation, the numerical method and results are presented for the propagation of the above mentioned wave considering the two extreme values of ripple height $h_r / d_0 = 0.02$ and 0.05.

2. FORMULATION

The incompressible, viscous, two-dimensional, free-surface flow is governed by the continuity

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} = 0 \tag{1}$$

and the Navier-Stokes equations

$$\frac{\partial u_1}{\partial t} + u_1 \frac{\partial u_1}{\partial x_1} + u_2 \frac{\partial u_1}{\partial x_2} = -\frac{\partial p}{\partial x_1} + \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 u_1}{\partial x_1^{\,2}} + \frac{\partial^2 u_1}{\partial x_2^{\,2}} \right) \tag{2}$$

$$\frac{\partial u_2}{\partial t} + u_1 \frac{\partial u_2}{\partial x_1} + u_2 \frac{\partial u_2}{\partial x_2} = -\frac{\partial p}{\partial x_2} + \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 u_2}{\partial x_1^{\,2}} + \frac{\partial^2 u_2}{\partial x_2^{\,2}} \right) \tag{3}$$

where t is time, x_1 and x_2 are the horizontal and vertical coordinates, respectively; u_1 and u_2 are the horizontal and vertical velocity components, respectively; p is the dynamic pressure and Re is the Reynolds number. The equations (1)-(3) are in dimensionless form utilizing the inflow depth d_0, the gravity acceleration g and the fluid density ρ, therefore, $\mathrm{Re} = d_0 (g d_0)^{1/2} / \nu$ where ν is the kinematic fluid viscosity.

The kinematic, dynamic and stress boundary conditions at the free surface, respectively, are

$$u_2 = \frac{d\eta}{dt} = \frac{\partial \eta}{\partial t} + u_1 \frac{\partial \eta}{\partial x_1}, \quad p = \frac{\eta}{\mathrm{Fr}^2} \quad \text{and} \quad \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right)\left(1 - \left(\frac{\partial \eta}{\partial x_1} \right)^2 \right) + 4 \frac{\partial u_2}{\partial x_2} \frac{\partial \eta}{\partial x_1} = 0 \tag{4}$$

where η is the free surface elevation and Fr is the Froude number, which under the present dimensionless formulation is equal to one. In addition, the no-slip and non-penetration boundary conditions at the bottom, respectively, are

$$u_1 - u_2 \frac{\partial d}{\partial x_1} = 0 \quad \text{and} \quad u_2 + u_1 \frac{\partial d}{\partial x_1} = 0 \tag{5}$$

where $d = d\left(x_1\right)$ is the bottom depth measured from the still free surface level.

Given that the free surface is time-dependent, the Cartesian coordinates are transformed, in order for the computational domain to become time-independent, according to the expressions

$$s_1 = x_1 \quad \text{and} \quad s_2 = \frac{2x_2 + d - \eta}{d + \eta} \tag{6}$$

According to (6), in the transformed domain, the free surface corresponds to $s_2 = 1$, while the bottom to $s_2 = -1$. The velocity components are also transformed in the following way

$$u_1 = v_1 \quad \text{and} \quad u_2 = v_2 + \frac{r}{2} v_1 = v_2 + v_\eta \tag{7}$$

where

$$r = \frac{\partial \eta}{\partial s_1}\left(1 + s_2\right) + \frac{\partial h}{\partial s_1}\left(1 - s_2\right) \tag{8}$$

and $\partial h / \partial s_1 = -\partial d / \partial s_1$ is the bottom slope.

Taking into account (6) and (7), the transformed continuity equation and the Navier-Stokes equations in rotational form, respectively, are

$$\frac{\partial v_1}{\partial s_1} + \frac{2}{d + \eta}\left[\frac{\partial v_2}{\partial s_2} + \frac{v_1}{2}\left(\frac{\partial \eta}{\partial s_1} - \frac{\partial h}{\partial s_1}\right)\right] = 0 \tag{9}$$

$$\frac{\partial v_1}{\partial t} = v_2 \zeta + \frac{1 + s_2}{d + \eta}\frac{\partial \eta}{\partial t}\frac{\partial v_1}{\partial s_2} + \frac{r}{d + \eta}\frac{\partial p}{\partial s_2} - \frac{\partial \Pi}{\partial s_1} + \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 v_1}{\partial s_1^{\,2}} + \frac{r^2 + 4}{\left(d + \eta\right)^2}\frac{\partial^2 v_1}{\partial s_2^{\,2}} + A_1\right) \tag{10}$$

$$\frac{\partial v_2}{\partial t} = -v_1 \zeta + \frac{1 + s_2}{d + \eta}\frac{\partial \eta}{\partial t}\frac{\partial \left(v_2 + v_\eta\right)}{\partial s_2} - a_\eta - \frac{2}{d + \eta}\frac{\partial \Pi}{\partial s_2} + \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 \left(v_2 + v_\eta\right)}{\partial s_1^{\,2}} + \frac{r^2 + 4}{\left(d + \eta\right)^2}\frac{\partial^2 \left(v_2 + v_\eta\right)}{\partial s_2^{\,2}} + A_2\right) \tag{11}$$

where $A_i = -\frac{2r}{d + \eta}\frac{\partial^2 u_i}{\partial s_1 \partial s_2} - \left(\frac{1 + s_2}{d + \eta}\frac{\partial^2 \eta}{\partial s_1^{\,2}} + \frac{1 - s_2}{d + \eta}\frac{\partial^2 h}{\partial s_1^{\,2}} - \frac{2r}{\left(d + \eta\right)^2}\left(\frac{\partial \eta}{\partial s_1} - \frac{\partial h}{\partial s_1}\right)\right)\frac{\partial u_i}{\partial s_2}$,

$a_\eta = \frac{\partial v_\eta}{\partial t} + v_1 \frac{\partial v_\eta}{\partial s_1} + v_2 \frac{2}{d + \eta}\frac{\partial v_\eta}{\partial s_2}$, $\Pi = p + \frac{1}{2}\left(v_1^{\,2} + v_2^{\,2}\right)$ is the transformed dynamic pressure head and $\zeta = \frac{\partial v_2}{\partial s_1} - \frac{2}{d + \eta}\frac{\partial v_1}{\partial s_2}$ is the transformed vorticity.

Finally, the free surface boundary conditions, obtained by the transformation of (4), are

$$v_2 = \frac{\partial \eta}{\partial t}, \quad \Pi = \frac{\eta}{\mathrm{Fr}^2} + \frac{1}{2}\left(v_1^{\,2} + v_2^{\,2}\right) \quad \text{and} \quad \frac{2}{d+\eta}\left(1+\left(\frac{\partial \eta}{\partial s_1}\right)^2\right)^2 \frac{\partial v_1}{\partial s_2} +$$

$$\left(1-\left(\frac{\partial \eta}{\partial s_1}\right)^2\right)\frac{\partial v_2}{\partial s_1} - 2\frac{\partial \eta}{\partial s_1}\left(1+\left(\frac{\partial \eta}{\partial s_1}\right)^2\right)\frac{\partial v_1}{\partial s_1} + \frac{\partial^2 \eta}{\partial s_1^{\,2}}\left(1-\left(\frac{\partial \eta}{\partial s_1}\right)^2\right) v_1 = 0 \quad (12)$$

while the bottom boundary conditions, obtained by the transformation of (5), are

$$v_1 = 0 \quad \text{and} \quad v_2 = 0\,. \quad (13)$$

3. NUMERICAL METHOD

The numerical solution of the unsteady, two-dimensional, Navier-Stokes equations is based on a fractional time-step scheme, which consists of three stages, for the temporal discretization and a hybrid scheme for the spatial discretization. In the latter, central finite-differences are applied along s_1 and a pseudo-spectral approximation with Chebyshev polynomials along s_2. Along s_1, the spatial interval Δs is constant, while along s_2, the number of collocation points N is equal to the highest order of the Chebyshev polynomials.

At the first stage of each time step, the nonlinear and part of the viscous terms (A_i) of the equations of motion (10) and (11), are treated explicitly using a third-order Adams-Bashforth scheme without imposing any boundary conditions. At the second stage, an implicit Euler scheme is utilized for the treatment of pressure head terms of (10) and (11). At this stage, the generalized Poisson's equation for the transformed pressure head Π is derived by satisfying the continuity equation (9) as well. The dynamic boundary condition (12), at the free surface, and the non-penetration condition (13), at the bottom, are imposed at this stage. The remaining viscous terms are treated implicitly, at the third stage of the time-step, using again an Euler scheme. The velocity components are computed by numerical solution of the corresponding generalized Poisson's equations subject to the stress boundary condition (12), at the free surface, and the boundary conditions (13), at the bottom. Finally, the free-surface elevation is computed by satisfying the kinematic boundary condition (12).

At the inflow boundary, velocity field, dynamic pressure and free-surface elevation are defined according to second order Stokes wave theory. A damping zone is placed at the outflow region in order to efficiently minimize reflection of waves by the outflow boundary [Dimakopoulos and Dimas, 2006].

4. RESULTS

The numerical model is validated by simulating the case of a laminar, oscillatory, current flow, which develops a uniform boundary layer over horizontal bottom of constant water depth $d = d_0$. In this case, the driving velocity is $u = U\sin(\omega t)$ where ω is the radial frequency, and the free surface is modeled as a rigid lid symmetry plane, therefore, $x_2 = 0$ corresponds to the free surface and $x_2 = -1$ to the bottom. The computational domain length is equal to $4d_0$, in the x_1 direction, while the numerical parameters are $\Delta t = 0.0005$, $\Delta s = 0.02$ and $N = 64$. The Reynolds and Strouhal numbers defined with respect to the Stokes length $\delta = (2\nu/\omega)^{1/2}$ are $\mathrm{Re}_\delta = U\delta/\nu = 8$ and $St_\delta = \omega\delta/U = 0.25$, and correspond to $\delta = 0.25\cdot d_0$. The streamwise velocity, obtained by the analytical solution

$$u/U = \sin\left(St\cdot t\right) - e^{-\frac{x_2}{\delta}} \sin\left(St\cdot t - \frac{x_2}{\delta}\right), \tag{14}$$

is in excellent agreement to the numerical solution (Figure 1).

For the simulation of finite-amplitude wave propagation over a rigid rippled bed, a sketch of the computational domain is depicted in Figure 2. As mentioned before, the case with $\lambda / d_0 = 6$ and $H_0 / \lambda = 0.05$ is considered, where $\mathrm{Re} = 250000$ and $\mathrm{Re}_\delta = 80$, while the domain length is set equal to 6.5λ. The shape of the ripples is parabolic

$$h\left(x_1\right) = h_r \left(2\frac{x_1}{L_r} - 1\right)^2 \tag{15}$$

and we consider two cases with ripple length $L_r / d_0 = 0.25$ and heights $h_r / d_0 = 0.02$ and 0.05. The ripple train spans half a wavelength (from $x_1 = 6.125$ to $x_1 = 9.125$) and consists of 12 ripples. The numerical parameters are $\Delta t = 0.002$, $\Delta s = 0.025$ and $N = 128$.

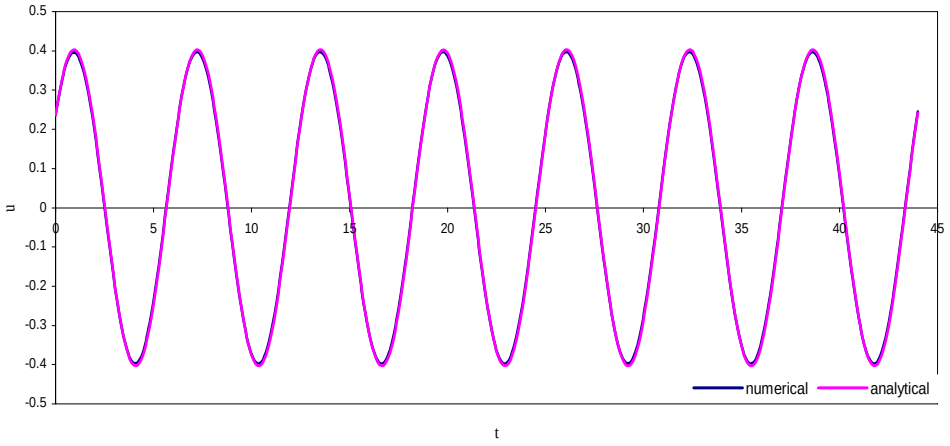


Figure 1. Time evolution of normalized streamwise velocity, at $x_2 = -0.916$, of oscillatory current over horizontal bottom.

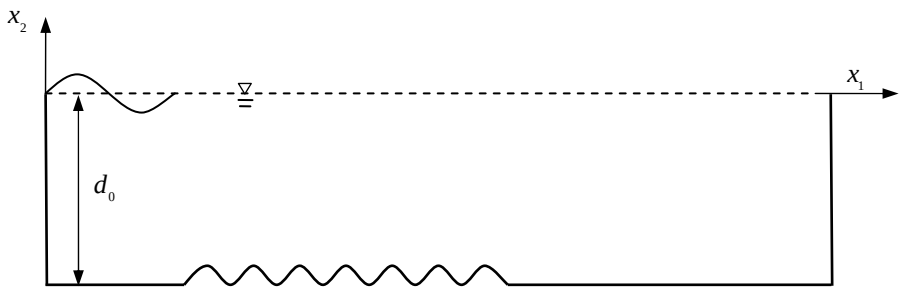


Figure 2. Sketch of computational domain for simulation of wave propagation over rippled bed.

Typical snapshots of pressure, velocity and vorticity fields at time $t = 2.2T$, for the case $h_r / d_0 = 0.05$, are shown in Figures 3 and 4. The presence of the ripples does not alter the wave propagation but affects the wave boundary layer, i.e. its thickness increases to about 5δ, due to flow separation at the ripples crest. The wall shear stress distribution at several time instants during a wave period are shown in Figures 5 and 6 for both cases of ripple height. The presence of the ripples increases the amplitude of the wall stress distribution proportionally to the ripple height, while the boundary layer separation presents a non-symmetric wall stress profile along each ripple. Then, shear stress, dynamic pressure and hydrostatic pressure are integrated along the surface of one ripple, from $x_1 = 7.5$ to $x_1 = 7.75$, to get the horizontal components of the friction force F_f, form drag force F_P and form drag force F_s, respectively, applied on the ripples. The time evolution of these forces is shown in Figure 7. The amplitude of the form drag forces, F_P and F_s, increases with increasing ripple height, while the amplitude of the friction force F_f is insensitive to the increase of ripple height. Therefore, the percentage of friction in the total drag force decreases with increasing ripple height.

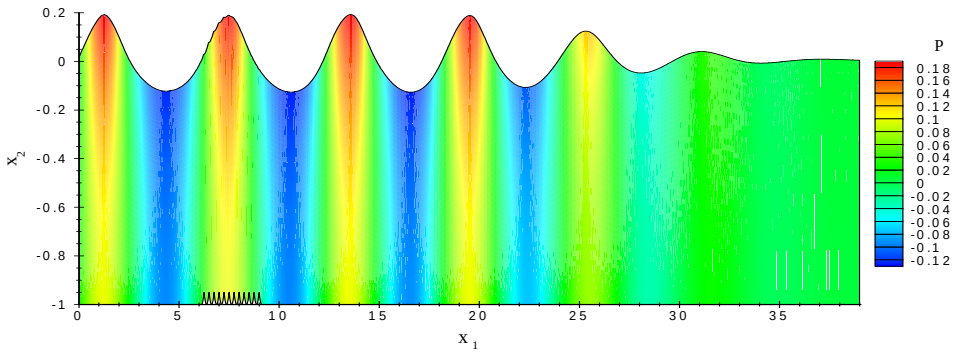


Figure 3. Dynamic pressure p, during wave propagation over rippled bed, for ripple height $h_r / d_0 = 0.05$.

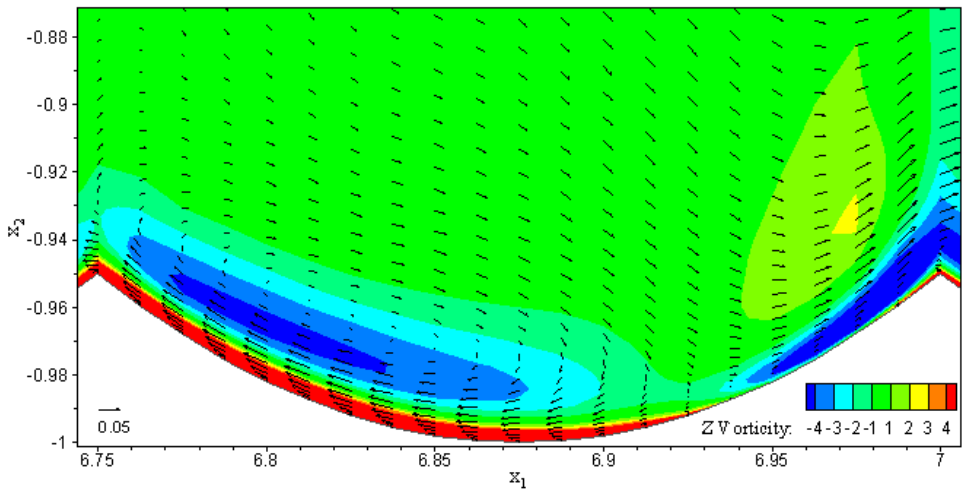


Figure 4. Vorticity and velocity field, close to the bottom between two successive ripple crests, for ripple height $h_r / d_0 = 0.05$.

5. CONCLUSIONS

The propagation of finite-amplitude waves is not but the wave boundary layer is affected by the presence of a rippled bed. Flow separation at the ripple crests generates a circulation region, which alters the wall shear stress. The amplitude of the wall shear stress over the ripples increases with increasing ripple height, while the corresponding friction force is insensitive to this increase. The amplitude of the form drag forces due to dynamic and hydrostatic pressures both increase with increasing ripple height.

ACKNOWLEDGEMENTS

This paper is part of the 03ED617 research project, implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Fund). Funding by the Patras Port Authority is also greatly appreciated.

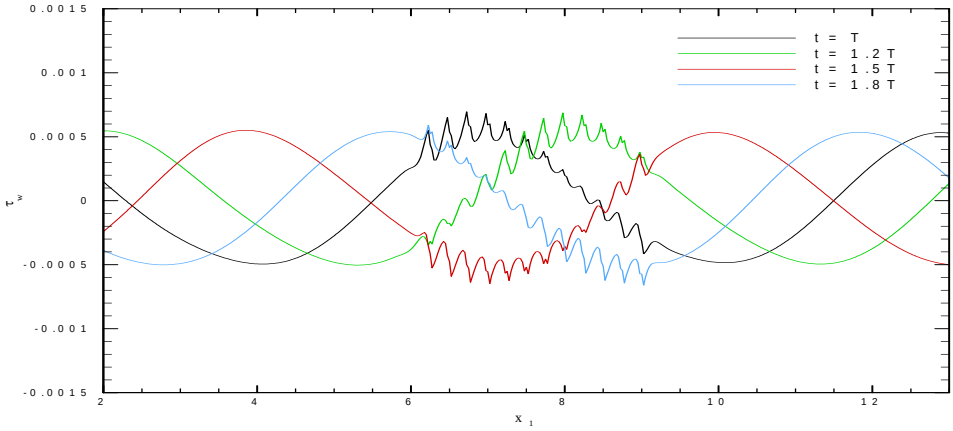


Figure 5. Wall shear stress τ_w, at time instants during a wave period, for ripple height $h_r / d_0 = 0.02$.

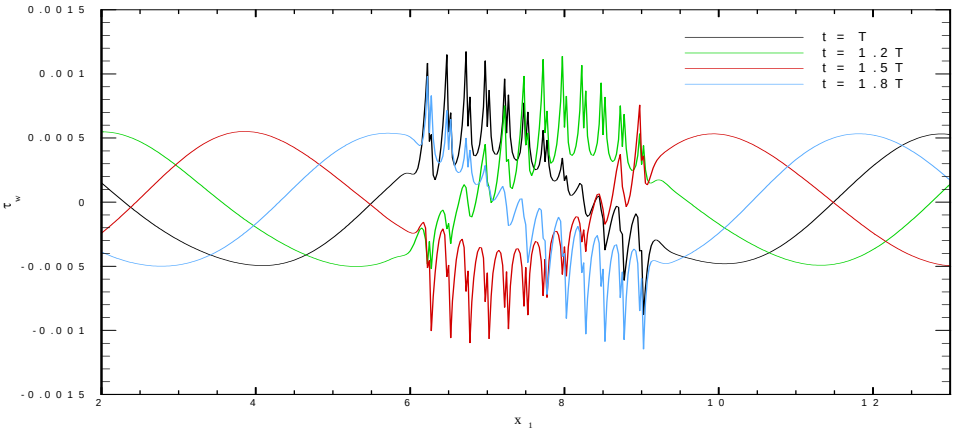


Figure 6. Wall shear stress τ_w, at time instants during a wave period, for ripple height $h_r / d_0 = 0.05$.

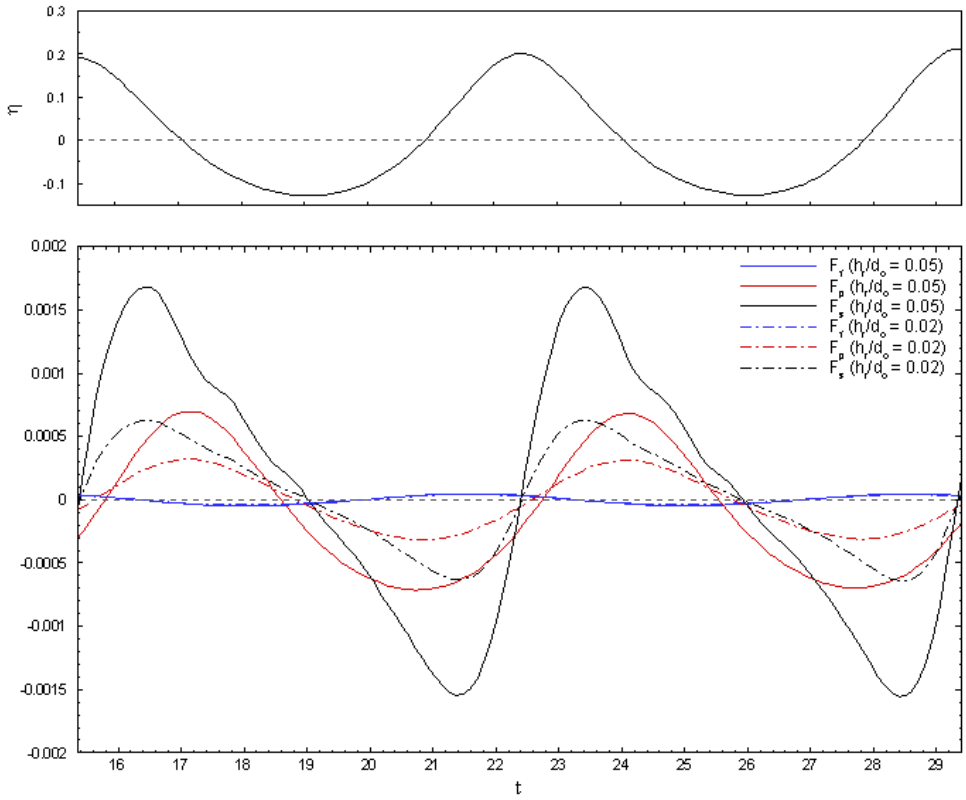


Figure 7. Time evolution of free surface elevation η at ripple valley ($x_1 = 7.625$) and friction force F_f , form drag force F_p due to dynamic pressure, and form drag force F_s due to hydrostatic pressure, over a ripple length L_r ($7.5 \le x_1 \le 7.75$) for Re = 250000 .

REFERENCES

Barr, B.C., D.N. Sinn, T. Pierro, and K.B. Winters, Numerical simulation of turbulent, oscillatory flow over sand ripples, *Journal of Geophysical Research*, 109(C09009), 2004.

Chang, Y.S., and A. Scotti, Modeling unsteady turbulent flow over ripples: Reynolds-averaged Navier-Stokes equations (RANS) versus large-eddy simulation, *Journal of Geophysical Research*, 109(C09012), 2004.

Dimakopoulos, A.S., and A.A. Dimas, Numerical simulation of nonlinear wave propagation and breaking over constant-slope bottom, *Proceedings of 25th International Conference on Offshore Mechanics and Arctic Engineering*, OMAE2006-92163, Hamburg, Germany, 2006.

Fredsøe, J., and R. Deigaard, *Mechanics of coastal sediment transport*, World Scientific, Singapore, 1992.

Fredsøe, J., K.H. Anderson, and B.M. Sumer, Wave plus current over a ripple-covered bed, *Coastal Engineering*, 38, 177-221, 1999.

Huang, C.J., and C.M. Dong, Propagation of water waves over rigid rippled beds, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 128(5), 190-201, 2002.

Malarkey, J., and A.G. Davies, An eddy viscosity formulation for oscillatory flow over vortex ripples, *Journal of Geophysical Research*, 109(C09016), 2004.

Ranasoma, K.I.M., and F.A. Sleath, Combined oscillatory and steady flow over ripples, *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 120(4), 331-346, 1994.

Wiberg, P.L., and J.M. Nelson, Unidirectional flow over asymmetric and symmetric ripples, *Journal of Geophysical Research*, 97(C8), 12745-12761, 1992.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A simplified experimental method to evaluate equivalent roughness of vegetated river beds

S. De Felice[a], P. Gualtieri[a] and G. Pulci Doria[a]

[a] Department of Hydraulic, Geotechnical and Environmental Engineering, University of Naples Federico II, Naples, Italy (sdefelic/pgualtie/pulci@unina.it)

Abstract: In this paper a new method to experimentally evaluate roughness of vegetated river beds is proposed. Unlike the usual method of attaining uniform flow, the new method is based on measurements performed in boundary layer flows, conditions that in flow experiments with plants can be attained easier then in uniform flows. The described method can be applied either in case of smooth or vegetated walls. The friction factor f and Manning's n values so obtained are in agreement with literature results.

Keywords: flow resistance; uniform flow; boundary layer; vegetated river beds.

1. INTRODUCTION

In the past, vegetation on river beds was considered an unwanted source of flow resistance, and for this reason vegetation was commonly removed to improve the water conveyance.

Nowadays, vegetation is regarded as a means for providing stabilization for banks and channels, habitat and food for animals, and pleasing landscapes for recreational use. Therefore the preservation of vegetation is of great relevance for the ecology of water systems.

For this reason, the study of the effects of vegetation on the hydrodynamics of rivers represents one of the most important topics studied by hydraulic engineers today. Characteristics of flows over vegetated surfaces have been deepened with either experimental or numerical methods. Numerous studies have been carried out to examine the flow resistance of the streams on a vegetated bottom, and the main hydrodynamic characteristics of these streams, as the mean flow, the turbulent structures, and sediment transport (Lopez & Garcia 2001, Nezu & Onitsura, 2002, Tsuijimoto 1999, Tsuijimoto & Kitamura, 1990).

2. AIM OF THE PAPER

In a water current, vegetation may be regarded as a kind of bottom roughness that produces a resistance against the flow (Bettess 2003, Kouwen et al. 1969, Kouwen & Unny, 1973, Kouwen & Unny, 1981, Lopez and Garcia, 1997, Stone & Shen, 2002).

The main procedure to experimentally obtain the resistance characteristics of a bottom is to use a very long channel and generate within it a uniform flow. Therefore, long experimental laboratory channels must be realized in order to evaluate resistance characteristics of vegetated river bottoms.

Motivated by this observation, and by authors' experience of boundary layer flow over vegetated beds (De Felice and Gualtieri 2005, Gualtieri and Pulci Doria 2008), a simplified method is proposed that enables determination of the resistance coefficient by investigation

of the boundary layer velocity profile. As the channel length necessary to produce a boundary layer is shorter, therefore, shorter laboratory channels are needed, and therefore this method allows easier determination of the resistance coefficient in a uniform flow.

Therefore, in this paper an experimental modelling of a particular kind of bottom roughness is described and a simplified method to evaluate the resistance coefficient is proposed.

3. GENERALITIES ABOUT RESISTANCE FORMULAS AND COEFFICIENTS

In a recent review (Yen, 2002) it is observed that each resistance coefficient can be considered either as cross section coefficient or as reach one, and that the usual resistance formulas can be considered as reach formulas to apply to uniform flows. It is possible to apply them also to not uniform flows, although to a very short reach, practically to a single cross section: therefore, in this case, the resistance coefficient can vary in the subsequent cross sections of the flow.

Referring to the reach formulas, the most frequently used, relating open-channel flow velocity V to resistance coefficients, are the Darcy-Weisbach, Manning and Chézy ones:

$$V = \sqrt{\frac{8g}{f}}\sqrt{RS}\,; \qquad V = \frac{1}{n}R^{2/3}S^{1/2}\,; \quad \mathrm{V} = \mathrm{C}\sqrt{\mathrm{RS}} \tag{1}$$

where f, n, and C are the Weisbach, Manning and Chézy resistance coefficients, and R = hydraulic radius, S = slope, g = gravitational acceleration. In case of cross section formulas, the slope S must be substituted by the head slope **J.** Comparing these formulas, it is possible to obtain the following expressions:

$$\sqrt{\frac{f}{8}} = \frac{n\sqrt{g}}{R^{1/6}} = \frac{\sqrt{g}}{C} \tag{2}$$

Among these resistance formulas, the authors choose the Darcy-Weisbach approach, because it is the most suitable for an exact evaluation of the flow resistances. Within this approach, the resistance coefficient f is related to the equivalent wall surface roughness, that here will be called ε, through the Colebrook-White formula:

$$\frac{1}{\sqrt{f}} = -K\,log\left(\frac{\varepsilon/4R}{a} + \frac{b}{Re\sqrt{f}}\right) \tag{3}$$

with Re Reynolds number defined as Re = 4VR/ν, and ν kinematic viscosity of the fluid.

The values of the constants a and b have been the object of many experimental surveys. In particular, in (Yen 2002) values for a and b, obtained by other researchers either for open channels with different aspect ratios or for wide open channels, are suggested. In Moody-type diagram, relative to circular full pipes, the values K = 2, a = 3.71 and b = 2.51 are used.

According to Marchi (1961) it is always possible to use the values of Moody-type diagram relative to circular full pipes also for open channels flows. This requires adoption of a suitable shape parameter ψ depending on the aspect ratio, so that the Colebrook-White formula becomes:

$$\frac{1}{\sqrt{f}} = -2\,log\,\frac{1}{\psi}\left(\frac{\varepsilon/4R}{3.71} + \frac{2.51}{Re\sqrt{f}}\right) \tag{4}$$

In particular, ψ assumes the value 0.83 for wide rectangular channels. The values of f so obtained are very similar to those suggested by Yen for wide channels.

4. PREVIOUS AUTHORS EXPERIMENTS ON BOUNDARY LAYER FLOWS

The elaborations performed in this paper are based on experimental data described in (De Felice and Gualtieri 2005, Gualtieri and Pulci Doria 2008).

The experiments were carried out in the plant sketched in figure 1. The main device was a channel 4m long and 15cm (b) wide, with variable slope, with plexiglas walls and bottom, coming out from a feeding tank supplied by a circulation pump, which took water from the drain tank downstream. The tank fed the channel through a rectangular adjustable sluice gate.

In the first sections of the channel, a zero piezometric head gradient boundary layer with horizontal free surface stream was generated. The boundary layer thickness increased in the subsequent sections along the channel, till it reached the same value as the height of the circulating flow, at a distance, from the inlet of the channel, depending on the dynamic characteristics of the flow itself (generally at least 50 cm).

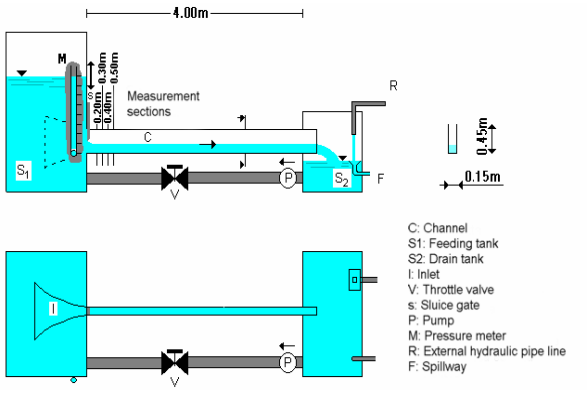


Figure 1. Scheme of the experimental plant.

Measurements were performed either with a smooth or a vegetated bottom. In particular, the vegetation was modelled by means of brass 4mm diameter cylinders with three different heights (5mm, 10mm, 15mm) placed according to two different regular geometries (respectively, a rectangular mesh 5*2.5cm^2 and a square mesh 2.5*2.5cm^2), pointed out synthetically as single and double density. Consequently, the projected area of vegetation per unit volume of water in the flow direction (Tsujimoto *et al.*, 1992) were, respectively, 3.2m^{-1} and 6.4m^{-1}. Combinations of three different heights and two different densities produced six different vegetated bottoms.

In all the seven considered flow conditions (smooth bottom and six different vegetated bottoms), the same experimental setting was considered. In particular, the height of the sluice gate was set at 7.49cm so that the height in the vena contracta was at 4.62cm; the load on the vena contracta was at 10.34cm, and the resultant velocity of the free-stream was at 1.424 m/s.

Moreover, it was necessary to ensure the zero value of piezometric head gradient of the boundary layer, in each one of the seven considered flow conditions. This corresponded to hold the free surface of the current horizontal, at least in the first 50cm where the boundary layer developed. Therefore, it was necessary to suitably incline the channel, taking into account the vegetation type, whose possible increase in height and density would generate a corresponding increase of head losses, which would need to be balanced by a suitable increase of channel slope. The chosen slopes' values are reported in both the subsequent tables 3 and 4. The test sections were set at 20, 30, 40 and 50 centimetres from the channel inlet. In each test section two measuring verticals were considered, differently positioned with respect to the cylinders. The first one was set at the centre of either a rectangular or a square mesh. The second one was set along a cylinder row and at the centre of the lateral

side of the same rectangular or square mesh. It is clear that in the case of a smooth bottom there was no need of such a second measurement location.

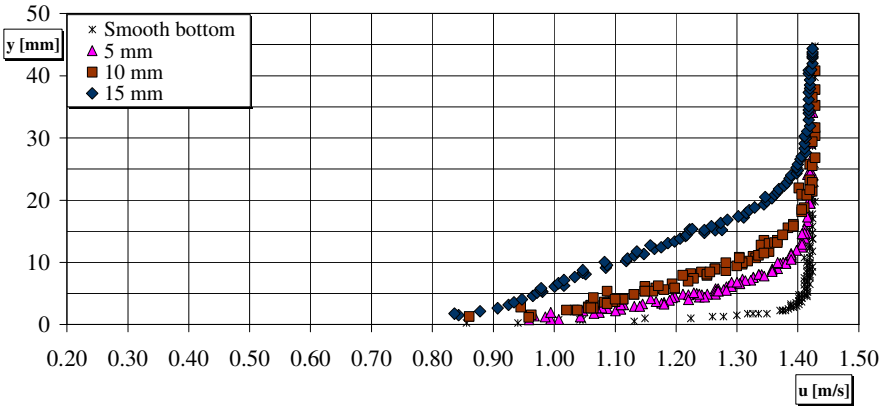


Figure 2. Local mean velocity distributions in test section n.1, central vertical (single density).

Either in the case of smooth or vegetated bottom, in each test section, along the chosen verticals, instantaneous velocities were measured in 20÷30 experimental points through a suitable LDA system, and in each point the local mean velocity was obtained.

Finally, in the figures 2 and 3, examples of the experimentally obtained local mean velocity distributions are reported.

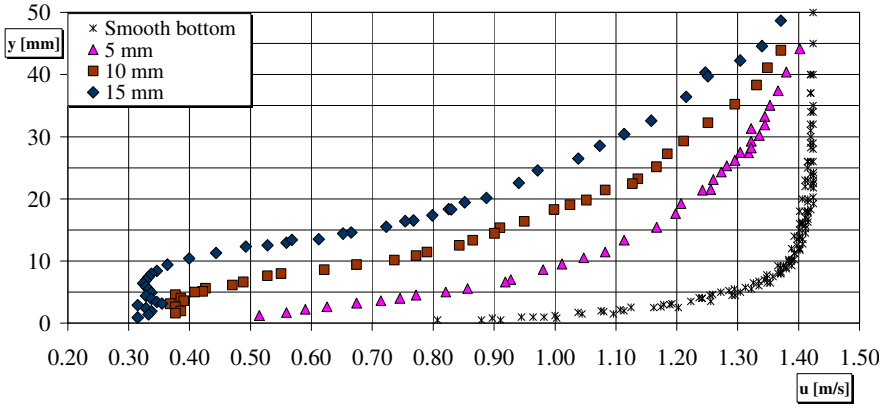


Figure 3. Local mean velocity distributions in test section n.4, lateral vertical (double density).

5. CONTROL OF FLOW RATE EXPERIMENTAL VALUES

Once the experimental local mean velocity distributions had been obtained, some preliminary elaboration of the data were carried out.

First of all, the local mean velocity distributions were completed through two further values of the velocity, respectively relative to the bottom and to the free surface, which could not be obtained directly through the LDA measurements. These values were defined through suitable extrapolation techniques.

The second preliminary elaboration was to control, in each test section and in each flow condition, the flow-rate experimental values per unit width q_u. This control was performed

starting from the integral definition of the flow rate per unit width and transforming it in a summation:

$$q_u = \frac{Q}{b} = \int_h V dh \cong \sum_0^{h\,max} V_{mean} \cdot \Delta h \qquad (5)$$

The summation was carried out with reference to each height step bounded by two following experimental points (included the bottom and the free surface ones) and the term V_{mean} represents the mean value of the superior and inferior velocity values in each summation step. The result of the summation represents directly the unit flow rates in case of smooth bottom, while in the cases of vegetated bottom instead, the unit flow rates were obtained for each vegetated bottom, through a mean between the results of the summation relative to central and lateral measurements.

The results of such a calculation are reported in the table 1. In the same table, the column of theoretical flow-rates represents the values computed directly multiplying the theoretical velocity in the vena contracta (1.424m/s) by the vena contracta cross section suitably reduced of the cross section relative to the cylinders. Moreover, also a column with the mean experimental flow-rates among the four test sections of each considered flow condition are shown.

In the last column of the table 1, the ratios between experimental flow rates and theoretical ones are reported. The maximum difference appears to be 2.5%: it can be attributed to experimental discrepancies. Therefore, the goodness of the fit between experimental and theoretical flow-rate values is sufficiently proved.

Flow-rate values are expressed per unit width, they show that the influence of lateral walls can be neglected similar to the wide channel case. Consequently in following calculations equation (4) with the Marchi's coefficient 0.83 will be considered.

Table 1. Experimental and theoretical unit flow-rates

		Experimental flow rates Q m^3/s				Theoretical flow-rates Q (m^3/s)	Experimental flow-rates means Q (m^3/s)	Ratio
		sec.1	sec.2	sec.3	sec.4			
Smooth bottom		0,00984	0,00988	0,00986	0,00983	0,00985	0,00985	1,000
Single Density	**5mm**	0,00981	0,00975	0,00974	0,00970	0,00968	0,00975	1,007
	10mm	0,00981	0,00974	0,00969	0,00975	0,00951	0,00975	1,025
	15mm	0,00958	0,00941	0,00931	0,00943	0,00934	0,00943	1,010
Double Density	**5mm**	0,00972	0,00966	0,00957	0,00956	0,00968	0,00962	0,994
	10mm	0,00952	0,00941	0,00924	0,00948	0,00951	0,00941	0,990
	15mm	0,00917	0,00895	0,00890	0,00939	0,00934	0,00910	0,975

6. BASIC METHODOLOGY TO OBTAIN THE f FRICTION FACTOR STARTING FROM BOUNDARY LAYER MEASUREMENTS

In the following paragraphs the methodology to obtain the f friction factor starting from boundary layer measurements will be described. In fact, as the available stream is a boundary layer one, the flow is not uniform and the f friction factor values that can be obtained are the cross section ones, and define a cross section f(s) function. In subsequent paragraph it will be shown how to obtain the uniform flow reach value of f friction factor. This one will be subsequently used to obtain the equivalent roughness ε and/or the Manning coefficient n of the considered kind of bottom.

Let us suppose to know the function f(s) relative to a certain kind of vegetated bottom and to the specific stream flowing over. Taking into account the circumstance that also the currents' height h(s) and the flow velocity $V(s) = q_u / h(s)$ are functions of the same distance,

it will be possible to write the Darcy-Weisbach equation (1) in the following way, where $\overline{V}$ is the mean velocity in cross section:

$$J(s) = \frac{f(s)}{4h(s)} \frac{\left[\overline{V}(s)\right]^2}{2g} \tag{6}$$

as in an wide rectangular cross section of height h the hydraulic radius coincides with h. Now it is also possible to write:

$$J(s) = -\frac{dH(s)}{ds} = -\frac{d}{ds}\left\{\alpha(s)\frac{\left[\overline{V}(s)\right]^2}{2g}\right\} \tag{7}$$

where α is the Coriolis coefficient, which in boundary layers is also a function of s. In fact, the hydraulic head H is the sum of the piezometric head and the kinetic head: the first one is constant in all our boundary layers currents as their free surface are always horizontal. Consequently, equations (6) and (7) give:

$$f(s) = -\frac{8gh(s)}{\left[\overline{V}(s)\right]^2}\frac{d}{ds}\left(\alpha(s)\frac{\left[\overline{V}(s)\right]^2}{2g}\right) = -\frac{4h(s)}{\left[\overline{V}(s)\right]^2}\frac{d}{ds}\left\{\alpha(s)\left[\overline{V}(s)\right]^2\right\} \tag{8}$$

Remembering now that $\overline{V} = q_u / h$ is the mean velocity in cross section and that consequently $\overline{V}$ is variable with s, whereas q_u is constant, rearranging of terms yields:

$$\begin{aligned} f &= -\frac{4h^3}{{q_u}^2}\frac{d}{ds}\left(\alpha\frac{{q_u}^2}{h^2}\right) = -4h^3\frac{d}{ds}\left(\alpha\frac{1}{h^2}\right) = -4h^3\left(\frac{1}{h^2}\frac{d\alpha}{ds} - \frac{2\alpha}{h^3}\frac{dh}{ds}\right) = \\ &= 4h^3\left(\frac{2\alpha}{h^3}S - \frac{1}{h^2}\frac{d\alpha}{ds}\right) = 8\alpha S - 4h\frac{d\alpha}{ds} \end{aligned} \tag{9}$$

In previous equation, for sake of simplicity, the functional dependence on s has been discarded. Now, as it will be shown afterwards, a particular cross-section can be identified where the condition dα/ds = 0 holds. Flow in this cross-section most closely resembles uniform flow (α=const). In this section we assume that the simplified resistance equation, f = 8αS, is valid. The corresponding resistance coefficient f is assumed to be a representative reach-value of a uniform field, having the same flow rate and water depth as the aforementioned selected channel cross-section.

At this point the situation is the following one. The f friction factor value relative to the uniform flow can be computed through the simple formulation f = 8αS utilized in the suitable cross section where dα/ds = 0. In order to obtain the unknown abscissa of this cross section and also the unknown value of α in the same cross section, it is necessary to known the analytical expression of the α(s) function. The value of α at the cross-section will be obtained by interpolation of α-values determined from velocity profiles in the test sections.

7. CALCULATION OF α VALUES, α(s) FUNCTIONS, AND f FRICTION FACTOR VALUES

In the case of smooth bottom, as the velocity is the same along any horizontal line of the cross section, the expression of α can be easily transformed in an equation holding integrals of only one parameter (instead of surface integrals) which at last can be computed in the same way that in equation (5), i.e.:

$$\alpha = \frac{\int_0^{h\,max} V^3 dh}{h\overline{V}^3} = \frac{\sum_0^{h\,max} \frac{(V^i c^3 + V^{i+1} c^3)}{2} \cdot \Delta h}{h\overline{V}^3} \tag{10}$$

where V^i_c is the measured velocity at measurement point i.

In contrast, in the case of our model of vegetation, the flow velocity is variable along any horizontal line of the cross section, and therefore α can not be calculated as easily. To solve this difficulty, we make use of the available velocity distributions in the test sections in the following way.

First of all, if we consider any horizontal line in a cross section, it happens that the horizontal velocity distribution has an oscillating behaviour between a minimum value corresponding to lateral measurements and a maximum value corresponding to central measurements. Among the possible functions, the sine function is the most suitable that can represent this oscillating behaviour. Consequently, we can start from this equation:

$$\alpha = \frac{\int_0^{h\,max} \overline{V^3_{mean}} dh}{h\overline{V}^3} \tag{11}$$

where the symbol $\overline{V^3_{mean}}$ is defined as the cubic mean of the velocity values along an horizontal line. In order to compute this term taking into account a sinusoidal velocity distribution between a minimum value (V_l) and a maximum value (V_c), it is necessary to start from the following representation of the horizontal velocity distribution: $V = V_0 + \Delta V \sin x$, where:

$$V_0 = \frac{V_c + V_l}{2} \qquad \Delta V = \frac{V_c - V_l}{2} \tag{12}$$

Consequently, it is possible to write in the generic point of the considered horizontal line:

$$\begin{aligned} &\overline{V^3_{mean}} = \frac{1}{2\pi} \int_0^{2\pi} (V_0 + \Delta V \sin x)^3 dx = \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left(V_0^3 + 3V_0^2 \Delta V \sin x + 3V_0 \Delta V^2 \sin^2 x + \Delta V^3 \sin^3 x\right) dx = V_0^3 + \frac{3}{2} V_0 \Delta V^2 \end{aligned} \tag{13}$$

Substituting now equation (12) in equation (13), it is possible to obtain, through simple calculations:

$$\begin{aligned} &\overline{V^3_{mean}} = V_0^3 + \frac{3}{2} V_0 \Delta V^2 = \left(\frac{V_c + V_l}{2}\right)^3 + \frac{3}{2}\left(\frac{V_c + V_l}{2}\right)\left(\frac{V_c - V_l}{2}\right)^2 = \\ &= \frac{V_c^3}{2} + \frac{V_l^3}{2} - \frac{3}{16}(V_c - V_l)\left(V_c^2 - V_l^2\right) \end{aligned} \tag{14}$$

Now, this expression must be inserted in equation (11) and this one must be suitably integrated obtaining:

$$\alpha = \frac{\int_0^{h\,max} \overline{V^3_{mean}} dh}{h \cdot \overline{V}^3} = \frac{\int_0^{h\,max} \frac{V_c^3}{2} dh + \int_0^{h\,max} \frac{V_l^3}{2} dh - \int_0^{hmax} \frac{3}{16}(V_c - V_l)\left(V_c^2 - V_l^2\right) dh}{h \cdot \overline{V}^3} \tag{15}$$

Finally, we transform integrals in summations obtaining:

$$\alpha = \frac{\sum_{0}^{h_{max}} \frac{(V^{i}{}_{c}{}^{3} + V^{i+1}{}_{c}{}^{3})}{2} \frac{\Delta h_i}{2} + \sum_{0}^{h_{max}} \frac{(V^{i}{}_{l}{}^{3} + V^{i+1}{}_{l}{}^{3})}{2} \frac{\Delta h_i}{2} - \sum_{0}^{h_{max}} \eta_j \Delta h_j}{h \cdot \overline{V}^{3}} \tag{16}$$

with:

$$\eta_j = \frac{3}{16}\left(V_{cj} - V_{lj}\right)\left(V_{cj}{}^{2} - V_{lj}{}^{2}\right) \tag{17}$$

where j represent the generic height between two following measurement points. Finally, equations (16) and (17) allow calculation of α-values in all the test sections of all the seven considered flow conditions. These values are reported in table 2.

The values from the table 2 are also presented in Figure 4. In this figure it is shown that the experimentally obtained α values can be interpolated through second order polynomials which start all from the same s_0 section placed at about 0.10m from the leading edge of the boundary layer. Up to this section, the α-value may be considered equal to unity.

Table 3 shows the results of the aforementioned analysis: the second order polynomial curves coefficients a, b, c; the slope of the current S; the abscissa s_0 of the test section where dα/ds = 0; the height of the test section where dα/ds = 0.

Table 2. Coriolis α values for every flow conditions and in each test section

		Final α values			
		sec.1	**sec.2**	**sec.3**	**sec.4**
Smooth bottom		1,0081	1,0069	1,0127	1,0161
Single Density	**5mm**	1,0289	1,0416	1,0507	1,0590
	10mm	1,0470	1,0790	1,0998	1,1132
	15mm	1,0884	1,1411	1,1796	1,1833
Double Density	**5mm**	1,0343	1,0369	1,0458	1,0402
	10mm	1,0743	1,1027	1,1298	1,1071
	15mm	1,1371	1,1890	1,2178	1,1749

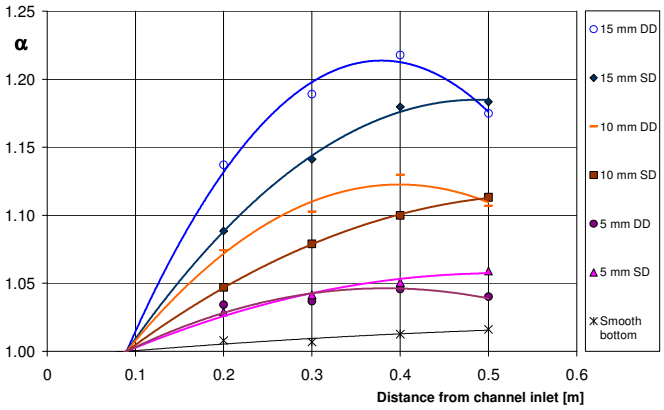


Figure 4. Values of parameter α and interpolating functions

Once the maximum values of α for each flow condition have been obtained, it is finally possible to use equation (9) to obtain the f values relative to the seven different considered bottoms. The results of such calculations are reported also in table 3.

Table 3. Interpolating curves coefficients for α values, consequent maximum values of same α and f correspondent values for each considered bottom.

		$\alpha=as^2+bs+c$			S	s_0 [m]	h_0 [m]	α [$s=s_0$]	$f=8S\alpha$ [$s=s_0$]
		a	b	c					
Smooth bottom		-0,03630	0,05930	0,99500	0,0025	0,817	0,04814	1,01922	0,02038
Single Density	**5mm**	-0,31910	0,32820	0,97310	0,0092	0,514	0,05083	1,05749	0,07783
	10mm	-0,49900	0,55930	0,95280	0,0160	0,560	0,05507	1,10952	0,14202
	15mm	-1,17120	1,14140	0,90680	0,0227	0,487	0,05716	1,18489	0,21518
Double Density	**5mm**	-0,54110	0,41410	0,96710	0,0115	0,383	0,05050	1,04633	0,09626
	10mm	-1,28530	1,02520	0,91820	0,0205	0,399	0,05428	1,12263	0,18411
	15mm	-2,56320	1,94120	0,84610	0,0295	0,379	0,05727	1,21363	0,28642

8. FINAL CALCULATIONS OF n AND ε VALUES

Finally, starting from the obtained f values, and employing the aforementioned methodologies, it is possible to obtain both the n Manning coefficients and the equivalent sand roughness ε corresponding to the seven considered bottoms. In particular the procedures are the following ones.

For the n coefficients, it is simple to use equation (2). For the ε values it is necessary to invert equation (4), with the shape parameter $\psi = 0.83$. The results of such calculations are reported in table 4. For the smooth bottom case, the value of $\varepsilon = 0.14$mm, and Manning's n = 0.0097 are consistent with values for plexiglass roughness cited in literature (Ven te Chow, 1988). As expected, both the values of ε and n in case of vegetated bottom follow a clear trend of growth either with increase of density or with increase of cylinders height.

Table 4. Bed's slopes and final f values and for each flow condition, and consequent equivalent sand roughness and n Manning parameter.

		$f=8\alpha i$ [$s=s_0$]	S	h_0 [m]	$\varepsilon/4h_0$ [$s=s_0$]	ε [$s=s_0$]	n
Smooth bottom		0,02038	0,0025	0,04814	0,000720	0,000139	0,00972
Single Density	**5mm**	0,07783	0,0092	0,05083	0,049685	0,010102	0,01917
	10mm	0,14202	0,0160	0,05507	0,145103	0,031961	0,02624
	15mm	0,21518	0,0227	0,05716	0,257374	0,058847	0,03250
Double Density	**5mm**	0,09626	0,0115	0,05050	0,075319	0,015215	0,02129
	10mm	0,18411	0,0205	0,05428	0,210463	0,045692	0,02980
	15mm	0,28642	0,0295	0,05727	0,358251	0,082069	0,03751

Finally, in figure 5, the obtained roughness of each bottom is reported as a function of the cylinders height and density, including also the case of smooth bottom. The experimental points are very well aligned along straight lines which join themselves in a single point on the abscissa axis. Probably the ε value is almost zero if the cylinders height is no more than 2 or 3mm, due to the presence of the laminar substratum. The two straight lines show an increasing trend with the cylinders height, more accentuated in case of double density with respect to single density. Besides the 3mm cylinders' height ε values can be computed, in relation to the two vegetation densities, through the formulas reported into the diagram.

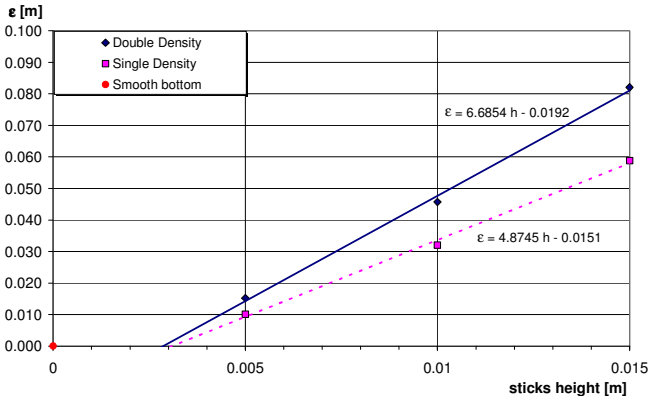


Figure 5. Roughness values of the different bottom and their interpolating functions.

9. COMPARISON TO DATA FROM LITERATURE

As it has been already stressed, the very little value of roughness obtained in the case of smooth bottom represents a first confirmation of the validity of the considered procedure. But, in any case, another literature control has been performed with the Lopez and Garcia experimental data (1997). In particular, to perform such a comparison, a graph of Lopez and Garcia has been considered. In this graph, in the abscissa, the projected area of vegetation per unit volume of water in the flow direction and in the ordinate the Manning's n values, are reported. In the cases of this papers' measurements, the projected area of vegetation per unit volume of water in the flow direction parameter is worth $3.2m^{-1}$ in the case of single density and $6.4m^{-1}$ in the case of double density. Consequently, six experimental points have been obtained, relative to both two densities' values and three cylinder' heights. For each one of these six different conditions, the values of the Manning's n reported in table 4 have been represented through a point into the Lopez and Garcia graph, so obtaining the final sketch of figure 6.

The obtained results are qualitatively comfortable. The Manning' s n value depends mainly on the bottom roughness, so that, with equal cylinders density, the n coefficient must grow with the height of the cylinders. Figure 6 shows that for both cylinders densities, higher cylinders yield higher Manning's n values, and that the obtained roughness values from the current study remain lower than the Lopez and Garcia interpolating line which corresponds to 10cm cylinder height. Referring to the flow depth, the use of equations (2) and (4) can show that for a given ε value, the Manning's n value must slowly decrease with it. Now, the flow in (Lopez and Garcia, 1997) is deeper than in authors experiments; consequently, the effect of the flow depth would be to shift the Lopez and Garcia line from the authors points.

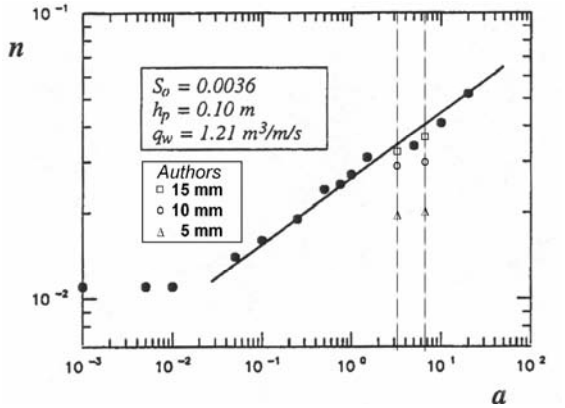


Figure 6. Comparison between Lopez and Garcia data and authors data

10. CONCLUSIONS

The aim of this paper was to present a simplified procedure of determining the resistance coefficients of water currents, and in particular of vegetated bottom current. The proposed procedure allows the estimation of the f friction factor of a uniform flow through experimental measurements carried out on a boundary layer stream. Consequently, instead of requiring a very long experimental channel, the whole procedure can be carried within the boundary layer of the flow in a shorter experimental channel and the results can be extrapolated to the uniform flow.

The roughness values obtained with the described procedure are reasonable either in case of smooth or vegetated bottom.

In a historic period in which vegetated bottoms will attain more and more importance in engineering practice, the possibility of obtaining such results through the use of simplified experimental facilities can play a positive role in the field of environmental hydraulics.

REFERENCES

Bettess, R., On the hydraulic roughness of vegetation in channels XXX *I.A.H.R., Congress Tessaloniki (Greece)* Theme C 25-30, 2003

De Felice, S. and P. Gualtieri, Interactions between turbulent Boundary Layer and rigid vegetated beds: Comparison among hydrodynamic characteristics in different points around the cylinders, *I.A.H.R. Congress, Seoul (Korea),* pp.2434-2445, 2005

Gualtieri, P. and G. Pulci Doria, Boundary layer development over a bottom covered with rigid submerged vegetation, *Fluid Mechanics of Environmental Interfaces*, Taylor&Francis, Leiden, The Netherlands pag 241-298, 2008

Kouwen, N., T.E. Unny and H.M. Hill, Flow retardance in vegetated channels *Journal Hydraulic Division ASCE* 95(2), 329-342, 1969

Kouwen, N. and T.E. Unny, Flexible roughness in open channels *Journal Hydraulic Division ASCE* 99(5), 684-698, 1973

Kouwen, N. and T.E. Unny, Flow resistance in vegetated waterways *Trans. ASEE* 24(3) 684-698, 1981

Lopez, F. and M. Garcia, Open-Channel Flow Through Simulated Vegetation: Turbulence Modeling and Sediment Transport, *US Army of Engineers Waterway Experiment Station Wetlands Research Program Technical Report WRP-CP-10*, 1997

Marchi, E., Il moto uniforme delle correnti liquide nei condotti chiusi ed aperti, *L'Energia Elettrica,* **23** (4)(5), 1961

Nezu, I. And K. Onitsura, Turbulent structures in partly vegetated open-channels flows with LDA and PIV measurements *Journal of Hydraulic Research* 2001 no 6 vol.39 629-642, 2002

Stone, B. M. and H. T. Shen, Hydraulic Resistance of Flow in Channels with Cylindrical Roughness *Journal of Hydraulic Engineering* 2002 (may) 500-506, 2002

Tsuijimoto, T., Fluvial processes in stream with vegetation *Journal of Hydraulic Research 37(6) 789-803*, 1999

Tsuijimoto, T. and T. Kitamura, Velocity profile of flow in vegetated-bed channels *KHL Progressive Report n.1 Kanazawa University Japan*, 1990

Tsuijimoto, T., Y. Shimizu, T. Kitamura and T. Okada, Turbulent open-channel flow over bed covered by rigid vegetation, *Journal of Hydroscience and Hydraulic Engineering*, **10**(2), pp.13-25, 1992

Chow, V.T., *Open channel hydraulics*, Mc Graw-Hill, U.S.A., 1988

Yen, B.C., Open channel flow resistance, *Journal of Hydraulic Engineering*, **128**(1), pp. 20-39, 2002

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

3D ecological modelling of the Aveiro coast (Portugal)

A. C. Cardoso [a,b] , J. F. Lopes [a,b],

[a] Departamento de Física, Campus Universitário Santiago
Universidade de Aveiro, 3810-193 Aveiro, Portugal (acardoso@fis.ua.pt)
[b] CESAM, Departamento de Física, Campus Universitário Santiago
Universidade de Aveiro, 3810-193 Aveiro, Portugal (jflopes@ua.pt)

Abstract: Until now the most of studies concerning the dynamics of the Portuguese coastal ecosystem focused in estuarine, lagoon and rivers systems, whereas more knowledge about the spatial and the seasonal distribution of phytoplankton population at the coast is still needed. The present work is aimed to study some aspects of ecological modelling of the nearshore Aveiro coastal system, namely the influence of physical forcings in the distribution of phytoplankton biomass. In order to implement the ecological model, the validation of the model were assessed by comparing simulation to data to a vertical and horizontal distribution of temperature, nutrients and phytoplankton biomass, obtained from the several surveys performed along the Portuguese coast during the last decades (Moita, 2001). The results evidence, that the physical processes play an important role in the understanding the phytoplankton dynamics and the close relationship between the physical forcings and the temperature and the phytoplankton distributions.

Keywords: Aveiro coast; Modelling; Ecological; Upwelling; Nutrients: Chlorophyll-a.

1. INTRODUCTION

The Iberian Atlantic coast is characterized by the occurrence of important phytoplankton productivity, both in space and time. In fact, upwelling is the most important hydrodynamic process that occurs there, typically between April and October. The coastal upwelling change the general phytoplankton cycle due to additional energy input, associated to the nutrients upwelled into the euphotic zone.
The knowledge of rich variety of the dynamics along the Aveiro coastal waters is the primary key for the understanding of the evolution of this ecosystem. This region is characterized by intense mesoscale activity, with generation and evolution of complex meanders, eddies and filaments along the coast. The processes responsible for the underlying dynamics of many of these features are not yet well understood.

The ocean off the Iberian Peninsular is one of high seasonal variability and is a demanding area to model realistically (Stevens et al., 2000). Therefore the validation of an ecological model for the Aveiro coastal ecosystem will contribute to better understand its dynamic.
The model was setup for the study area in order to modelling the temperature, nutrients and chlorophyll-*a* distributions, in situations of upwelling favourable winds.

Scenarios representing, typical summer/winter conditions corresponding generally to upwelling/downwelling situations along the western Iberian coast and the influence of the Ria de Aveiro lagoon are presented. The results show that the model results, when compared to the data, reproduced correctly the vertical profiles of temperature (T), nutrients (NO_3) and chlorophyll-a (CHL-a), and that the physical forcings determine the ecosystem state of the water column. The model is now a valuable tool for assessing the ecosystem variability of the Aveiro coast.

2. NUMERICAL MODEL

2. 1 Introduction

The model used in this study is a three-dimensional coupled hydrodynamic and ecological model for coastal and shelf seas (Luyten, 1999, Proctor, 1997). It is constituted by four sub-models: the hydrodynamical/physical model, the biological model, the sediment model and the contaminant model. The physical model includes the main processes, namely, the advective and the diffusive transport of momentum, as well as the interaction with the atmosphere, through the momentum (wind stress at the surface) and the radiative exchanges. The biological sub-model is based on Tett conceptual model (Tett, 1998; Tett P. and Walne A., 1995) and has the following structure: a) internal non-conservative biological or chemical processes; b) photosynthesis by the absorption of PAR; c) physical transport by advection and diffusion; d) vertical sinking; e) deposition and erosion via a "fluff" layer; f) exchanges between the water column and the sediment layer. Its distinguishing features are: (1) the use of a microplankton compartment to include the biomasses and the microbial loop organisms activity, as well as those for the phytoplankton, and (2) the use of variations in the chemical composition (especially, the nitrogen: carbon ratio) and of the microplankton and the detrital components to control many of the biological processes.

2.2 The ecological model description

The ecological model is forced by a physical model which is represented by a set of equations including 3D hydrodynamic and transport models. The hydrodynamic model is very similar to the *Blumberg and Mellor* primitive equation models (1987), which use the sigma coordinates and an embedded turbulent closure scheme (Luyten, 1999, Proctor, 1997).

The biological sub-model is based on Tett conceptual model (Tett, 1998; Tett and Grenz, 1994; Tett and Walne, 1995) and has the following structure: a) internal non-conservative biological or chemical processes; b) photosynthesis by the absorption of PAR; c) physical transport by advection and diffusion; d) vertical sinking; e) deposition and erosion via a "fluff" layer; f) exchanges between the water column and the sediment layer.

The model is forced by seasonal cycles of vertical transport and mixing, light and zooplankton grazing pressure, and boundary conditions. The autotrophs (i.e. phytoplankton) growth rate is calculated from the cell quota threshold limitation (CQTL) theory (Droop et al., 1982), as the least growth rate predicted from light or nutrient controlled growth:

$$\mu_a = \min\left[\mu_{a1}(I_p), \mu_{a2}(Q)\right] \qquad (1)$$

where $Q=N/B$ is the nitrogen quota defined as the ration of the organic nitrogen N (mmol N) and the nitrogen microplankton carbon(mmol C) and I_p is the photosynthetically active radiation (PAR).

The phytoplankton needs light for growth, and under some conditions (e.g. during winter) light, rather than nitrogen supply, limits the microplankton growth. The light intensity controlled growth are calculated as the sum of the photosynthetic production and respiration loss, by a formalism described by Droop et al. (1982) and Tett (1989), which takes into account the photosynthetical efficiency and the ratio of chlorophyll to autotrophic carbon:

$$\mu_1(I_p) = (\alpha \chi_a I_p - r_a)(1-\eta) - r_h \eta \qquad (2)$$

where $\alpha=k\,\varepsilon^X\,\Phi$ ((mmol C)(mg Chl)$^{-1}$(day^{-1} W^{-1} m^2)) is the photosynthetic efficiency, ε^X (m^2 mg^{-1} Chl^{-1}) the phytoplankton attenuation cross-section, Φ (nmol C μE^{-1}) the photosynthetic quantum yield, k a conversion factor, χ_a the ratio of chlorophyll to autotroph carbon and r_a the autotrophic respirations and r_h the heterotrophic respirations losses, respectively:

$$r_a = r_{0a} + b_a \mu_{a_1}(I_p) \qquad r_h = r_{0h} + b_h\, \mu_1(I_p) \qquad (3)$$

with r_{0a} and $.r_{0h}$ corresponding to the basal respiration rate for autorothrops and heterothrops, respectively, and b_a and b_h the autotrophic and heterotrophic respiration-growth slopes, respectively. The heterotrophic fraction η is defined as:

$$\eta = \frac{B_h}{B_a + B_h} \tag{4}$$

with B_a and B_h, respectively, the autotrophs and the heterotrophs carbon biomass.

The nutrients controlled growth is calculated as a function of the nutrients cell quota and the respiration loss. The nutrient uptake is calculated by a Michaelis-Menten type kinetics function. Under nitrogen-limiting conditions, growth rate should, ideally, be calculated as a fraction of the maximum specific growth rate depending on the cell nutrient quota, less respiration:

$$\mu_2(Q) = \mu_{max_a} f(T)(1 - \frac{Q_{min_a}}{Q_a})(1-\eta) - r_h\eta \tag{5}$$

where Q_a is the (variable) autotrophic cell quota and f(T) the temperature growth.

The biological model cycles the concentrations of organic carbon and nitrogen through the microplankton and the detrital compartments with associated changes in dissolved concentrations of nitrate, ammonium and oxygen. The concentrations are updated in time by solving the transport equation for each state variable, whereby the biological interactions are included as source or sink terms and which takes account of vertical sinking and the physical transport by advection and diffusion.

3. STUDY AREA

The study area is located in the Atlantic upwelling system of the western coast of the Iberian Peninsula, characterised by meteorological conditions of strong north/northwest prevailing, during almost the year which favours the upwelling of nutrient enriched cold deep waters, the phytoplankton development and the primary production. The continental shelf of the Aveiro coast is relatively wide (~60km) and gently sloping with an edge defined by the 200-m isobath, where the Aveiro Canyon (40º42'N) is the most significant topographic feature due to the fact that the slope gets very steep in just a few kilometres (Peliz et. al., 2002).

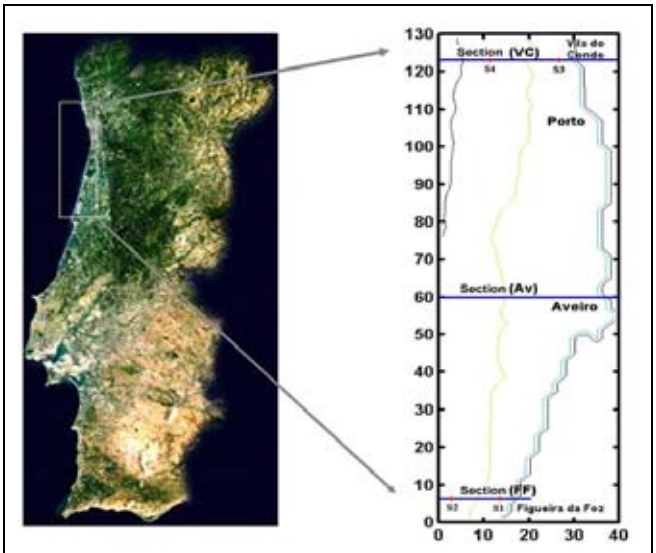


Figure 1: Study area. Costal area between Vila do Conde and Figueira da Foz, centred in Aveiro. Vertical sections W-E and offshore (S1 and S3) and inshore (S2 and S4) stations.

The bathymetry of the Portuguese coast for the study area has a 2 km resolution in both the latitude and the longitude directions (Cardoso, 2003). The computational domain, with the origin situated at 40º38'N and 8º27'W, corresponds to an area centred at the Aveiro station, extending 40 km offshore and 130 km alongshore (Fig. 1). It is represented by a uniform horizontal grid composed by 40x20 cells and a uniform 22 sigma levels distributed in vertical direction. The wind intensity and direction were taken from the WRF model (Weather Research and Forecast Model) a nesting mesoscale and assimilation forecast model.

4. MODEL VALIDATION

The model implementation included the calibration (Cardoso, 2005) and the validation. The last one may be defined as a procedure consisting in verifying if the model is able to reproduce data, independently of those involved in its calibration. Fig 2 shows vertical profiles for temperature, nitrates (NO_3) and chlorophyll-a (Chl-a) for two stations S_1 (near coast) and S_2 (offshore), both located at Figueira da Foz. It can be observed a good agreement between simulated temperature and data for S_1, whereas for S2 there is an overestimation by the model namely between 10 and 70m deep, corresponding to the intermediated layers (fig 2b). The simulated temperature profile presents a shallower thermocline, which in accordance to data should be between 30 and 70 m deep (fig 2b) and more smooth. Concerning the NO_3, (fig 2c), it can be observed that data and simulations are of the same order but the simulation shows a sharper profile than data. The model simulates well the near surface concentration (10 m) for the S_1 station but overestimates the concentration at deeper layers. For the offshore station (S_2) there is a good agreement for the near surface and bottom concentrations, but less for the intermediate layers. The Chl-*a*-simulated values for the S_1 station, (fig 2 (e)), are in the same range of data. The simulated profile presents a good agreement with data, although it overestimates it at 10 m deep (0.06 mg/m^3 instead of 0.09 mg/m^3 for data).

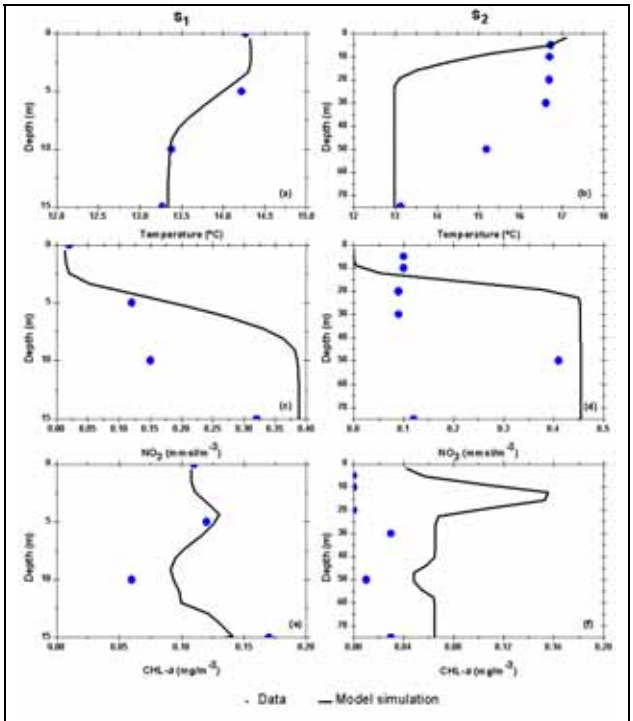


Figure 2: Vertical profiles of: (a)-(b) temperature; (c)-(d); NO_3 (mmol N. m^{-3}); (e)-(f) and Chl-*a* (mg Chl-*a*.m^{-3}), for two stations, S1 e S2, near Figueira da Foz, for the validation.

The results lead to conclude that, in general, the simulations values are of the same order of data. The model reproduces better data near the coast. Offshore station the simulation presents some deviations relatively to data, which may be attributed to the influence of boundary conditions.

5. MODEL APPLICATION IN THE STUDY OF SCENARIOS

The model was applied to study the following scenarios for the Aveiro coast: summer of 2005; winter of 2005; the influence of the Aveiro lagoon (Ria de Aveiro) input.

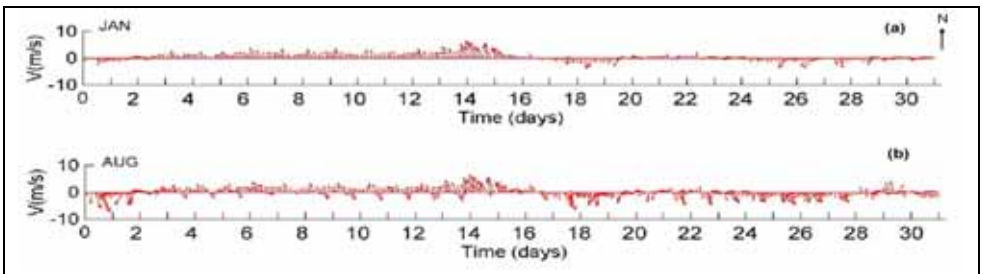


Figure 3: Wind Intensity and direction for: (a) January (winter) and (b) August (Summer) of 2005 year, given by WRF model (Weather Research and Forecast Model).

The wind intensity and direction (fig 3), air temperature, pressure and humidity were taken from the WRF model (Weather Research and Forecast Model), a nesting mesoscale and assimilation forecast model (Skamarock et al., 2007; Michalakes et al., 2005), which allows a mesh refinement, enabling a high resolution output focused over the Aveiro region.

5.1. The 2005 Summer Scenario

This scenario has been performed considering a situation for which winds are predominantly northerly along the Aveiro coast (fig 3(b)), which are typical situation of the upwelling at the western Iberian coast. It was chosen due to the fact that the wind intensity for this summer was relatively higher for this season, when compared with typical situation.

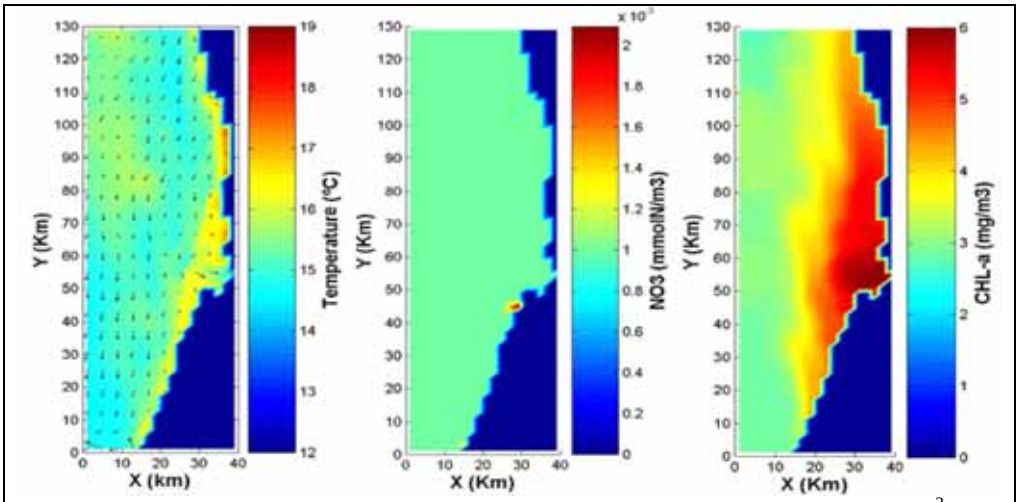


Figure 4: Simulated maps of surface: temperature, NO_3 concentration (mmol N. m^{-3}) and Chl-*a* concentration (mg Chl-*a*.m^{-3}), for 2005 summer. The arrows presented at the temperature map represent the current vectors.

The surface temperature distribution is almost homogeneous (14.5-15ºC). This uniform surface temperature distribution differs from a typical summer situation observed in August 1985 by Moita (2001), which may be attributed by the stronger upwelling, which brings more cold water to the surface, induced by the strong northeasterly wind. Indeed, the temperature vertical sections (Fig. 5(a-c)) shows a cold thin superficial layer (5-10m), with temperatures of about 14.5ºC-16ºC, which was established in the east-wet direction, evidencing the influence of the strong upwelling. It can be observed, as well, the presence of a layer cold water and the changes of declivity of the isotherms close to the coast, which are signatures of the upwelling. Very close to coastal line it can be observed some spot of relatively high surface temperature (16ºC), due the shallowness of the area (fig 4a).

The NO_3 responses of the model are presented in the horizontal maps and vertical sections (Figs. 4b and 5(d-f), respectively). Fig. 4 shows that NO_3 surface concentrations are practically null ($1.0x10^{-3}$ mmolN/m^3).These values are quite similar to those observed by Moita (2001) in August 1985, for both the offshore and the inshore surface waters. The NO_3 vertical section, fig 5 (d-f), shows a depleted layer, within the first 20 m depth, and the NO_3 concentration rapidly increase toward the bottom. Near the coast the NO_3 concentration is practically null through the entire water column, which reflects a depletion by phytoplankton, as, illustrated in Chl-*a* vertical distribution of fig 5(g-i) .

Fig 4c (surface concentration) and Fig. 5g (vertical section) show that Chl-*a* surface concentration is extremely high for the season, reaching a maximum value of 6.0 mg Chl-*a*.m^{-3}, near the coast (Av), and 2.5-3.0 mg Chl-*a*.m^{-3} offshore. Inshore, the Av and the VC sections show that the maximum Chl-*a* concentration is located near the bottom. A plume of a thin layer of maximum Chl-*a* concentration emerges from the coast, stretches and deepens offshore, between 20m and 40 m depth. At the section FF the plume of maximum Chl-*a* concentration is shallower and sinks offshore between 15 and 20 m depth. The maximum Chl-*a* concentration is located within the euphotic zone, around 10 m bellow the

surface, deepening offshore to values close to 40 m, at VC and Av stations, and 20m at FF station. Furthermore, the zone of maximum Chl-*a* concentration nearly follows the bottom profile. A rough estimation shows that the euphotic zone is, therefore, of the order of 5-10 m, near the coast, and 20-40 m offshore.

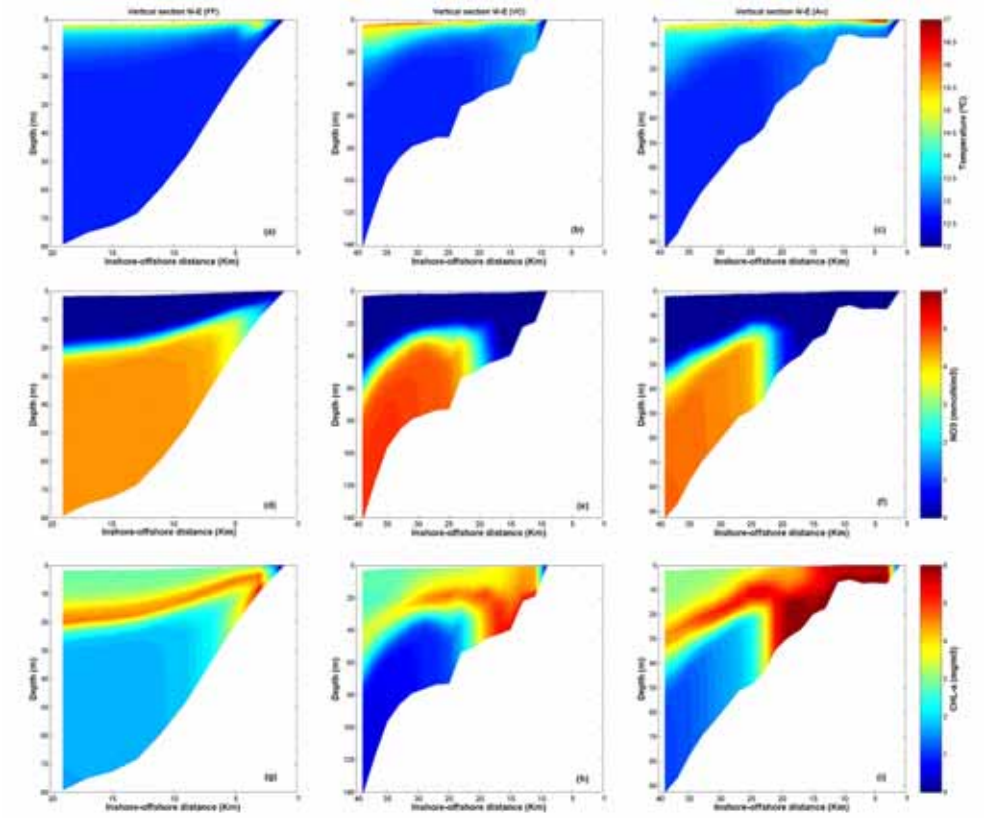


Figure 5: Vertical sections of (a-c) temperature (ºC); (d-f) NO_3 concentration (mmol N. m^{-3}); and (g-i) Chl-*a* concentration (mg Chl-*a*.m^{-3}), at Figueira da Foz (FF), Aveiro (Av) and Vila do Conde (VC) respectively, to the 2005 summer.

The results show the reinforcement of the summer upwelling induced by the strong northerly winds conditions, which mobilise more nutrients to the shallow layers. This condition associated to condition of high temperature and no light limitation, favour the phytoplankton development and high Chl -*a* concentrations. The near depletion of the NO_3 at the surface denotes the intense phytoplankton activity.

5.2. The 2005 Winter Scenario

This scenario corresponds to a situation for which winds are predominantly southerly along the Aveiro coast (fig 3(a)), which is a typical situation of downwelling.
The analysis of fig 6 shows that the surface temperature varies between 14 ºC (offshore) and (12ºC inshore). At southern part of the domain the temperature distribution is more uniform, about 14ºC. It can be observed for the surface, near the coast, the establishment of a warmer water layer, which extending offshore along the bottom. This warm water results from the convergence associated to the Ekman transport due to southerly winds. The vertical distribution of temperature presented in fig 7(a-c) shows as well that temperature is nearly homogeneous within the water column and it is, as well, of the order of 14ºC. It can be observed that the termocline is almost inexistent.
The surface NO_3 concentration is homogeneous with values of 3.5-4 mmolN/m^3 (fig 6), however the concentrations of the northern section (VC) are smaller than that of the southern section (FF), where a maximum concentration of 4 mmolN/m^3, is observed at 40m depth.

At the Av section the NO_3 concentration is smaller (3.5 mmolN/m^3) than that for the remaining stations. Contrarily to the previous scenario (Summer situation) where the surface concentration is almost null, the winter scenario show high NO_3 surface concentrations , since the nutrients consumption by phytoplankton is lower due to its low concentration (fig 6). This result is in agreement with Moita observations (2001). In general, near the coast the nitrate concentrations are lower than offshore, which reflects a downwelling situation.

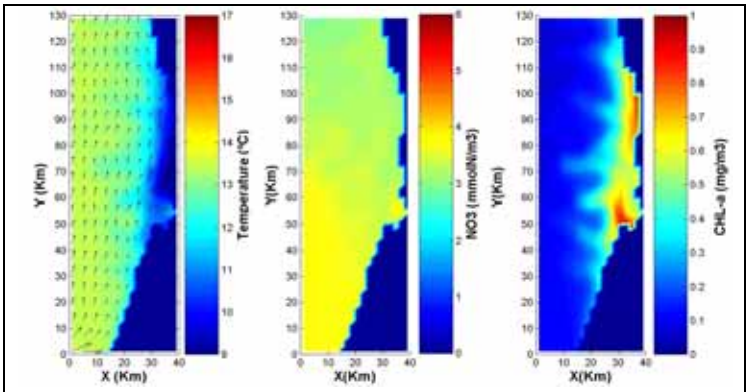


Figure 6: Simulated maps of surface temperature, NO_3 concentration (mmol N. m^{-3}); and Chl-*a* concentration (mg Chl-*a*.m^{-3}), for 2005 winter. The arrows presented at the temperature map represent the current vectors.

The surface Chl-*a* concentrations stretches in a band along the coast mainly between Av and VC sections, where the maximum concentrations of about 1mg/m^3 is observed (very low values when compared to the summer situation) (fig 6). The maximum Chl-*a* concentration deepens following a typical downwelling path as the phytoplankton must be supplied in nutrients (fig 7(g-i). Offshore the Chl-*a* concentrations are low and, therefore, the nutrients availability increases (fig 7 (e-f). In general, the maximum Chl-*a* concentration appears inshore, near the surface (fig 7 (h-i)).

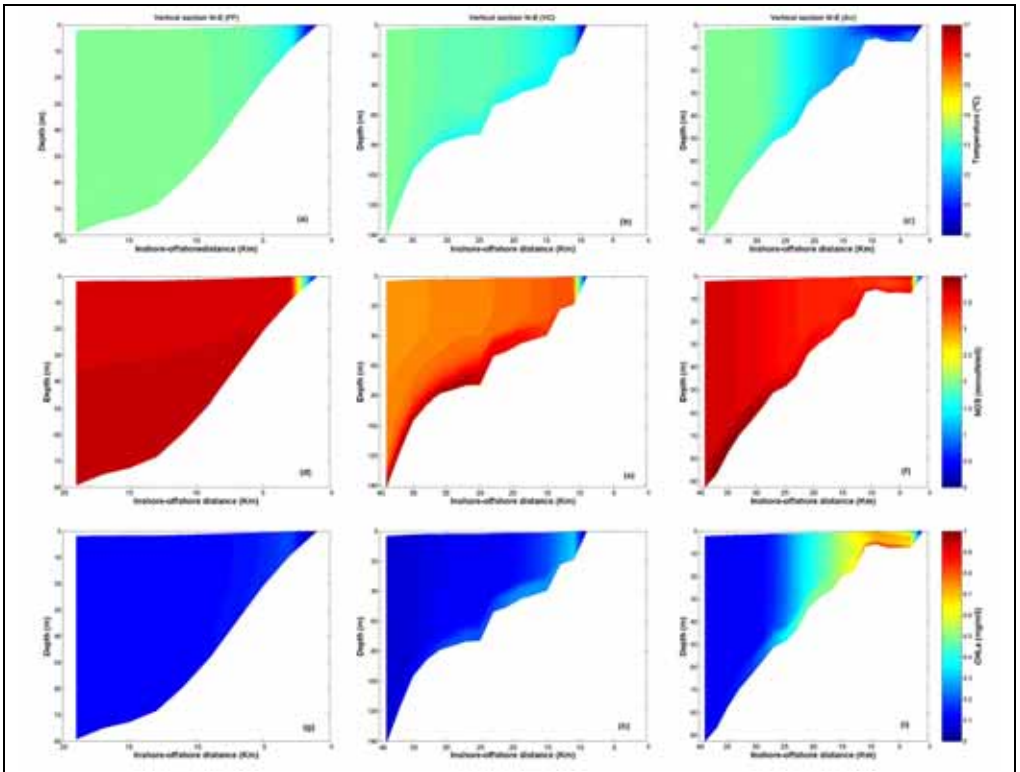


Figure 7: Vertical sections of (a-c) temperature (ºC); (d-f) NO_3 concentration (mmol N. m^{-3}); and (g-i) Chl-*a* concentration (mg Chl-*a*.m^{-3}); at Figueira da Foz (FF), Aveiro (Av) and Vila do Conde (VC) respectively, to the 2005 winter.

This result is, as well, in agreement with Moita observations (2001), and lead to conclude that in southerly wind conditions a downwelling situation is setup forcing, the nutrients to be concentrated near the bottom and offshore. In overall the surface phytoplankton

concentration is low, as the phytoplankton tends to sinks in order to reduce the limitation by light and nutrients.

5.3. The influence of the lagoon outflow

To assess the influence of the Ria de Aveiro lagoon (situated at the center of the eastern boundary), a scenario identical to the 2005 summer scenario was considered, where it has been imposed a constant outflow from the lagoon. All the other conditions were maintained equal.

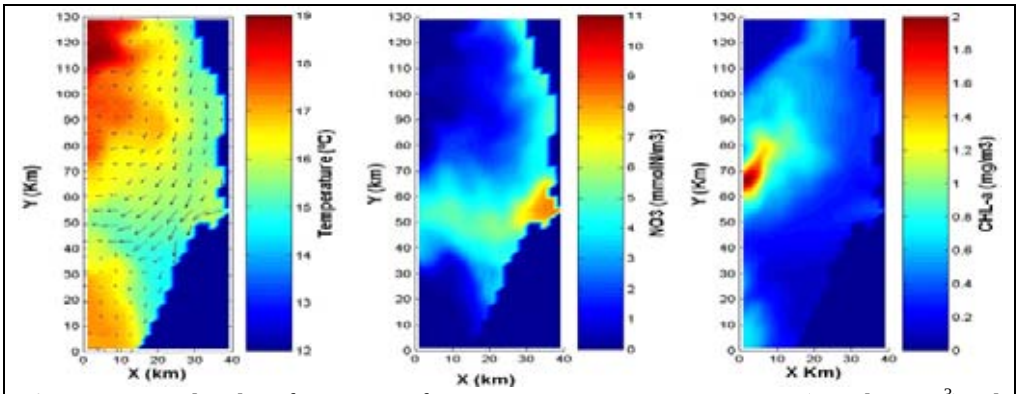


Figure 8: Simulated surface maps of: temperature, NO_3 concentration (mmol N. m^{-3})and chlorophyll-a and Chl-*a* concentration (mg Chl-*a*.m^{-3}). The arrows presented at the temperature map represent currents vector.

Fig 8 shows the influence of the lagoon input. Comparing to fig 4(a) it can be observed that the lagoon induces an intensification of currents near the eastern boundary of the study area.

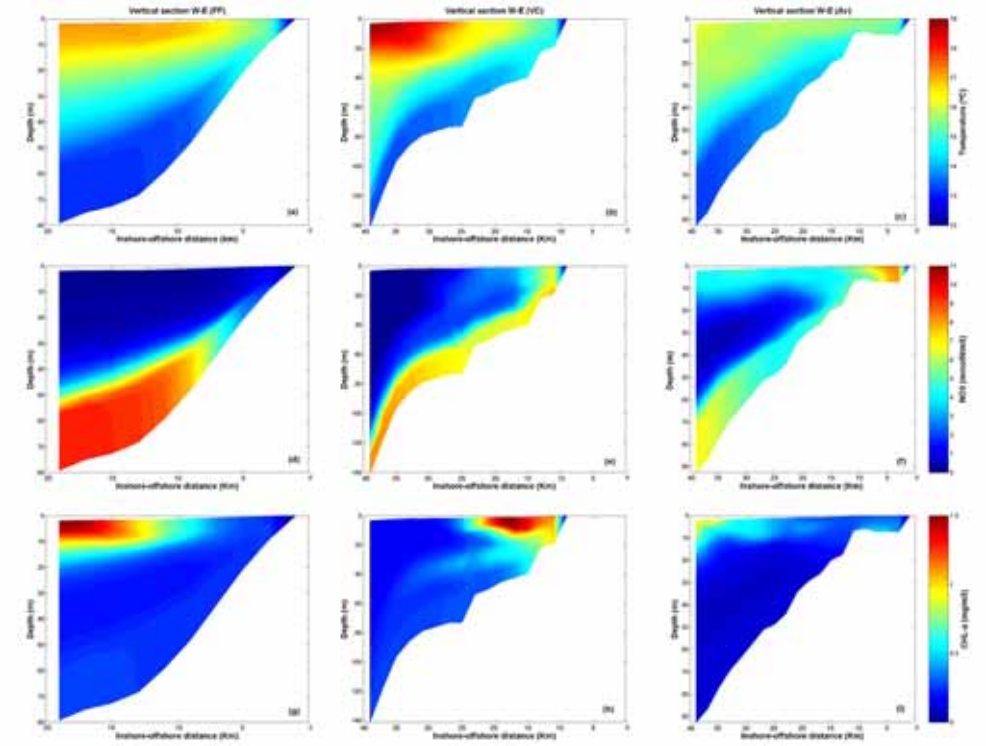


Figure 9: Vertical sections of (a-c) temperature (ºC); (d-f) NO_3 concentration (mmol N. m^{-3}); and (g-i) Chl-*a* concentration (mg Chl-a.m^{-3}); (j-l), at Figueira da Foz (FF), Aveiro (Av) and Vila do Conde (VC) respectively (Ria de Aveiro input).

The surface temperature at the northern area are higher (17-19ºC) than that of the central and the southern areas, which are about to 17-19ºC. The vertical sections shows that a cool water band (15-16ºC) has developed at the eastern boundary, while offshore a warm water layer of about 20m thickness has formed at FF (fig 9 (a)), reaching a value close to 40 m at Av and VC sections (fig 9(b-c). This warm water, located offshore, is due to the intensification of the divergence associated to the Ekman transport.
At the eastern boundary, close to the lagoon mouth, the NO_3 surface concentration is higher relatively to the summer scenario, with concentrations of about 8 mmolN/m^3, and stretches offshore suggesting an enhancing of the nutrients upwelling due to the currents intensification (fig 8 e 9(f)). On the other hand, at Av and VC sections, the maxima concentrations are located near bottom (fig 9 (d-e).
The surface Chl-*a* maximum concentration is occurs offshore, indicating that Chl-*a* distribution reflects this time the influence of the lagoon outflow (Fig 8 and fig 9 (g and i).
At Vila do Conde the surface Chl-*a* maximum concentration is located close to the coast. A secondary Chl-*a* concentration appears stretching offshore at about 40m depth (fig 9 (h)).
This scenario leads to conclude that the lagoon outflow influences the coastal circulation, as well as the intensity of the upwelling. As a consequence, the transport of nutrients and phytoplankton biomass distributions are affected.

6. CONCLUSIONS

The model reproduced correctly the vertical profiles of temperature (T), nutrients (NO_3) and chlorophyll-a (Chl-*a*) at the Aveiro coastal system. The scenario results evidence the influence of the northerly winds in the upwelling intensity, as well as in the availability of nutrients and the phytoplankton development, whereas the southerly winds induce a downwelling situation and lower surface phytoplankton concentration. On the other hand, we has evidenced that the lagoon outflow influences the upwelling intensity and, therefore, phytoplankton biomass distributions. In overall, the results confirm that the model is a valuable tool for assessing the ecosystem variability of the Aveiro coast and confirm the crucial role played by the physical processes in the phytoplankton in this system.

ACKNOWLEDGEMENTS

This work was supported by the *Fundação para a Ciência e Tecnologia, Portugal* (FCT), through the project SIMCLAVE (POCI/MAR/56296/2004)

REFERENCES

Cardoso A, C., J.F., Lopes *Validation of an ecological model for the Aveiro near coastal zone, Portugal.* Submited to the Journal of Ecological Modelling, 2007.

Droop M. R., Mickelson M. J., Scott J. M. and Turner M. F., Light and nutrient status of algal cells. Journal of the Marine Biological Association of the United Kingdom, 62, 403-434, 1982.

Luyten P.J., COHERENS – Dissemination and exploitation of a coupled hydrodynamical-ecological model for regional and shelf seas, *MAS3-CT97-0088. Final Report. MUMM Internal Report, Management Unit of the Mathematical Models*, 76 pp, 1999.

Michalakes, J., J. Dudhia, D. Gill, T. Henderson, J. Klemp, W. Skamarock, and W. Wang, 2005. The Weather Research and Forecast Model: Software Architecture and Performance. Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology. Eds. Walter Zwieflhofer and George Mozdzynski. World Scientific, pp 156 – 168, 2005.

Moita, M.T., Estrutura, Variabilidade e Dinâmica do Fitoplâncton na Costa de Portugal Continental. *PhD thesis*. Faculdade de Ciências da Universidade de Lisboa, 272 pp, 2001.

Peliz, Á., Rosa, T., L., Santos, A. Miguel P., Pissarra, J. L. . Fronts, jets, and counter-f lows in the Western Iberian upwelling system. *Journal of Marine Systems*. 35, 61-77, 2002.

Proctor R., NOMADS –North Sea Model Advection Dispersion study, MAS2-CT94-0105.Technical Report IV: Model Intercomparison. POL *Internal Document Nº 107*, Proudman Oceanographic Laboratory, 143 pp, 1997.

Skamarock, W. C., and J. B. Klemp, 2007. A time-split nonhydrostatic atmospheric modelfor research and NWP applications. J. Comp. Phys., special issue on environmental modeling. In press.

Stevens, I., Hamann, M., Johnson, J. A., Fiuza, A. F. G.Comparisons between a fine resolution model and observations in the Iberian shelf–slope region Journal of Marine Systems 26, 53-74, 2000.

Tett P. and Walne A. Observations and simulations of hydrography, nutrients and plankton in the southern North Sea. *Ophelia*, 42, 371–416, 1995

Tett P., 1998. Parameterising a microplankton model. *Report*, Napier University, Edinburgh, 54 pp, 1998.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Modelling external intrusion into water distribution systems

P.A. López[a], J.J. Mora[a], V. Fuertes[a] and F.J. Martínez [a]

[a] Multidisciplinary Center for Fluid Modeling, Polytechnic University of Valencia, Cno. de Vera, s/n, Valencia, Spain (palopez@gmmf.upv.es, josmorod@doctor.upv.es, vfuertes@gmmf.upv.es, jmsolano@gmmf.upv.es)

Abstract: This article describes the representation of pathogen intrusion in water distribution systems through experimental and numerical modelling, in order to study one of the phenomena which can cause contamination of drinking water through leakage inside the networks. This situation can happen when negative pressure conditions are achieved in the system, allowing the entrance of water around a leak and thus generating pathogen intrusion, causing a problem of water quality. The modelling process is based on experimental and computational procedures. On the other hand, we present an analysis of the behaviour of intrusion considering the transportation of pollutant again by means of experimental measurements and computational simulations.

Keywords: Water quality modelling, Pathogen intrusion, Water distribution systems.

1. INTRODUCTION

Water distribution systems usually conduct good quality water, which is considered safe drinking water to supply the population and thus satisfy their basic needs of consumption. Water quality in the system depends on the quality level, habitually controlled by the water treatment plant. Moreover, the time of residence and the state of the pipes (corrosion, biofilms, and substances transported deposition) can modify the main parameters of water quality. In addition, the possible entry inside the system of elements surrounding the main itself (the so called pathogen intrusion) can become an additional problem (Lopez, 2001).

Water quality and hydraulic performance in networks are linked by the intrusion, as leaks represent a potential influx of undesirable substances in the system. Hydraulic performance of the system (relationship between the water produced and consumed) indicates the lost of volume and pressure, related to the sealing and conservation status (Fuertes et al, 2002, 2003). The risk of water pollution is related to several factors. In 1998, a classification of pathogens entry routes was made, focused on the level of risk considering the causes resulting by the intrusion (Kirmeyer et al, 2001). The routes that have high risk were water treatment breakthrough, transitory contamination, cross connection and water main repair/break.

The aim of this contribution is to make an approach to the problem of pollution occouring in water supply systems based on the pathogen intrusion, as it has been mentioned earlier. The modelling of the leakage from hydraulic point of view can be done as an analysis of flow through an orifice (Lopez, 2005), remaining valid in cases of overpressure and depression, which can favour the entrance of pollutants. The leakage flow depends on the differences in pressure, the resistant feature of the structural defect, the pipe material and

kind of defect (May, 1994). All these aspects have been considered to make the models presented here.

2. PHYSICAL AND COMPUTATIONAL MODEL FOR PATHOGEN INTRUSION IN WATER MAINS

2. 1 Introduction

The proposed study over external intrusion is based on two main aspects. On one hand, two sorts of hydrodynamic models (experimental and computational) have been developed. For the experimental model, a prototype was built in laboratory representing the leakage of the pipe with an orifice, placed in the throat of Venturi tube. Numerical modelling simulates the prototype in three dimensions, using a program based on Computational Fluid Dynamics (CFD), which displays the fields of hydrodynamic components. It gets the best calibration set for the prototype by comparing the velocity and pressure fields along the model. The contrast made with both models allows establishing the foundation for further pathogen intrusion simulations occurring in distribution systems.

On the other hand, an analysis of the behaviour of intrusion is presented considering the transportation of pollutant modelled through a conservative parameter, in this case water salinity. By this modelling, the mixture is studied, as well as the diffusion and transport of pollutants from the area of intrusion through the meshed structure of the model.

By both, mathematical and physical models, it is intended to have a better knowledge of quantities that can not be measured, such as velocity fields, aspects of turbulence, pressure fields, concentrations, etc. existing in mixing processes related to external intrusion. Through computational modelling, we want to study conditions under steady state of the leak and the subsequent mixture, entry and diffusion of the pollutant within the pipe to see in detail the phenomenon that occurs and complement the experiences that will be developed in the laboratory.

2.2 The physical Model

In order to represent the studied flows, and the intrusion phenomenon, the first component to be considered here is a suitable assembly developed in the laboratory (Figure 1). It is oriented to calibrate the computational model with a completely controlled experiment in which the measurements can be very precise.

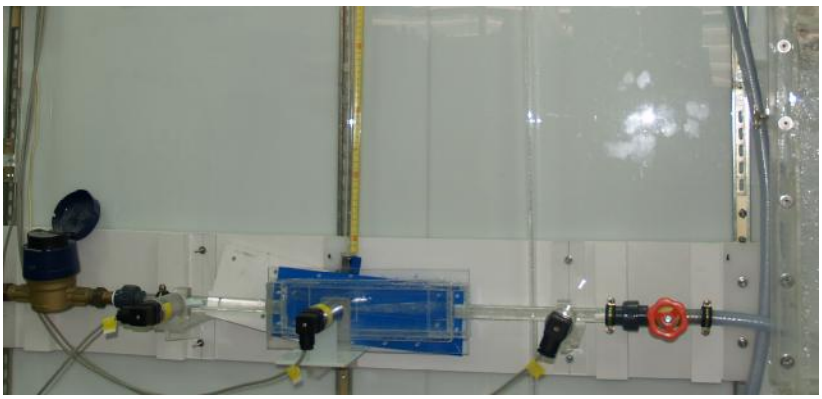


Figure 1. Experimental model for the analysis of the flow into a Venturi tube.

The depression has been represented with a Venturi tube in order to force the intrusion inside the main conduct. Thus, we can observe and model the entrance of external flow inside the principal flow. The Venturi effect is an example of Bernoulli's principle in the case of incompressible fluid flow through a tube or pipe with a constriction. The fluid velocity must increase through the constriction to satisfy the equation of continuity, while its pressure must decrease due to conservation of energy: the gain in kinetic energy is supplied by a drop in pressure or a pressure gradient force. Due to this correlation between the drop of pressure between the zones with different diameters and the circulating volume through the conduit, we can introduce and control de external flow inside the main water flow.

The experiment has been designed to measure pressure and velocity in certain points. A suitable interface has been designed by means of LABVIEW. In Figure 2, a computer screen shows the points in the laboratory assembly where flow and pressure measurements are taken. The intrusion flow was taken in a volumetric way; in each one of the simulations was establish a specific time in order to obtain this intrusion flow.

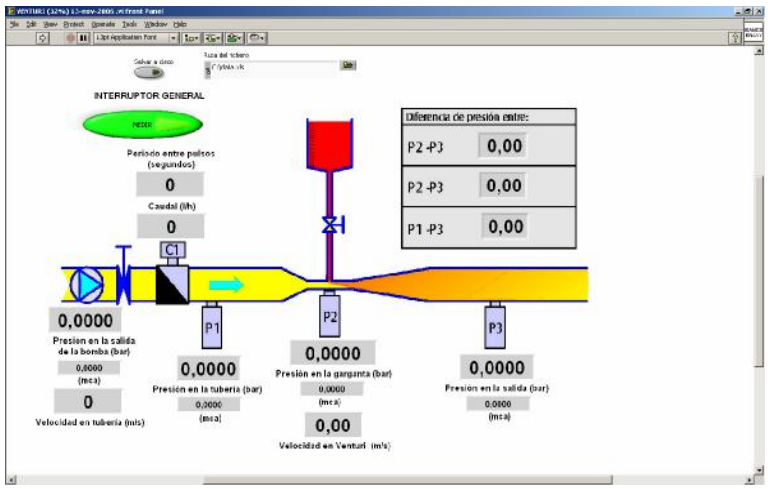


Figure 2. Design of control screen for the measurements in laboratory experience.

In this physical model each measurement was repeated seven times for each one of the nine different apertures of the main valve.

2.3 Computational Model

In order to compare the results of the measurements and to visualize many other aspects, this case has been implemented within the software ©FluentInc. As already indicated, the computational model solves numerically the governing laws of Fluid Dynamics. These equations, taking into account turbulent phenomena, are solved in a geometrical domain, given a number of suitable boundary conditions. In a CFD the relevant magnitudes (velocity, pressure and temperature) are calculated in a discrete manner at the nodes of a certain mesh or grid and they are represented along the mesh.

©FluentInc is a CFD code that uses a numerical method based on the discretization of the space by means of so-called finite volumes (Fluent, 2007) and it offers a number of CFD options for the steady state solution, which can be used in this case. Specifically, incompressible flow and steady state have been considered, together with water as the current fluid under turbulent regime.

The conservation equations solved by the code are those of mass and momentum. The continuity or mass conservation equation solved by ©FluentInc and it is used in the hydrodynamic study of the present problem is the following one:

$$\frac{\partial r}{\partial t} + \nabla r\vec{v} = S_m \quad (1)$$

Where ρ is the fluid density, v its velocity and S_m the mass source contained in the volume of control. For other geometries, suitable coordinates, namely spherical or cylindrical, should be used. Also, the momentum equation is considered by the following expression:

$$\frac{\partial(r\vec{v})}{\partial t} + \nabla r(\vec{v}\vec{v}) = -\nabla p + \nabla \bar{t} + r\vec{g} + \vec{F} \quad (2)$$

Here p is the static pressure, $\vec{g}$ and $\vec{F}$ the gravitational and outer forces defined on the control volume, respectively, and $\bar{t}$ the stress tensor defined by:

$$\bar{t} = m\left[\left(\nabla\vec{v} + \nabla\vec{v}^T\right) - \frac{2}{3}\nabla\vec{v}I\right] \quad (3)$$

Where μ is the eddy viscosity, I is the unit tensor and the third term accounts for the effect of the expansion of volume.

The geometry in study was constructed in AutoCAD® 2006 considering the sections where pressure was measure in the physic model. The mesh was designed on Gambit 2.2.30, in this software it established the volume control and the solid and liquid elements, defining the computational dominium and constructed a *Tgrid* mesh (Figure 3).

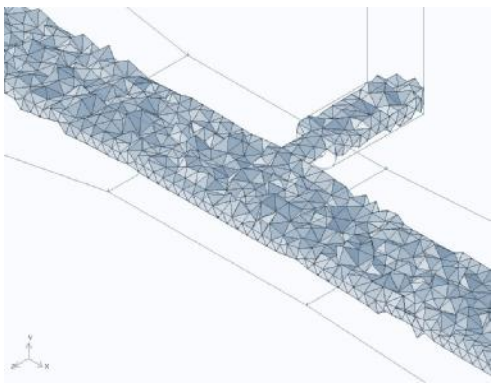


Figure 3. Details of the generated mesh.

Once the mesh has been built and suitably refined for the solution of the equations to be optimal, it is necessary to define the boundary conditions. These conditions must be compatible with the solution strategies of the calculation code.

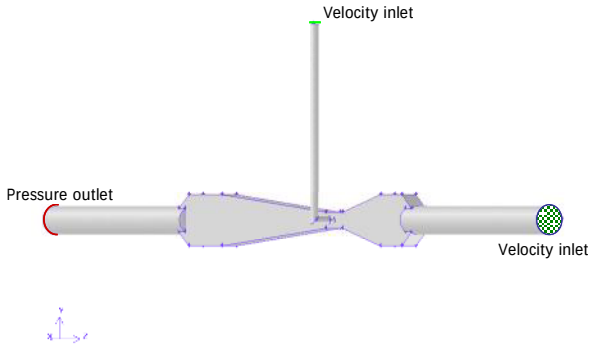


Figure 4. Boundary conditions.

On the ©FluentInc software, the boundary conditions shown in figure 4 were defined. The model works with velocity inlet boundaries at the entry of the flow on principal and intrusion sections; and pressure outlet boundary on the final section of the main tube. The hydraulic diameter method and the turbulence intensity were considered for boundaries.

Focus on the conditions of this model, the *RNG k-e* turbulent model is the best option. It considers in its equations some terms that improve the accuracy for flows that are filtered

rapidly, solve accurately the phenomenon of vortices in the flow, and also manage with differential effective viscosity to consider effects of small Reynolds numbers.

In order to improve the accuracy of the solution, second-order discretization was used for numerical method. The gradient option was modified from cell-based to node-based in order to optimize energy conservation; this option is more suitable for tri-elements meshes (Fluent, 2005). Finally, the calculation made requested approximately 300 iterations for every one of the experimental cases.

3. ANALYSIS OF RESULTS

3.1. Hydrodynamic model analysis

After making the experimental and computer simulations, we show the results which describe the agreement of both models. Table 1 shows the results obtained from the experimental simulation. These data are the necessary information to establish the boundary conditions for the numerical model.

In the Venturi throat section, the pressures were negatives as expected in almost all the different nine experiments for different flows. This forced negative pressure is the cause of the intrusion of flow in the main current.

Table 1. Experimental flows for the different experiments (l/h)

	Principal	Intrusion		Principal	Intrusion
1	341,7	45,4	6	2068,8	50,7
2	540,4	45,4	7	2316,7	53,2
3	1074,4	45,5	8	2548,6	55,9
4	1325,2	46,2	9	2857,1	60,1
5	1828,6	48,1			

Figure 5 shows the contours of velocity and pressure for maximum, intermediate and minimum flow in the intrusion zone. In those figures we can observe that the maximum flow occurs after the intrusion section, this maximum velocity registered on the computational model was 13,42m/s. On the same way the minimum pressure is established just on the corner of the intrusion tube, presenting a negative pressure of -70.000Pa. When the flow is reduced, also the velocity and the pressure tend to zero.

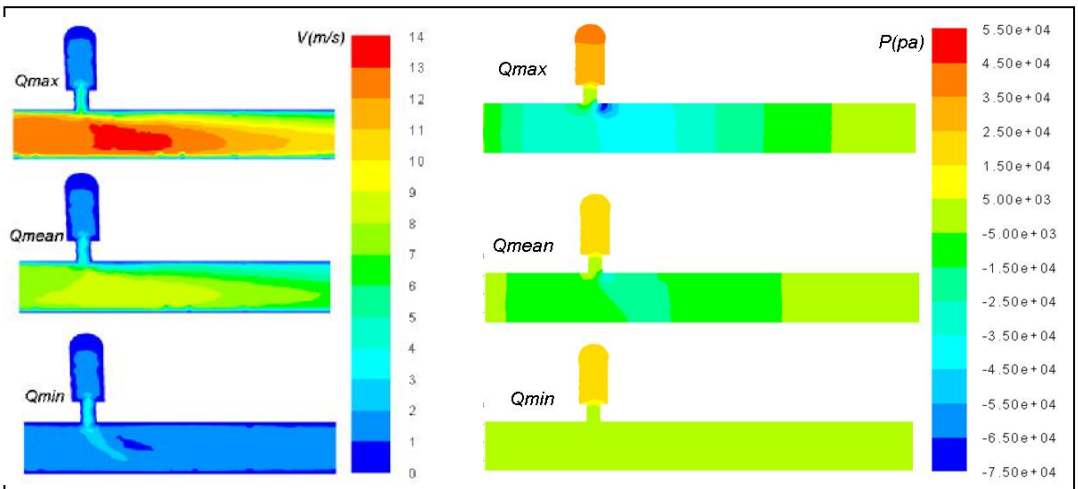


Figure 5. Velocity and pressure fields.

Once the laboratory experiment and the computational model of the problem have been performed, we can analyze both results together. These results, after computational model calibration, are shown in Figures 6 and 7. As it can be seen, the agreement for pressure and velocity values between measured and numerical results is satisfactorily achieved.

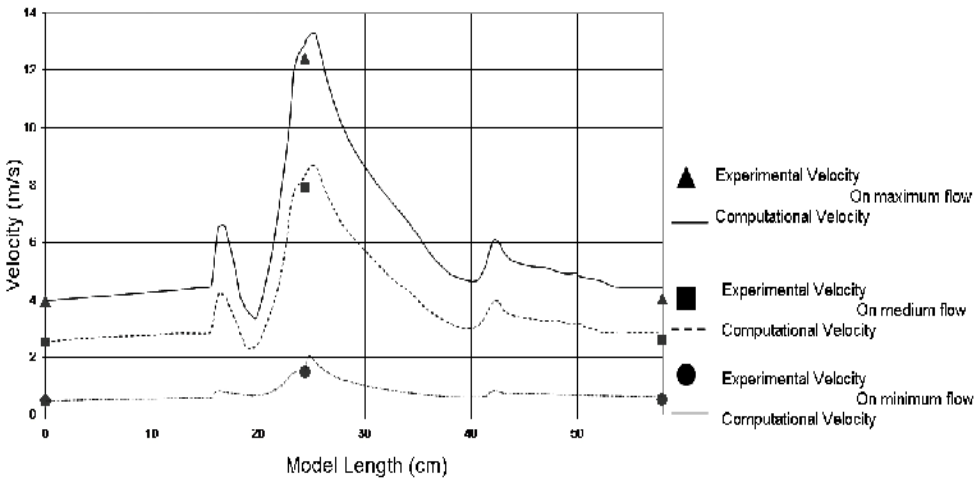


Figure 6. Velocity measurements vs. computational model predictions.

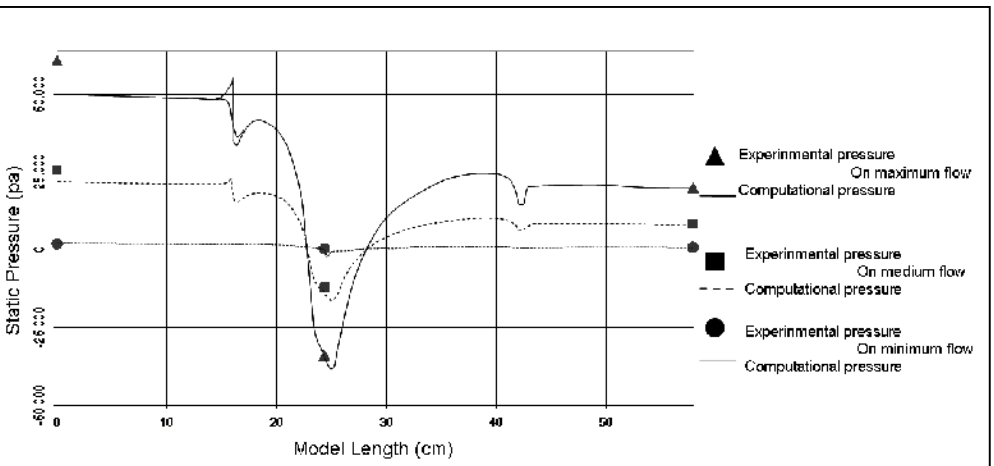


Figure 7. Pressure measurements vs. computational model predictions.

The error was established considering the Nash-Sutcliffe efficiency (*E*), that is one of the most used criteria for the hydrologic evaluation between simulated and observed variables (Krause et al, 2005). It is defined as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period under investigation. It is calculated as:

$$E = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O}_i)^2} \qquad (4)$$

With *O* observed and *P* predicted values. The range of *E* lies between 1,0 (perfect fit) and -∞. An efficiency of lower than zero indicates that the mean value of the observed would have been a better predictor than the model. In this case, we assess the velocity and pressure

of both models in every section of measurement. With respect to the velocity configuration, evaluated in the intrusion and the final sections of the models, the results obtained were 0,9883 of similitude from the intrusion section, and 0,9516 on the final section. On the same way, the pressure configuration was evaluated, in this case considering the first section and the intrusion sections of the models, the results were 0,9260 on the first section, and 0,9988 on the intrusion section. All of these results present more of the 92% of similitude between the computational and experimental models. Considering that the predicted values were adequate for the representation of the experimental model.

Computational modelling is consistent with the physical prototype adequate in the flow range considered. This conclusion leads us to consider the computational model as an entity that represents in detail the velocity and pressure fields, in those points where measurements are not available or in points that require contrast.

3.2. Quality model.

Based on the hydrodynamic model of the intrusion, some experiments to represent water quality in the intrusion mixture have been carried out. In this case, water quality intends to be represented through a conservative parameter such as electrical conductivity. A little volume of water with a huge electrical conductivity is used to represent the pollutant intrusion.

The physical experiments were conducted on the following schedule: the principal flow represents a main in the network water and the initial flow rate on the prototype, this higher volume has a low conductance. On the other hand, we represent the flow of intrusion with a small volume but with a high salinity concentration. Finally, we capture the information flow resulting from the mixture. The measurement of the volumetric flow rate of pollutant intrusion was carried out by three tests for each one of the 11 cases modelled for the main flow.

Two different conditions were considered in the mathematical model, first one represents the main flow with a low electrical conductivity; second one represents the flow generated through the intrusion, with a high concentration of salinity and thus a much higher electrical conductivity than that existing in the main flow.

For this case, in order to solve conservation equation for chemical species, ©FluentInc predicts the local mass fraction of each species, Y_i, through the solution of a convection-diffusion equation for the *i*th species (Fluent, 2007). This conservation equation takes the following general form:

$$\frac{\partial}{\partial t}\left(rY_i\right)+\nabla\cdot\left(r\mathbf{v}Y_i\right)=-\nabla\cdot\mathbf{J}_i+R_i+S_i \qquad (5)$$

Where R_i is the net rate of production of species *i* by chemical reaction and S_i is the rate of creation by addition from the dispersed phase plus any user-defined sources. In turbulent flows, ©FluentInc computes the mass diffusion in the following form:

$$J_i=-\left(rD_{i,m}+\frac{m_t}{Sc_t}\right)\nabla Y_i \qquad (6)$$

Where Sc_t is the effective Schmidt number for the turbulent flow (it varies with the turbulent viscosity, m_t, and the turbulent diffusivity), and $D_{i,m}$ the mass diffusion coefficient for species *i* in the mixture.

To specify boundary conditions, mass flow inlet condition was established for the principal and intrusion entrances. This boundary condition defines both flow and density of water and also the pressure. The density of water was measured in the experimental model in addition of hydrodynamic parameters indicated on the experimental simulations.

Table 2. Input parameters on the numerical model

Simulation	Principal inlet boundary	Instrusion inlet boundary	Oulet boundary
	Mass flow (kg/s)	Mass flow (kg/s)	velocity (m/s)
1	2,86E-05	1,59E-04	0,309
2	5,97E-05	8,36E-05	0,625
3	8,58E-05	1,15E-04	0,898
4	1,06E-04	3,59E-04	1,121
5	1,23E-04	3,78E-04	1,307
6	1,44E-04	3,66E-04	1,517
7	1,61E-04	3,66E-04	1,699
8	2,18E-04	3,01E-04	2,281
9	2,51E-04	3,64E-04	2,629
10	2,76E-04	3,71E-04	2,887
11	3,04E-04	3,80E-04	3,177

We simulated the same conditions established on the hydrodynamic model, the simulations were in steady state and converge were presented between 250 and 300 iterations. As a result we present the mixture in which can observe the percentage of maximum mass flow contaminant of 6,10% in the last section of the model. The configuration of the concentration of contaminant flow can be seen as follows (Figure 8).



Figure 8. Salinity configuration when the percent of max flux contaminated is 6,10% on the output of the model

To calibrate the transport model some comparisons between the theoretical, experimental and computational modelling have been done. The mean absolute error of computational model from the result of theoretical and experimental study is 0,090 and 0,082. The certainty of the computational modelling is 91,0 and 91,8%. Again in this case the agreement between the modelling predictions of the mixture concentration and the theoretical calculations were adequate. The computational model let us visualize the mixture region along the section of the tube.

4. CONCLUSIONS

Contamination of drinking water due to exposure to biological and chemical pollutants is a very important cause of diseases. One of the causes for this sort of water pollution is external intrusion of pathogen agents across defects in the pipes. In this contribution we have studied the phenomenon of external pathogen intrusion in water mains by means of simulation models.

We have developed two ways of solving the problem: by physical prototype and numerical computations; by both mathematical and physical models. We have intended to achieve a better knowledge of quantities that can not be measured, such as velocity fields, aspects of turbulence, pressure fields, concentrations, etc. presented in mixing processes when unusual situations occur in the system.

Through computational modelling, we want to study conditions under steady state of the leak and the subsequent mixture, entry and diffusion of the pollutant within the pipe to see in detail the phenomenon that occurs and complement the experiences that will be developed in the laboratory.

5. ACKNOWLEDGMENTS

This article has been made possible through actions of the CMMF researchers, involved in the following projects: *DANAIDES: Desarrollo de herramientas de simulación para la caracterización hidráulica de redes de abastecimiento a través de indicadores de calidad del agua. REF. DPI2007-63424. Ministerio de Educación y Ciencia de España.*

REFERENCES

Fluent, TUTORIAL 1, Introduction to Using FLUENT: Fluid Flor and Heat Transfer in a Mixing Elbow, © Fluent Inc. January, 2005.

Fluent User Manual. 2007. Internet url: http://www.fluent.com/.

Fuertes, V.S., García-Serra, J., Iglesias, P.L., López, G., Martínez, F.J., Pérez, R., Modelación y diseño de redes de abastecimiento de agua, Fluid Mechanics Group. Polytechnic University of Valence, Spain, 2002.

Fuertes, V.S., Iglesias, P.L., Izquierdo, J., López, P.A., López, G., Martínez, F.J., Pérez, R., Ingeniería hidráulica en los abastecimientos de agua, Fluid Mechanics Group. Polytechnic University of Valence, Spain, 2003.

Kirmeyer, G.J., LeChevalier, M.W., Pathogen Intrusion into Distribution Systems [Project #436], http://www.awwarf.com/exsums/436.htm, American Water Works Association Research Foundation (AWWARF), USA.

Krause, P., Boyle, D. P., Bäse, F., Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89-97, 2005.

López, P.A. Metodología para la calibración de modelos matemáticos de dispersión de contaminantes incluyendo regímenes no permanentes. PhD Thesis Disertation, Polytechnic University of Valence, 2001.

López P.A., Fuertes, V., Iglesias P.L., Martínez, F.J., Modelación mediante cfd de fugas en tuberías de redes de abastecimiento. Ponencia en V Seminario Iberoamericano de planificación, proyecto y operación de sistemas de abastecimiento de agua. ISBN 84-89487-20-0 Valence, Spain. December 2005.

May, J. Pressure-dependent leakage, World Water and Environmental Engineering, October 1994.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Has the ICT Revolution Run its Course?

K. I. Carlaw[a], R.G. Lipsey[b] and Ryan Webb[c]

[a] *Department of Economics, University of British Columbia, Kelowna, Canada* **(***kenneth.carlaw@ubc.ca***)**

[b] *Department of Economics, Simon Fraser University, Burnaby, Canada* **(***rlipsey@sfu.ca*)

c Department of Economics, Queen's University, Kingston, Canada
(*webbr@qed.econ.queensu.ca*)

Abstract: We assess the contention that the force of the ICT revolution was spent when the dotcom bubble burst. We identify the fundamental principle of *flexible machine logic*, the use of which distinguishes the group of modern ICTs — the electronic computer, the Internet and some related technologies — that make up the GPT of *programmable computing networks* (PCNs) and that have driven the ICT revolution. We seek to place PCN on two logistic trajectories that describe its evolving efficiency and applications. We argue that PCN still has significant scope for increasing its efficiency, that it is still creating an undiminished (potentially accelerating) flow of new applications, and conclude that the force of the ICT revolution will continue for some decades.

Keywords: ICT revolution, Technological change, Sustainable economic development**)**

1. INTRODUCTION

For some decades most of the world has been living through a period of revolutionary change induced by what is commonly called the information and communication (ICT) revolution — the economic, social and political transformations currently being driven by a cluster of technologies centered on the electronic computer and the Internet. Elsewhere (Lipsey 2002 and Lipsey Carlaw and Bekar 2005: 108-119, hereafter LCB), we have made the case for the existence of this revolution and replied to critics, such as Robert Gordon (2000 and 2003) and Triplett (1999) who deny its importance. We say no more about these issues here except to note that the profound transformation of the social, political, and economic aspects of our society wrought by the computer, the internet, and a few other related electronic technologies began in the late 1970s and continued through the year 2000.

When the dotcom bubble burst early in the 21st century, many observers argued that the ICT revolution had run it course and so ceased to be a source of economic change and new commercial opportunities. In this paper, we examine this contention. We identify what we call *programmable computing networks* (PCN) as the fundamental general purpose technology that is driving the modern ICT revolution and seek to determine how far along PCN is in its trajectories of efficiency and applications. We argue that the evidence shows that PCN still has significant scope for increasing the efficiency with which it delivers its services, and that it is still creating a vast array of new applications that provide opportunities for profitable economic exploitation, a process that shows no signs of slowing in near future.

1.1 Some Definitions

In this paper we follow LCB in the use of three concepts defined as follows

- A *general purpose technology (GPT)* is a single generic technology, recognizable as such over its whole lifetime, that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects.
- The *spillover effects* of a GPT are effects that spread through the economy well beyond the sector that produces the GPT itself.
- The *facilitating structure*, is the set of actual physical objects, people, structures and institutions in which technological knowledge is embodied, including plant and equipment — what it is, how it works, how it is organized, and where it is located — the internal organisation and industrial concentration of firms, all infrastructure, and all financial institutions.[1]

1.2 The Evolution of a Typical GPT

A new GPT typically begins in a crude form that is only slowly improved and adapted; later in its evolution, as it becomes well developed, its efficiency rises quickly; eventually, however, physical limits are approached, causing gains in efficiency to slow, finally coming to a halt if the GPT remains in use long enough.

As it is with efficiency, so it is with applications. A mature GPT is widely used for multiple purposes. There are, for example, few products and manufacturing processes that do not now use electronic computing power in one way or another. However, the evolution of the applications of a new GPT is driven by two distinct sources. First, as the GPT delivers its services with ever-increasing efficiency, many tasks that were prohibitively expensive with the displaced GPT now became economically viable. The second source of new applications arises because the new GPT makes possible goods, processes and forms of organisation that were technically impossible with the technology that it replaces. Indeed, most of their really transforming effects come because of the new technological possibilities that are created. Because these developments occur slowly at first, when the GPT is in fairly crude single-purpose form, then accelerate as its efficiency and number of uses increases, and finally slow as the potential of the GPT is more fully exploited and physical limits begin to be approached, the cumulative applications of each GPT also tend to follow a logistic time path.

A technology that eventually becomes a GPT typically is initially incorporated into a facilitating structure that has been designed for the incumbent technology that the new GPT is challenging. Slowly, this structure is redesigned to suit the new emerging GPT. When the changes required are deep and long lasting, it is common to refer to a ‘revolution’ being brought about by the GPT in question. For example, computers were first introduced into management organisations designed for handling information on hard copies. Later, as all levels of the structures of management and administration were redesigned to accommodate electronic means of communicating, analysing, and storing information, the organisation of the typical business was redesigned and only then did administrative efficiency rise. And a similar order of events was observed on the shop floor. However, some unforeseen and revolutionary changes wrought by the introduction of computers were things such as the changes to economic, social and cultural interaction, with whole communities and even new cultures emerging and existing strictly online unconstrained by geographic location.

1.3 The Evolution of Efficiency and Applications

LCB stylize the evolving increases in the efficiency, defined as the cost at which the GPT delivers a unit of its service, and the evolving increase in the GPT’s range of applications, defined as products, processes, forms of organization that include the GPT in one form or

[1] The full list is given by LCB.

another, as logistic time paths. Whether it is efficiency or applications that are being measured, the tendency for a logistic time path can be divided into phases.

Phase 1: A new limited-purpose technology that eventually evolves into a GPT is introduced into the facilitating structure that is designed for a pre-existing set of GPTs. Both its efficiency and the number of applications that it spins off increase only slowly.
Phase 2: The facilitating structure is slowly being redesigned to fit the new technology that is evolving into a GPT. This stage is often long drawn out, full of uncertainty, and conflict since the adjustments create many winners and losers. The growth rates of applications and efficiency tend to accelerate.
Phase 3: The principles of the new GPT are applied to produce many new applications — new products, new processes and new organizational forms — *within a newly evolved facilitating structure that is by now fairly well adapted to it*. The number of new applications grows rapidly and eventually reaches a maximum rate. Efficiency growth is also reaches its maximum rate of growth during Phase 3 of the efficiency trajectory.
Phase 4: The opportunities for applications of the GPT to create new product, processes and organizational technologies (and to improve existing ones) diminish as does the rate at which its efficiency is rising.
Phase 5: If the GPT remains in use long enough, the scope for further increases in efficiency and applications may be exhausted, often because physical limits of one sort of another are reached. The relevant part of the curve is then horizontal.[2]
Each of these phases varies from one GPT to another, depending on the productivity potential of each and how it is exploited. For example, Phases 1 and 2 may be very short if efficiency increases rapidly quite early in the GPT's evolution. It may enter its Phase 4, or even Phase 5, while applications are still in their Phase 3.

1.4 The Co-evolution of Applications and Diffusion

The evolving set of applications of a GPT covers both the development of new applications and the spreading use of existing applications, i.e., diffusion.

1.4.1 New applications

In some cases, new applications may depend solely on the GPT in question and a number of other lesser technologies, not themselves GPTs. In other cases, the GPT may cooperate with other GPTs. To consider such cooperation further we need to define two new terms.

- *A primary application of a GPT* is an application whose nature is mainly influenced by the characteristics of the GPT itself.
- *A background, enabling application of a GPT* is an application for which the GPT is a necessary condition but whose nature is not mainly determined by the characteristics of that GPT.

When a new GPT is evolving, its changing characteristics influence the nature of the innovations that depend on it — it is contributing primary applications. Eventually, however, it gives way to even newer GPTs which become the prime movers in the application generating process and while it becomes a background enabler.

1.4.2 Diffusion

When a new application is developed, its diffusion is often slow, costly and uncertain. Just to discover what is currently in use throughout the world is a daunting task, particularly for small firms. Even if a firm can identify best practice techniques, this (at best) provides it with a blueprint; learning how to produce successfully what is described in a blueprint implies acquiring all the tacit knowledge that goes with adopting something new. It follows

[2] In LCB, Phase 5 refers to the time at which a new GPT arrives to challenge the incumbent. For our purposes, it is more useful to use the term for the time at which either new applications cease to be developed or efficiency ceases to increase.

that the existing set of technologies does not provide a freely available pool of immediately useful knowledge. Furthermore, adapting technologies in use elsewhere to one's own purpose often requires innovation. As a result, innovation of new applications and the diffusion of existing ones interact in a system of mutual causation; they shade into each other rather than being clearly distinct activities.

As we will see in a later section, this blurring of the creation of new applications with the diffusion of existing applications creates a challenge when we seek to identify the applications curve for computer related technologies. The data we present does not permit a clear distinction between the creation of new applications and their diffusion although they do give us important information.

1.5 The Co-evolution of Efficiency and Applications

Next, we enquire into the co-evolution of the efficiency and the applications of any new GPT, which we illustrate with a stylized example. At one extreme, we can imagine a GPT that arrives with an initial level of efficiency (a fixed blue print) that cannot be improved. At this extreme, one initial set of potential applications has been enabled, with no possibility for it to be widened. Of course, many of these will not be known or even guessed at when the GPT first arrives. The set of applications will be developed over time after the arrival of the GPT, typically following a logistic pattern. At the other extreme, we can think of a GPT whose efficiency goes on increasing, and hence the cost of its services declining, indefinitely. As this happens, the application curve may still be logistic but Phase 3 will be prolonged because efficiency is continually increasing. If efficiency finally stabilizes at some given level, the applications will continue to evolve for some time, but will eventually enter Phase 4 as the pool of potential applications begins to be exhausted.

In most real cases, both of these developments are occurring simultaneously: the efficiency with which the GPT delivers its services and its range of applications are increasing together in a positive feedback loop where increases in one encourages increases in the other.

2. INFORMATION AND COMMUNICATION TECHNOLOGIES (ICTs)

We define ICTs as all technologies that communicate, analyse, transform and store information, including speech, writing, the printing press, and many more modern technologies, such as telegraph, telephone, radio and computer.

2.1 The Nature of Information

Because information is neither matter nor energy, the physical laws of conservation do not apply to it. It can be copied and used without loss or depletion and its use is *non-rivalrous* in the sense that one person’s use of a piece of information, does not preclude someone else from using it simultaneously. We refer to the physical, or objective form of information as a *signal*. These are of little use unless they can be transmitted from one user to another and interpreted by the receiver. Thus, an essential part of useable information is the process by which information is transmitted, analyzed, organized, and received. Communicating information requires a transmission medium such as the electromagnetic spectrum, or air for sound waves. Throughout history, humans have developed technologies to improve their ability to communicate, to a great extent by improving the transmission process. The physical limitations of speech led to the development of the principle of *inanimate transmission*, the use of inanimate objects to transmit information. The invention of writing enabled information to be encoded, stored and transported over long distances in an accessible form, bypassing the need for direct human-to-human speech. The printing press further improved transmission by enabling this encoded information to be duplicated and mass-produced efficiently, but with the important limitation that the information still needed to be physically moved from place to place.

2.2 Machine Logic

By the 19th century, knowledge of electricity had advanced to a stage where an electrical current could serve as a long-distance transmission medium. Although humans have senses for detecting and analyzing signals sent through many types of transmission media, such as sound waves and the visible light spectrum, they cannot monitor an electrical current to receive signals. This requires an application of the principle of *machine logic*, which is the process of applying an inanimate logical system to a transmission medium in order to analyze a signal for some purpose, such as encoding/decoding, tabulating, organizing or filtering, and/or converting it from one transmission medium to another. The ability to impose a logical structure on a transmission medium allows a transmitted signal to be analysed without direct human intervention.

The modern electronic computer makes use of the important principle of *flexible machine logic*, which is machine logic with the added characteristic that it can be altered without altering the internal physical structure of the device that uses it. For example, the early electronic calculating machines that were developed during World War II had their logic built into their physical design. In contrast, the important property of modern computers is that the logic of the system is alterable through the use of software without a physical re-arrangement of its circuits, which are hard wired according to the principles of machine logic.

2.3 Programmable Computing Networks (PNC)

A *network* is a system of flexible machine logic dispersed across multiple nodes. In much of the literature, a distinction is made between computers and the best known network, the Internet. However, a computer is itself a network since it is comprised of a set of logical components, each designed for specific tasks and communicating with each other via a system bus. Going to a lower level of aggregation, one of the computer's components is a pre-designed set of logical gates that communicate in series (or even in parallel). This is also a network. At higher levels of aggregation, modern supercomputers are comprised of a networked set of personal computers communicating in parallel. At an even higher level of aggregation, distributed computing networks span the entire Internet. All of these examples perform one fundamental function: they use a logical electronic system to communicate and to manipulate information.

This understanding of the interrelationship among computers and other networks based on flexible machine logic leads us to regard them not as two or more distinct GPTs, but as a single GPT of *programmable computing networks (PCN)*, which is composed of all logical processors of information that use flexible machine logic. This includes the computer plus all electronic information networks that use flexible machine logic, including the Internet, local area networks (LANs), wide area networks (WANs), and wireless networks. We use the term programmable computing networks to distinguish this grouping from both the narrower class of computers as they are usually defined, and the wider class of ICTs as defined earlier.

Two other GPTs, namely electricity and lasers, are important technologies used by PCN. The resulting applications are mostly background enabling applications of electricity and lasers and primary applications of PCN, i.e., they are generated by the evolution of PCN not by those of electricity or lasers. Of course, both electricity and lasers also have a variety of primary applications that extend well beyond PCN.

3. PLACING PCN IN ITS EFFICIENCY CURVE

In Section 3.1, we use some indexes of efficiency to argue that PCN is still in Phase 3 of its efficiency curve. We then disaggregate to study the major sources of this increasing efficiency. The efficiency of any computing device can be improved in two basic ways: *electrical engineering process advancements*, such as the ever-shrinking etching size of a transistor, and *logical advancements*, such as the stored program computer, multi-core processors, and software optimization — the latter including both the evolution of the

software instruction set *and* the logical arrangement of electronics on any piece of computer hardware. Engineering process advancements bring greater speed, miniaturization, and capability to the electronics, while logical advancements optimize the organization of the electronics. The effects of advancements in both of these ways are enhanced by the exploitation of scale effects. We study each of these sources of efficiency changes in sections 3.2, 3.3 and 3.4. In section 3.5 we use these discussions to look into the future.

3.1 Increasing Efficiency of PCN

Since there is no way to look at the overall efficiency of so complex a general purpose technology as PCN, we do this piecemeal by examining a selection of the many existing quality-adjusted price indexes for various devices that incorporate PCN. As a result, the indices for personal computers, microprocessors, software, and telecommunications equipment serve as a proxy for changes in the efficiency of each of these technologies that utilise PCN.

In their Figure 3.1 Carlaw, Lipsey and Webb (2007) (here after referred to as CLW)[3], show a hedonic real price index for computers purchased by business and government. The figure shows a constant rate of price decrease, which is about an order of magnitude every decade for at least four decades.

Similar trends are observable in U.S. real price data, compiled by the Bureau of Economic Analysis (BEA) for computer related prices, those of mainframes, PCs, disk storage devices, tape storage devices, terminals, printers and other peripherals. The price in indices for each of these products declines by several orders of magnitude from 1958 to 1994.

A matched-model price index for integrated circuits, both microprocessors and memory modules, is given in Figure 3.2 of CLW. The decrease in price over the period 1975 – 2001 has been roughly 5 orders of magnitude. Since uncorrected market prices did not change anything like as much, it follows that most of the marked decline in the price index is explained by quality increases.

Figures 3.3 and 3.4 in CLW show hedonic indices using U.S. Bureau of Economic Analysis data for both microprocessors and memory chips, albeit for a shorter time period. Here we can observe a price decrease close to two orders of magnitude per decade for micro processors and memory chips.

All of these indices show that the quality-adjusted prices of devices that embody and implement the GPT have been falling continuously. We conclude that PCN has had an efficiency increase of at least five orders of magnitude since its introduction. Since these increases as yet show no sign of slowing, it seems clear that PCN is still in Phase 3 of its efficiency curve.

3.2 Advancements in Engineering Processes

Early in the lifetime of the modern electronic computer, transistors and integrated circuits produced large efficiency gains. Later Moore's Law, first stated in 1965, predicted that the number of transistors per integrated circuit would increase steadily while the cost per transistor would fall, resulting in about a doubling of the performance/cost ratio every two years (see also Stokes, 2003). Indeed the past 40 years have seen great technological advances in all of the areas which determine the minimum average cost of producing integrated circuits—the maximum number of transistors per square inch, the size of the wafer, the average number of wafer defects per square inch, and the costs associated with producing and connecting many small integrated circuits to perform the function of a larger one.

Recently, the cost of further increasing the computational capacity of CMOS has risen significantly. An upper bound in wafer size may be reached before long due to inherent

[3] All figures referred to in the text are found in Carlaw, Lipsey and Webb (2007).

material limitations, such as crystal strength (Chaterjee and Doering, 1998). Serious technical problems related to transistor leakage (small amounts of current flowing through an "off" transistor), power density (the amount of heat generated in each area of the chip) and power usage (energy required to power the chip) impede further attempts to reduce etching size. (Chaterjee and Doering, 1998). These issues are primarily responsible for what has been popularly termed the "megahertz wall," referring to the recent difficulty chip manufacturers have had in increasing processor clock speeds. While a continued increase in the number of instructions that can be executed per second is still theoretically possible, this requires a major shift in how software developers handle the instruction flow sent to the processors to take advantage of multiple cores.

The semiconductor industry is moving on two tracks to combat these problems. One attempts to deal with the issues directly by engineering smaller transistors, new materials, and integrating new technological breakthroughs such as nanotechnology with CMOS-based devices. It appears, however, that we are nearing the final decades of efficiency gains realizable with such devices (ITRS, 2005). But note that a decade or two is not tomorrow! The second track is research into post-CMOS devices that we discuss in Section 3.5.

3.3 Advancements in Logic: optimisation and functionality[4]

Programmable computers have a hierarchy of software and hardware logic[5] One key to the device's efficiency is that more complex functions can be created from simpler ones by designing a higher level instruction that is composed of other instructions (including more basic ones from lower levels in the hierarchy). As one level in the instruction set hierarchy grows in functional complexity, further potential for efficiency gains arise by shifting some of those functions to lower levels of the hierarchy. However, efficiency can be reduced by having too many instructions either at lower levels where they increase the complexity of the hardware logic, or at higher levels where they increase the complexity of the software logic layers. A process called 'optimization' is used to maximise the overall efficiency of the device by balancing these factors. A subset of this process, software optimization, repeats this task within the software layers.[6] Although the feed back between adding functional complexity and optimizing the placement of instructions in the logic hierarchy tends to be positive on average, bottlenecks or salients can be created, causing efficiency improvements generated by logic advances to evolve irregularly.

Logical advances offer myriad ways of improving efficiency but are ultimately constrained by the physical engineering structures of PCN. Thus they will encounter limitations as CMOS itself encounters physical limitations. The development of multi-core processors and "grid" computing have potential to side-step this issue, however to be scaleable, they must be accompanied by advances in parallel programming. Software developers schooled in serial programming must first be re-trained to take advantage of the extra cores.

3.4 Exploitation of Scale Effects

A significant portion of the efficiency gains from PCN have resulted from dealing with information networks. Because these networks are just scale increases in the size and effectiveness of computing devices, they amplify the efficiency gains for computers by exploiting the latent scale effects of electronic information networks. These scale effects come from at least three sources.

---

[4] Although we simplify our discussion in this sub-section by focusing on the logical design of the electronic computer, we emphasize that the same analysis applies to the logical design of other networks that we include in PCN (or any other inanimate communications structure).

[5] This is really a "reverse hierarchy" because it is an upside down pyramid, with the basic instructions at the bottom rather than at the top.

[6] One reason why the instruction set may become overly complex is the demand for a device to perform a wider variety of functions. There are many examples of software designers adding instructions for functional reasons to the detriment of the device's performance, often described as "bloat." Adding functions usually alters the optimal distribution of instructions. Thus the design of a device's logic needs to create an effective balance between the allocation of functions (optimization) on one hand and functionality on the other (Stokes, 2005).

First, the physical delivery grid of the network is a source of increasing returns. Over the latter half of the 20th century, by far the largest proportion of our network traffic used electrical current as a transmission medium, usually via the telephone or cable distribution networks. Recently, however, fibre-optic technology, based on the GPT of the laser, has provided vast potential for gains in network bandwidth capacity, only some of which has been realised so far. While the control components of network infrastructure of PCN is still implemented by means of electronic components, efficiency improvements in fibre-optic cabling are driving a replacement of electricity as the transmission medium, resulting in scale effects from bandwidth increases.

Second, the information that is transmitted in a network is non-rivalrous in consumption. This allows for a positive externality in the sense that the cost of transmission does not rise with the number who receive it. When exploited, this is another source of increasing returns.

Third, there is a classic network externality. As new users join a network, all existing network members gain a benefit, creating a positive externality. This is a source of increasing returns since the larger the network the more is the total benefit from the externality.

The continuing exploitation of these scale effects is one of the major causes of the geometric progression that we observe in our indexes of PCN's efficiency and performance and there is no reason to believe that the potential for such exploitation has been fully exploited.

3.5 Looking into the Future

The overall judgment for PCN is that it is still well within Phase 3 of its efficiency curve, but in the absence of unexpected breakthroughs in engineering and logical processing, it will in a decade or two enter Phase 4 with steadily falling rates of efficiently growth. In contrast, we argued in Section 3.4 that there are unlimited potential scale effects that can be exploited by advancemans in either engineering or logical processes.

Note also that when efficiency gains do slow this need not herald an imminent slowing of new applications and a consequent slowing of the social and economic gains from PCN. First, as discussed in Section 1, new applications typically continue to be developed, often for decades, after a stabilisation in the efficiency with which the main GPT delivers its services. Second, new technologies may replace CMOS based devices with newly enabled applications. For example, the International Technology Report on Semiconductors has laid out a roadmap for moving beyond an "ultimately scaled" CMOS, "accomplished by … extending the CMOS platform via heterogeneous integration of new technologies and, later, via developing new technological and nano-architectural concepts." (ITRS, 2005) An example of such a possibility is the quantum computer.

Although estimates vary, the engineering consensus is that a move to post-CMOS devices will be made by about 2020. It is impossible to say at this stage whether the replacements will fall under our definition of PCN or be better regarded as a wholly new GPT. For instance, quantum computing may present a completely new technical architecture with its own tremendous potential for efficiency advancements and wide range of uses. So although CMOS-based devices may reach their efficiency limits in a decade or two, new engineering technologies that integrate well with existing and future logical structures of PCN are on the horizon with promises of further engineering gains and new applications. So even if the GPT of PCN moves into its phase 4, the ICT revolution may continue, driven by some non-CMOS-based GPT.

4. PLACING PCN IN ITS APPLICATIONS CURVE

In Section 4.1, we measure the diffusion of some of the major technologies spun off from PCN. In Section 4.2 we deal with current and perspective new applications.

4.1 Diffusion

The quantitative data available to us mainly show market diffusion of specific aggregate categories, such as "personal computers" and it is impossible to get sufficient sectoral data to aggregate up to a measure of PCN's overall diffusion.[7]

Figure 4.1 of CLW shows that the number of personal computers, one of the most prolific embodiments of PCN, sold each year in Canada began to climb in the early 1980's with an accelerating growth rate through the 1990's and mid 2000's. There was a temporary slowdown in diffusion in the early 2000's, which paralleled the economic slowdown during this period. International data for this aggregate category reflects a similar temporary slowdown, followed by a quick resurgence.

Figure 4.2 of CLW shows the stock of PCs held. Fully 30% of Canadian households did not have a personal computer at the terminal observation and the absolute growth in household adoption remained relatively constant over the time period. Although the majority of Canadian households have PCs, the rising sales suggest a mixture of faster replacement due to faster obsolesce and a rising number of multi-computer households.

Figure 4.3 of CLW shows a pronounced growth in Internet users comparable to the growth in personal computer ownership. Since 2000, the gap between PC ownership and Internet use has closed because a larger number of people are using the same personal computer as more members of a household go online and because more devices with Internet access are becoming available, the mobile phone being a primary example.

Figure 4.4 of CLW shows that the growth in the number of unique domains hosted on the Internet over the past two decades has been close to exponential. Figure 4.5 of CLW shows the same data for worldwide hostnames, adding data for active names. These data are useful for examining the collapse of the "Internet bubble" on Internet hosts. Web usage appears to have grown in spite of the economic slump in the early 2000's.

Figure 4.6 of CLW gives data for digital mobile (cellular) phones, which undertake a growing number of the same tasks as a standard personal computer. Digital phones were introduced in 1996 and ten years later they accounted for 85% of the total mobile phone market and rising. International data shows similar exponential growth and a larger percentage of the total mobile phone market in many European and Asian countries.

Another generic product that embodies PCN technology is the digital camera. The growth of the market for such cameras highlights the role of PCN in replacing earlier technologies, in this case the film-based camera, as shown in Figure 4.7 of CLW.

For yet another example of diffusion, 45 percent of multi-channel TV households had either digital cable or satellite TV service, according to the 2004 edition of an annual report from media researcher Horowitz Associates. This presents many opportunities for new products and services that utilise this medium.

Finally, according to the "2004 Ownership and Trend Report from The Home Technology Monitor", 4% of homes with TV report owning a DVR (such as TiVO) – a figure that had doubled in the previous 6 months; 6% had an HDTV set, up 50% six months previously; 18% a VCR/DVD dual deck; and 5% a PC with a TV tuner. These applications are obviously in their infancy in terms of diffusion. It seems most likely that they will diffuse rapidly over the next few years as more networks convert to HD broadcasts. This diffusion will in turn require more sophisticated home theatre setups and such things as DVRs to fully exploit the new, higher audio and signal and signal quality.

---

[7] In measuring the diffusion of various applications of PCN, we must rely on the evidence collected by various statistical agencies. Much of the useful sector-specific data is collected by industry groups or commercial organizations for commercial purposes and is not available for unfunded academic research. A significant proportion of what is available is given in terms of dollar values, a metric that is unsuited for enumerating applications in the absence of category specific price data. Much of our data comes from the International Telecommunications Union, which is responsible for coordinating the operations of telecommunications networks and services. Where possible, ITU data was corroborated by other international sources including the CIA World Factbook, the OECD, and national statistical agencies including Statistics Canada.

We have reported on those generic products for which we have been able to obtain useful data. Although we are sure that similar patterns exist in many other home electronics industries such as video cameras, home theatre and audio, as well as household appliances, such industry data are not available for academic purposes without a fee.[8] However, on the basis of what we have shown above, it seems safe to assume that although many major applications of PCN have already deeply penetrated many markets, they have yet further to go in most of these.

4.2 New Applications

PCN began to be applied to the creation of new products, processes and organisational forms soon after the first electronic computers were developed in the 1940s. While computers became more powerful but remained large cumbersome machines, the number of application grew slowly. The number of potential applications then increased greatly with miniaturisation. Computing power began to be added to many existing products and processes, as well as enabling the development of wholly new ones. Dramatic ICT-driven changes began to be felt throughout the economy in a major way in the late 1970s, with impacts that grew exponentially in the 1980s and 1990s. Lipsey (2002) fills several pages with an illustrative list of new products, processes and organisational forms that were computer driven during the last three decades of the 20th century.

We enumerate illustrative cases to show that the pace of new applications is still rapid and that many of these suggest further applications that build on them.[9] In contrast, when any GPT is entering Phase 4 of its applications curve, the pace of new applications slows appreciably and many of those that are developed are dead ends in the sense that they do not suggest further applications that build on them. The following cases represent only a sampling, but they should sufficient to show that new applications are still being invented and innovated at a rapid rate and are of the type that in their turn enable yet further applications. Here we briefly mention those on our list. The details are posted on the website sfu.ca/~rlipsey.

- Computing power is still in the process or being added to just about every imaginable kind of consumers' good, from washing machines to children's' toys. Sensors and controls are being built into clothing in a first step towards a new realm of "smart fabrics." Carmakers are putting artificial neural networks into engines to increase fuel-efficiency and reduce pollution. Smart houses and smart office building are being built with of all kinds of newly developed automatic controls that add to comfort, safety, and efficiency.
- Video games, often denounced for their supposed ill effects, are being shown to have a surprising range of therapeutic uses, opening opportunities to develop games specifically designed for such purposes.
- Objects are being sprayed with thousands of tiny microdots that, when read by a computer, give them a unique identity just as finger prints do for humans.
- Researchers have developed a revolutionary new way to control computers by thought alone, opening myriad possible applications including the control of artificial limbs by a computer that intercepts brain impulses and converts them into movement commands for artificial muscles.
- Home buyers with internet access who are looking for finance are no longer at the mercy of their own bank or agent, the thoroughness of whose advice is hard to monitor.

---

[8] The most noteworthy of these are data collected by organisations in the semiconductor industry that detail semiconductor usage, and OECD data for diffusion statistics covering such health applications as MRI and CT scanners. Indeed, most diffusion data are controlled by industry groups, including those covering the majority of home electronics applications.

[9] Our case data have come from various sources, including traditional media outlets (both online and print), relevant mailing lists, press releases, technology websites and the quarterly technology review section of the *Economist* Magazine.

- Authors can publish books and articles and post them on the net for downloading, with free access or as a credit-card purchase, thus opening markets for limited edition publications that were unavailable to authors when all publications had to be in hard copy.
- A nationwide vehicle-tracking service is allowing fleet operators to monitor the performance and location of their vehicles, making it easier to manage the performance of their fleets, reduce fuel costs, analyze driving behaviour and improve delivery time.
- Dairy framers can remain in their living rooms while controlling the movements of their herds, including milking and monitoring health. Diagnostic practices have been greatly aided by computers; more such developments are in the pipeline; others are still in researchers' imaginations.
- A group of national weather centres across Europe is creating a global weather forecasting system that allows meteorologists to make more accurate and more timely predictions. Indeed, virtually anything that has sufficient regularities to allow prediction can be better predicted by high powered computers than methods that were state-of-the-art 10 or 15 years ago.
- User-generated content, best known for fuelling the popularity of Web sites such as You Tube and MySpace, is rapidly taking hold in advertising representing a fundamental shift in the democratization of content.
- MacDonald's has pioneered the centralized handling of orders where an operator located in a central clearing house hears instructions given by drive-in clients from around the continental United States and Hawaii and routes them back to the local kitchen for handling.

The illustrations repeatedly show several key features. First, most of these developments are new and have much scope for direct improvements and further applications. Second, most of them suggest many spinoffs in terms of other new technologies that can exploit those on the list to create different new products and different new processes. Third, many of the items would have seemed like science fiction a mere few years before they were developed, illustrating how difficult it is to predict what new applications of PCN are around the corner. What is clear is that the pace of new developments and applications has not slackened and there is nothing in the nature of these to suggest that it will slacken in the near future.

Yet this is not the end of the story. So far, we have concentrated on developments that are mainly enabled by PCN on its own. Also in sight are many present and myriad foreseeable future applications based on a union of PCN with biotechnology and nanotechnology. A sampling of these is given in Lipsey (1999). They make it clear that the union of PCN with biological and nano science has already become highly fruitful and is spawning a mass of new applications in fields that span most of the economy.

Eventually, as it is with electricity today, PCN will become a mere background, enabling input that is used everywhere in bio- and nano-technology applications that owe their form to biology and physics and not to the evolving structure of PCN. But that is still some way off because further efficiency developments of the PCN are needed before some of these applications can be realised.

4.3 CONCLUSION

The data presented in this section strongly suggest that PCN is still well within Phase 3 of its applications trajectory. Diffusion of existing applications is continuing apace and new applications are being developed almost daily. Inspection of our sample of these strongly suggests that many of them will spawn (or have already spawned) further new applications. We elaborate briefly on each of the sources of future potential for PCN.

- The efficiency of PCN has increased rapidly over an extended period of time (beginning in the middle of last century), and has extended over a large number of dimensions, enabling a succession of wider and wider possibilities for applications.

Although it is possible that, absent a major breakthrough such as the perfection of quantum computing, or a dramatic scaling of multi-core processors, PCN may be approaching Phase 4 of its efficiency trajectory, further increases in efficiency can still be expected for years, possibly a decade or two to come. These efficiency gains will, if past experience is any guide, enable a host of new applications that are either too costly or technically infeasible with today's ICT technology.

- Even if PCN evolves well into Phase 4 of its efficiency curve, new non-CMOS based technologies may take over as the drivers of an on-going ICT revolution.
- We have seen from the sample of applications for which we could get reliable data that the diffusion of PNC applications is far from complete. Since diffusion often goes along with the discovery of new opportunities for innovations, applications that come from this source have yet to be fully exploited.
- Given the logistic behaviour of new applications, even if efficiency stopped increasing today (PCN reached late Phase 4 or Phase 5 on its efficiency curve), many applications would remain to be exploited — a list that no one can enumerate in full since it is in the nature of new knowledge that it cannot be fully anticipated until it is discovered.
- Finally, the union of PCN with biotechnology and nanotechnology will spawn an almost unlimited set of new opportunities for inventions and innovations over at least the next half century. Gradually, these applications will become more and more background enabling applications from the point of view of PCN, and primary applications from the points of view of biotechnology and nanotechnology. But this will be a slow evolution and for some time to come many developments in these two fields will arise from, and will create opportunities for, new developments in PCN.

Given all this evidence, it seems clear that PCN will continue to have a profound and formative influence on the world's technologically driven economic and social change for at least several decades to come, offering countless opportunities for the development and exploitation of new applications throughout much of the economy.

We also conducted a comparison with the GPT of electricity the results of which are also posted on the website sfu.ca/~rlipsey This led us to two further conclusions. First, PCN seems to be about where electricity was on its efficiency curve in the early 1920s, with a decade or two more of efficiency gains in store. Second, if primary applications follow the same path as charted by electricity, then there are also at least two decades of new primary applications in store for PCN and another decade or two of high demand based on the diffusion of already innovated applications.

REFERENCES

Aizcorbe, Corrado and Doms (2003), "Why Do Matched-model and Hedonic Techniques Yield Similar Price Measures?" *Federal Reserve Board Working Paper 2003-14*.

Aizcorbe, (2004), "Product Introductions and Price Measures for Microprocessor Chips in the 1990s," http://129.3.20.41/eps/io/papers/0502/0502004.pdf

Canadian Imaging Trade Association (2006). "Industry Data," http://www.citacanada.ca/News/industry.htm.

Chaterjee and Doering, (1998), "The Future of Microelectronics," *Proceedings of the IEEE*, Vol. 86, No. 1.

Diener, R. A. V. (1989), "Information Science. What is it? … what should it be?" *American Society for Information Science*, 15(4), 17.

Dominion Bureau of Statistics. *Cost of electricity for domestic service and monthly bills for domestic service, commercial light and small power*. Ottawa: Queen's Printer, 1926-1957.

Goldstine, H. H. (1993), *The Computer from Pascal to von Neumann*, Princeton University Press, Princeton, NJ.

Goodall R., D. Fandel, A. Allan, P. Landler, H. Ruff (2002), "Long Term Productivity Mechanisms of the Semiconductor Industry," *American Electrochemical Society Semiconductor Silicon 2002 Proceedings*, 9th Edition.

Gordon, Robert, J. (2000), "Does the 'New Economy' Measure up to the Great Inventions of the Past?", *NBER Working Paper*, 7833.

_____ (2003) "Hi-tech Innovation and Productivity Growth: Does Supply Create Its Own Demand?", NBER Working Paper No. W9437

Grimm, (1998), "Price Indexes for Selected Semiconductors, 1974-1996," *Survey of Current Business*, February.

Home Technology Monitor (2006) "Ownership and Trend Report"
http://www.knowledgenetworks.com/htm/index.html.

International Telecommunications Union, World Telecommunications Indicators Database (2005). http://www.itu.int/ITU-D/ict/statistics/

Internet Systems Consortium (2006), "ISC Internet Domain Survey," http://www.isc.org/index.pl.

ITRS (2005), *International Technology Roadmap for Semiconductors for 2005*, http://www.itrs.net/Links/2005ITRS/Home2005.htm.

Lipsey, Richard G. (2002) "The Productivity Paradox: A Case of the Emperor's New Clothes", *ISUMA: Canadian Journal of Policy Research*, **3** (1) 120-126.

_____ (1999) "Sources of Continued Long-Run Economic Dynamism in the 21st Century" in *The Future of the Global Economy: Towards a Long Boom*, (Paris: OECD), 33-75.

Lipsey, Richard G., Kenneth I. Carlaw and Cliff Bekar (2005), *Economic Transformations: General Purpose Technologies and Long-run Economic Growth*, Oxford University Press, Oxford, UK.

Moschovitis, C.P. (2003), *History of the Internet : A chronology, 1843 to the Present*, ABC-CLIO, Santa Barbara, CA.

Netcraft (2006), "Netcraft Websurver Survey,"
http://news.netcraft.com/archives/web_server_survey.html

Schement, J. and B. Ruben, Eds. (1993), *Between Communication and Information*. Information and Behaviour. New Brunswick and London, Transaction Publishers.

Schurr, S. et al. (1990), *Electricity in the American Economy : Agent of Technological Progress*, Greenwood Press, New York, NY.

Stokes, Jon (2005), "Introducing the IBM/Sony/Toshiba Cell Processor," *Ars Technica*, http://arstechnica.com/articles/paedia/cpu/cell-1.ars/.

Stokes, Jon (2003), "Understanding Moore's Law," *Ars Technica*, http://arstechnica.com/articles/paedia/cpu/moore.ars/1.

Triplett, J.E. (1999), "Economic Statistics, the New Economy, and the Productivity Slowdown", *Business Economics*, 34(2): 13–17.

Triplett, J.E. (1996), "High-Tech Industry Productivity and Hedonic Price Indices," in Organization for Economic Cooperation and Development, *Industry Productivty*, Paris, pp. 119-42.

www.SiliconFarEast.com, 2000.

Wylie, P. (1989), "Technological Adapatation in Canadian Manufacturing, 1900-1929," *Journal of Economic History*, Vol. 49, No.3.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

National Accounting Frameworks for Measuring Information and Knowledge Economies

David Thorns[a] and Les Oxley[b]

[a]School of Sociology and Anthropology, University of Canterbury, Christchurch, New Zealand (david.thorns@canterbury.ac.nz)

[b]Department of Economics, University of Canterbury, Christchurch, New Zealand (les.oxley@canterbury.ac.nz)

Abstract:

The paper explores the problems and challenges posed by the growth of the information and knowledge economies for national accounting frameworks and measurements. The conventional approach has been to add to rather than reconstruct such indices. One of the consequences is the focus this leads to on "information technologies" and their prevalence and fails to fully appreciate the transformation that is required to enable "information" to be transformed into "new knowledge" and therefore create the "new economy. The paper argues for a recasting of the analysis of the new economy around an appreciation of the changing nature of physical capital brought about by the creation of new global data bases and transaction information. However alongside of this is an increased importance to intellectual assets, human and cultural capital and the rise of new networked relationships none of which are easily incorporated into national accounts

Keywords: Information Technologies; New Economy; Human Capital; eResearch

1. INTRODUCTION

One of the key questions that persists in the literature and discussions about the arrival or otherwise of the knowledge economy relates to the lack of consensus around what would constitute such an economy and whether there is indeed a significant transformation taking place in the basis of advanced economies so that knowledge is increasingly the driver of production [Carlaw et al., 2006]. Economists typically (but not exclusively) focus more narrowly than sociologists upon the changed role of knowledge in economic activity. For example the OECD defined a KBE as "*economies, which are directly based on the production, distribution and use of knowledge and information* [OECD, 1996]. In both the work of sociologists and economists it is the importance of the digital technologies, the Internet, computers, information and globalized networks that these technologies enable that have been stressed. It is now the "age of speed" time and space have been compressed [Harvey, 1989; Virillio, 2004]. There is an increasing shift of activities to computers rather than these being carried out in specific locations. Testing of products can now be done through simulation on the computer. People can work from home [Felstead et al., 2005]. People can create virtual worlds in "my space" and live out their lives in cyberspace [Castronova, 2001]. Whilst not all are involved in these activities it does extend the range of possibilities and gives more prominence to 'mental' labour rather than physical labour carried out in discrete places. Knowledge is now seen as the primary source of competitiveness and the desire of governments is increasingly to create innovative and 'smart citizens' to work in the "smart economy". Incorporating into what constitutes knowledge the "cultural and creative" sector extends the discourse on the knowledge

economy and gives recognition to this sector as a potential contributor to economic growth. Through this emphasis on knowledge as the driver of economic activity knowledge itself has become a commodity and becomes traded across global networks creating new patterns of international migration – the brain drains and brain gains. New Zealand data shows that there is a brain exchange taking place with both inflows of highly qualified and losses of New Zealand nationals, especially those under 30 [MRST, 2006; Wang, 2007].

Given that the idea of the knowledge economy is somewhat imprecise it is little surprising that the measurement of this change to the basis of economic activity has similarly been problematic. One of the problems that is encountered here is the shift back to an emphasis on information when measurement is required. Consequently we have a raft of "measures" that are now widely used by national and international organisations of "information based activity and these usually comprise distributional measures of such things as the ownership and spread of computers and internet connections, attempts to plot changes in the occupational structure and the rise of "information or ICT related workers". These measures are then generally used to infer that we are becoming more "knowledge dependant". However, at a more robust theoretical level it is necessary to identify the transformation that takes place between the possession of information and the creation of knowledge. Here educational level and skills are crucial and hence a focus has been upon the levels of formal education and qualifications within advanced economies. Though here again simply counting the number and courses covered does not in itself indicate that the recipients had obtained or continue to utilise "knowledge" gained. The problem here is that quantifiable measures are not sensitive to the qualitative changes that are part of moving from the acquisition of information to the deployment of "knowledge". Nor are they sensitive to the relationship between explicit – formal and codefiable knowledge and tacit – more informal knowledge often acquired through experience

The knowledge economy we see as involving a change in the relationship between physical and human capital in production with a greater emphasis moving to ideas and innovation rather than the creation of large scale physical plant. This encourages new forms of economic organisation such as the partnership rather than joint stock companies and may create the possibilities of firms which do not have physical assets – all their wealth lying in the collective knowledge and experience they have accumulated. This raises interesting questions about the role of intellectual property and how ideas would be traded in such changed markets. In this new "weightless" economy intellectual assets usually seen by economist in terms of human capital, but maybe in such economies there is now also the need to appreciate the role of social and even cultural capital as knowledge creation may engage people in new forms of association and networking – both virtual and real. Such networks can be facilitated by ICT technologies that are extending our capabilities to interact face to face with people in real time across the globe. Here the expansion of "social networking" has been an important illustration of the power of such connectivity to create new forms of social connection and new commercial opportunities as people begin to trade in virtual space and virtual goods.

For the national and international bodies trying to provide "measures" of this changing economic and social world and at the same time retain the existing frameworks of measurement such as in national accounting systems many problems are presented. Often national and international bodies are slow to change as they are protective of their data and the need to ensure comparability between data sets over time. This can lead to resistance to shift outside of the existing measurement frameworks thus we tend to see additions being made to accommodate change rather than a more radical rethinking of the framework themselves. This raises the critical question of whether or not this can be achieved.

In the next section we ask the question whether the present system of national accounting introduced during WW11 as a measure of wartime production capacity, can be adapted to allow us to see the emergence of information and then knowledge economies.

2. NATIONAL ACCOUNTING FRAMEWORKS AND MEASUREMENT OF THE KNOWLEDGE ECONOMY

Many authors have discussed the requirements for, and problems with, measuring the knowledge\information economy most of the data on which the measures are based comes from the national accounts of the various countries involved. This does raise the question as to whether or not these accounts are suitably designed for this purpose. There are a number of authors who suggest that in fact the national accounts are not the appropriate vehicle for this task. For example Howitt argues that ". . . the theoretical foundation on which national income accounting is based is one in which knowledge is fixed and common, where only prices and quantities of commodities need to be measured. Likewise, we have no generally accepted empirical measures of such key theoretical concepts as the stock of technological knowledge, human capital, the resource cost of knowledge acquisition, the rate of innovation or the rate of obsolescence of old knowledge" [Howitt, 1996: 10]. Howitt goes on to make the case that because we cannot measure correctly the input to and the output of, the creation and use of knowledge, our traditional measures of GNP and productivity give a misleading picture of the state of the economy. Howitt further claims that the failure to develop a separate investment account for knowledge, in much the same manner as we do for physical capital, results in much of the economy's output being missed by the national income accounts.

In Carter [1996] six problems in measuring the knowledge economy were identified: 1) The properties of knowledge itself make measuring it difficult, 2) Qualitative changes in conventional goods: the knowledge component of a good or service can change making it difficult to evaluate their "levels of output" over time, 3) Changing boundaries of producing units: for firms within a knowledge economy, the boundaries between firms and markets are becoming harder to distinguish, 4) Changing externalities and the externalities of change: spillovers are increasingly important in an knowledge economy, 5) Distinguishing 'meta-investments' from the current account: some investments are general purpose investments in the sense that they allow all employees to be more effcient,5 6) Creative destruction and the "useful life" of capital: knowledge can become obsolete very quickly and as it does so the value of the old stock drops to zero. Carter argues that these issues result in it being problematic to measure knowledge at the level of the individual firm. This results in it being difficult to measure knowledge at the national level as well since the individual firms' accounts are the basis for the aggregate statistics and thus any inaccuracies in the firms' accounts will compromise the national accounts.

Haltiwanger and Jarmin [2000] examine the data requirement for the proper measuring of the information economy. They point out that changes are needed in the statistical accounts which countries use if we are to deal with the information\knowledge economy. They begin by noting that improved measurement of many "traditional" items in the national accounts is crucial if we are to understand fully IT's impact on the economy. It is only by relating changes in traditional measures such as productivity and wages to the quality and use of IT that a comprehensive assessment of IT's economic impact can be made. For them, three main areas related to the information economy require attention: 1) the investigation of the impact of IT on key indicators of aggregate activity, such as productivity and living standards, 2) the impact of IT on labour markets and income distribution and 3) the impact of IT on firm and industry structures. Haltiwanger and Jarmin outline five areas where good data are needed: 1) measures of the IT infrastructure, 2) measures of e-commerce, 3) measures of firm and industry organisation, 4) demographic and labour market characteristics of individuals using IT, and 5) price behaviour.
In Moulton [2000] the question is asked as to what improvements we can make to the measurement of the information economy.In Moulton's view additional effort is needed on price indices and better concepts and measures of output are needed for financial and insurance services and other "hard-to-measure" services. Just as serious are the problems of measuring changes in real output and prices of the industries that intensively use computer services. In some cases output, even if defined, is not directly priced and sold but takes the form of implicit services which at best have to be indirectly measured and valued. How to do so is not obvious. In the information economy, additional problems arise. The provision of information is a service which in some situations is provided at little or no cost via media such as the web. Thus on the web there may be less of a connection between information provision and business sales. The dividing line between goods and services becomes

fuzzier in the case of e-commerce. When internet prices differ from those of brick-and-mortar stores do we need different price indices for the different outlets? Also the information economy may affect the growth of Business-to-Consumer sales, new business formation and in cross-border trade. Standard government surveys may not fully capture these phenomena. Meanwhile the availability of IT hardware and software results in the variety and nature of products being provided changing rapidly. Moulton also argues that the measures of the capital stock used need to be strengthened, especially for high-tech equipment. He notes that one issue with measuring the effects of IT on the economy is that IT enters the production process often in the form of capital equipment. Much of the data entering inventory and cost calculations are rather meager and needs to be expanded to improve capital stock estimates. Yet another issue with the capital stock measure is that a number of the components of capital are not completely captured by current methods, an obvious example being intellectual property. Also research and development and other intellectual property should be treated as capital investment though they currently are not. In addition to all this Moulton argues that the increased importance of electronic commerce means that the economic surveys used to capture its effects need to be expanded and updated.

In Howitt's view there are four main measurement problems for the knowledge economy: 1)the "knowledge-input problem", 2) the "knowledge-investment problem", 3) the "quality improvement problem", 4) the "obsolescence problem". To deal with these problems Howitt makes a call for better data. But it's not clear that better data alone is the answer, to both Howitt's problems and the other issues outlined here. Without a better theory of what the "knowledge economy" is and the use of this theory to guide changes to the whole national accounting framework, it is far from obvious that much improvement can be expected in the current situation. As Howitt himself says, "[i]f the critical component of this paper has been larger than the constructive component, this is mainly attributable to the fact that the issue at hand is not likely to be fixed by minor tinkering with national income accounting practice. The underlying problem is that the very conceptual foundations on which national income accounting is based assume away the mainspring of long-term economic growth, by taking knowledge as unchanging and freely available. In such a world, market prices and quantities are all one needs to measure economic activity. In a world where growth is based on the creation, acquisition and use of knowledge, however, we need to look at other magnitudes, and a better conceptual foundation is needed before we know just what magnitudes to look at and how" [Howitt, 1996: 26].

One simple question is to which industry or industries and\or sector or sectors of the economy can we tie knowledge\information production? When considering this question several problems arise. One is that the "technology" of information creation, transmission and communication pervades all human activities so cannot fit easily into the national accounts categories. It is language, art, shared thought, and so on. It is not just production of a given quantifiable commodity. Another issue is that because ICT exists along several different quantitative and qualitative dimensions production cannot be added up. In addition if much of the knowledge in society is tacit, known only to individuals, then it may not be possible to measure in any meaningful way. If on the other hand knowledge is embedded in an organisation via organisational routines then again it may not be measurable. Organisational routines may allow the knowledge of individual agents to be efficiently aggregated, much like markets aggregate information, even though no one person has a detailed understanding of the entire operation. In this sense, the organisation "possesses" knowledge which may not exist at the level of the individual member of the organisation. Indeed if, as Hayek can be interpreted as saying, much of the individual knowledge used by the organisation is tacit, it may not even be possible for one person to obtain the knowledge embodied in a large corporation.

3. KNOWLEDGE ECONOMY AND KNOWLEDGE SOCIETY

A further measurement complexity is introduced by the lack of clarity between the concepts of the knowledge economy and society. This has resulted in many of the same variables being used to measure both these entities yet they are differentiated in their content with the Knowledge society being more encompassing and largely about normative aspirations to

create a society where there is universal access to information and information technology on the grounds that these are essential enabling technologies that increase equality of opportunity and democratic citizen participation.

Considerable effort has been made to measure the information\knowledge society by national and international organisation such as UNESCO, the UN and the EU. Despite the difference in frameworks and indicators, there are some common themes. These include human capital, innovation, ICT development and the context dimension. The human capital theme includes variables on the levels of people's skills and education which reflects the size of the pool of educated people. Included in the innovation theme are variables showing innovation investment, procedures, capacities and networks. There are diverse indicators under the ICT theme; yet, they can be categorised as either resources or access. The former refers to the information infrastructure while the latter is related to the accessibility of information in people's life and work. The context dimension always includes variables on socio-economic, political and institutional conditions for knowledge production.

Obviously, these themes are crucial for measuring the knowledge society. However, these measures are not without their pitfalls. One basic problem for these measures is caused by the "knowledge problem". In some cases, knowledge is understood partially and information and knowledge are treated as exchangeable terms. As a result, some documents focused entirely on measuring the information economy while talking about the knowledge economy and society. Other documents mentioned the difference between tacit and explicit knowledge, the distinction between information and knowledge, and thus, the distinction between the information society and the knowledge society while they failed to employ appropriate variables to reflect the distinctions, due to data availability. Among these documents, we do see a gradually shifting understanding and discourse on the knowledge society. For example, "UNESCO World Report: Towards Knowledge Societies" could be seen as a leading document in initiating the paradigm shift from the information society to the knowledge one. It acknowledges that "the idea of the information society is based on technological breakthroughs. The concept of knowledge societies encompasses much broader social, ethical and political dimensions" [UNESCO, 2005: 17]. At the same time, another document prepared by UNESCO on statistical challenges shows difficulties in identifying the relevant data within the existing measurement frameworks.

In addition the knowledge problem raises other issues to do with the choice of indicators in each of the major themes. For example, human capital is measured according to people's formal education and skills based on human capital approaches. This inevitably ignores people's tacit knowledge and knowledge between people. There are a number of sociological studies which show that even within the economic domain people are not rational actors but their economic performance is significantly affected by social, cultural and political structures in which they are embedded. Thus, social capital is a crucial complement of any attempt to measure human capital.

Similarly, the measurement of innovation in these documents seems to focus mainly on the production of scientific knowledge in laboratories. This is inconsistent with the Mode-2 knowledge production initiated by Gibbons [1994] in the knowledge society in which science and society co-evolve. Also the measurement of innovation fails to distinguish the role of inventions from that of innovations. Consequently, it is difficult to see how they can measure the economic value of innovation and at the same time attach a social value to it. Regarding ICTs, it seems that the widely accepted practice is to enumerate the physical infrastructure or, at best, measure access to information. There is a misunderstanding on the relationship between technology and human beings here. It is not technology but human beings and their interactions that constitute so-called society and its institutions. Thus, the function of ICTs is not only their capacity to provide additional new connections but also their potential for opening or closing forms of personal, social and economic capacities, relationships and power-plays [Dutton, 2004]. Mansell and Wehn's INEXSK approach would be a valuable endeavour to integrate the dimension of human beings, their knowledge and ICTs in the knowledge society measurement [Mansell and Wehn, 1998].

Another problem with measures of the knowledge society is confusing the knowledge economy with the knowledge society. Generally, there are two kinds of documents on the

measurement of the knowledge society. One group focuses on measuring the knowledge economy although they mentioned the concept of the knowledge society. The foci of the measurement are human capital, innovation and ICT development. A representative document is "Measuring a Knowledge-based Economy and Society: An Australian Framework" prepared by the Australian Bureau of Statistics. The document's author claims that this framework "does not attempt to cover all knowledge in the economy and society . . . [and] over a comprehensive treatment of a knowledge-based society although it does address those social elements which potentially affect economic change or are affected by it" [Australian Bureau of Statistics 2002, 15]

Another group of documents considers both economic and technological features and social conditions and outcomes of the knowledge society. Two representative documents here would be, "Advancement of the Knowledge Society: Comparing Europe, the US and Japan" [European Foundation for the Improvement of Living and Working Conditions, 2004] and "Knowledge Society Barometer" [European Foundation for the Improvement of Living and Working Conditions, 2004a] published by the European Foundation for the Improvement of Living and Working Conditions. There are some variables reflecting social issues such as social inclusion, quality of life and gender equality in the two documents. However, they failed to see that both the economic and the social are equally important and integrated components in the measurement frameworks. Instead, the social is still treated as the 'leftover' after having identified 'significant' and 'measurable' components for national accounting.

In light of these issues it would seem that a necessary first step along the path towards the correct measurement of the knowledge society\economy would entail the development of a theory of the knowledge society\economy. Such a theory would tell us, among other things, what the knowledge economy is, how - if at all - the knowledge economy\knowledge society differ, how they change and grow, and what are the important measurable characteristics. Based on this, a measurement framework could be developed to deal with, at least some of, the problems outlined.

3. CONCLUSIONS: POINTS OF DEPARTURE AND POTENTIAL FOR PROGRESS

Some features and issues that arise with the KBE are:

1. The increasing (but not exclusive) importance of ICTs in economic and social activities. The access to such technologies will be both enabling (GPTs) and have the potential for exclusion and differential effects - 'winners and losers'. This is not new - the introduction of new technologies has lead historically to winners and losers, both in the short and longer term.
2. The increasing proportion of 'human capital' involved in productive activities.
3. The changing role and importance (and ownership) of intellectual property in productive activities. When information is a key input in economic activity its ownership and control will assume greater importance. Intellectual 'property' v. Intellectual 'commons, will affect access to this resource and there will be winners and losers in the Information Economy as a consequence.
4. The changing nature of the 'theory of the firm'. The current economic theory of the firm is based upon firms having ownership and control of physical capital where 'workers' are employed by owner-managers to work with this owned physical capital. As we move to 'human- capital intensive' firms, the modern theory of the firm is left without much of its theoretical substance. The distinction between owners (of capital) and workers (human capital) becomes blurred and economic theory stumbles.
5. The Information Economy has typically focused upon ICTs, however, the KBE surely stretches into bio\genetic engineering issues where intellectual property, human capital and 'knowledge' have a dominant role. The economic and social impacts of genetic engineering and nanotechnology have typically not been established.

6. Measuring the extent and effects of the KBE can be done either directly (physical\ monetary values of its effects - both positive and negative) and\or indirectly via the consequences of the growth of knowledge where it impinges upon areas such as work practices, employment patterns, social inclusion, wealth\wellbeing\crime\surveillance, environmental and especially energy use consequences etc.

From this it is clear that the key to a quantitative measure of the size, extent and effects of a KBE is a theoretical definition which would necessarily have an important (but not exclusive) role for ICTs., a measure of the size and distribution of 'human capital' and some boundaries where knowledge does and does not contribute to the economy\society.

On the basis of currently available data from the System of National Accounts (SNA) we can measure "OECD-type" measures of the Information Economy. It is generally quite easy to measure the extent and growth of ICT-related goods and services such as, computers per person; internet hosts; IP addresses; email addresses: internet companies; mobile phones, etc., but without other data on things such as speed of connection; use of computers, these summary measures remain simply that, a largely uninformative summary. Furthermore, what are the critical levels of these measures that determine when the economy is to be deemed an "Information Economy" or a "developing Information Economy"? To answer these questions requires a clearer theoretical base to inform the statistical measurement.

We can focus more explicitly on some possible implications of a more networked economy\society. Possible indicators include the composition of labour force transitions including; Hours worked; Flexible work environment (i.e., home based work); Service workers, Knowledge intensive versus non-knowledge intensive sectors. In addition the size, composition and growth of new and emerging sectors could be measured and tracked. The role of ICTs as General Purpose Technologies (GPTs) is postulated to have had effects (mostly delayed) on productivity. One way to track the influence of knowledge on the economy is to measure productivity effects via Total Factor Productivity (TFP) as well as direct measures of technological change.

A key feature of the KBE is the prominence of human capital, Discussion of the role of human capital has a long history in economics going back to at least Smith [1776] and its measurement goes back even further, beginning with Petty [1690]. For up-to-date surveys on the area see Wößmann [2003] and Le et al. [2003]. However, these types of measures remain mainly 'academic' and the national systems of accounts would typically only include such things as occupational and industry level employment\participation\ hours of work data; employment by ethnicity; regional employment differences; qualifications of the employed workforce. These educationally related data can be enhanced directly from University Calendars., educational attainment and curricula composition shift data (from say arts to computer science etc.), however, these are typically anonymous and not linked to specific industry employment\ output effects.

ACKNOWLEDGEMENTS

We wish to acknowledge support from the NZ Marsden Fund and NZ Foundation for Research, Science and Technology which aided completion of this research.

REFERENCES

Australian Bureau of Statistics, *Measuring a Knowledge-Based Economy and Society: An Australian Framework, Discussion Paper 1375.0*, Canberra, 2002.

Carlaw, K., L. Oxley, P. Walker, D. Thorns and M. Nuth, Beyond the hype: Intellectual property and the knowledge society/knowledge economy, *Journal of Economic Surveys*, 20(4) September, 633-90, 2006.

Carter, A. P., Measuring the performance of a knowledge-based economy, In D. Foray and B. A. Lundvall (eds.), *Employment and Growth in the Knowledge-Based Economy*, Paris, 1996.

Castronova, E., *Virtual Worlds: A Firsthand Account of Market and Society on the Cyberian Frontier*, Social Science Research Network, 2001.

Dutton, W. H., *Social Transformation in an Information Society: Rethinking Access to You and the World*. Paris, 2004.

European Foundation for the Improvement of Living and Working Conditions, *Advancement of the Knowledge Society: Comparing Europe, the US and Japan*. Dublin, 2004.

European Foundation for the Improvement of Living and Working Conditions, *Knowledge Society Barometer*, Dublin, 2004a.

Felstead, A., N. Jewson and S. Walters, *Changing Places of Work*, Palgrave Macmillan, New York, 2005.

Gibbons, M., *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, SAGE Publications, London and Thousand Oaks, 1994.

Haltiwanger, J., and R. S. Jarmin, Measuring the digital economy, In E. Brynjolfssonand, and B. Kahin (eds.), *Understanding the Digital Economy: Data, Tools, and Research*, The MIT Press, Cambridge and Mass., 2000.

Harvey, D., *The Condition of Post Modernity*, Basil Blackwell, Oxford, 1989.

Howitt, P., On some problems in measuring knowledge-based growth, In P. Howitt (ed.), *Implications of Knowledge-Based Growth for Micro-Economic Policies*, University of Calgary Press, Calgary, 1996.

Le, T., G. John, and L. Oxley, Cost- and income-based measures of human capital, *Journal of Economic Surveys*, 17(3), 271-307, 2003.

Mansell, R. and U. Wehn (eds.), *Knowledge Societies: Information Technology for Sustainable Development*, Oxford University Press, Oxford, 1998.

Ministry of Research Science and Technology (MRST), *Research and Development in New Zealand: A Decade in Review*, Wellington, New Zealand, 2006.

Moulton, D., and R. Brent, GDP and the digital economy: Keeping up with the changes, In E. Brynjolfsson and B. Kahin (eds.), *Understanding the Digital Economy: Data, Tools, and Research*, The MIT Press, Cambridge and Mass., 2000.

Organisation for Economic Cooperation and Development (OECD), *The Knowledge-based Economy*, Paris, 1996.

Petty, W., and P. Arithmetik, Reprinted in C. H. Hull (1899), *The Economic Writings of Sir William Petty*, Cambridge University Press, Cambridge, 1690.

Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Volumes I and II, R. H. Campbell and A. S. Skinner (general eds.), W. B. Todd (textual ed.) (1981), Liberty Classics, Indianapolis, 1776.

UNESCO, *Towards Knowledge Societies*, UNESCO Publishing, Paris, 2005.

Virillio, P., The third interval, In S. Graham (ed.), *Cyber Cities*, Routledge, London, 2004.

Wang, H., The yellow dragon, the black box and the golden coin: New Chinese immigrants and their contributions to New Zealand's knowledge society. M. A. Thesis, Va. University of Canterbury, Christchurch, New Zealand, 2007.

Wößmann, L., Specifying human capital, *Journal of Economic Surveys*, 17(3), 239-70, 2003.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

ICT Diffusion and Productivity in OECD Economies and New Zealand's Industrial Sectors[1]

Kenneth I. Carlaw[a]

[a] Department of Economics, University of British Columbia, Canada and Department of the Economics, University of Waikato, New Zealand (kenneth.carlaw@ubc.ca)

Abstract: We compare measures of total factor productivity (TFP) to independent measures of ICT diffusion and technological change in New Zealand and a collection of OECD countries. The measures of technological change are investment quality (investment specific technological change or IST) and direct measures of ICT diffusion available in aggregate and in the industrial data for New Zealand. TFP growth is generally uncorrelated and sometimes negatively correlated with the independent measures of technological change. These findings are consistent with two views. First that TFP is not a measure of technological change but ideally (often not operationally) is a measure of free lunches associated with such change. Second, that the productivity bonuses from technological change usually occur with significant temporal lags. Thus, TFP is not a contemporaneous measure of the free lunches from technological change. We conclude that New Zealand is experiencing a similar growth phenomenon to that of many other OECD economies. The general purpose technology of electronic ICT has been diffusing since the mid- - 1970's through a costly investment process that requires many complementary investments to exploit the economic potential of this GPT and the productivity benefits are only recently emerging in most cases. We also conclude that the recent low productivity growth in New Zealand is not necessarily cause for alarm from policy makers. This is because New Zealand like many other OECD economies is investing in technology that has a potentially high future output payoff even if it is not measurable in the TFP statistic.

Keywords: ICT diffusion, Total Factor Productivity, Investment specific technological change

1. INTRODUCTION

A fundamental question for those who study and make policy decisions with respect to economic growth is whether the GPT of modern electronically derived ICT and its multitude of ancillary and complementary technologies together have caused a revolution in production, communication and more broadly societies and cultures, globally? The issue is contentious. In this paper we argue using a combination of pre-existing theory and our analysis of empirical evidence available in New Zealand and the OECD that the answer is yes, at least with respect to productivity and communication. The evidence for this revolution is not found in any single scalar measure such as productivity, but, rather in a number of independent measures (qualitative as well as quantitative) of the actual technological change and its impact. Separating the diffusion of this transforming general purpose technology (GPT) from measured output or productivity gains generated by it is therefore critical.[2] Such a separation is one of the novel contributions of this paper relative to others that have addressed the productivity slow down with traditional productivity analysis. By utilizing a theory of endogenous, technology driven growth which can interpret these separate measures in a coherent framework we also provide a tool that

[1] This research received funding support from the New Zealand Treasury, the Royal Society of New Zealand's Marsden Fund Grant number UOC 101 and Foundation for Research Science and Technology grant number UOWX0305.

[2] Lipsey, Carlaw and Bekar (2005) define transforming GPTs as those that "lead to massive changes in many, sometimes most, characteristics of the economic, social, and political structures."

potentially can be used to identify possible avenues for policy intervention designed to improve the economic performance.

There is little disagreement that computers, programmable computing networks (PCN)[3] (including the Internet) and the myriad supporting complementary technologies that they have enabled, have revolutionized production into the global economy.[4] However, debate persists on whether this technological revolution is creating a social and economic revolution similar to those of the First and Second Industrial Revolutions, with most dissenting views based on neoclassical theoretical interpretations of measures of productivity as the evidence that computers and ICT are not as important as the technologies that drove the earlier revolutions.[5]

Economic historians and students of technology agree that technological change is *the* major determinant of very long-term economic growth (i.e., across the millennia). Yet, over shorter periods, there is debate over what proportion of measured economic growth is due to technological change and what to other forces, such as the accumulation of physical and human capital. Such debates imply that we are able to separate the ultimate effects of technological change from their embodied and proximate effects.[6] This separation is only possible under the theoretical fiction of an aggregate production function because it assumes the existence of generic indexes of labour, capital and the state of technology, and facilitates the easy calculation of productivity growth. Often such assumptions are viewed as sufficient to interpret productivity growth to be a measure of technological change. However, as Carlaw and Lipsey (2003), Lipsey and Carlaw (2004) and Lipsey, Carlaw and Bekar (2005) note this is not the case. In their view, most technological knowledge must become embodied in some real physical component of the work whether it is physical or human capital (including all tacit skills), laws and legal institutions, or social and cultural norms in order to become productively useful. Furthermore, each of these embodiments requires costly investment. Thus, the separation of the contribution of technological change from the contribution of measured factors such as physical and human capital to economic growth is difficult.

Keys to connecting technological change to economic growth lie in identifying specific embodiments of new technology, finding independent measures of these embodiments and determining their contribution to economic growth over a long time horizon.

At the centre of the debate concerning ICT's impact on economic growth is the so called productivity paradox – a combination of a number of stylised and anecdotal observations about the proliferation of computers and ICT with the statistical observation of a decline in the growth rate of total factor productivity (TFP) in many OECD countries, starting in the early 1970s and running through to the middle of the 1990s. The erroneous presumption that underwrites the paradox is that TFP measures technological change in a perfectly, contemporaneously correlated fashion. This interpretation finds its theoretical footing in the neoclassical models of economic growth which employ the concept of the aggregate production function that date back to Solow (1956). In such models it is explicitly assumed that technological change acts as a parameter which instantaneously shifts the aggregate production when such change occurs.

This aggregate production function approach is a simplification of a much broader concept of the aggregate production function that allows for the resource consuming activities of R&D. Examples include, Jorgenson and Griliches (1967), Jorgenson (2001), and Barro (1999), whose approaches involve intermediate production functions, or a meta production possibilities frontier. Two critical features of these approaches are their treatment of R&D as an input and the returns to scale in production functions (or overall production possibilities set). Jorgenson and Griliches (1967) and Jorgenson (2001) treat all lines of

[3] For a full definition of PCN see Carlaw, Lipsey and Webb (2007). Throughout this paper we use the term ICT tp refer to the GPT, however, the GPT is more appropriately identified as PCN.
[4] However, some economist, such as, Young (1992) and Krugman (1996) commenting on Young argue that the lack of high total factor productivity in the Asian economies that experience exceptional growth in GDP per capita through the 1970s and 1980s is evidence that no technological revolution occurred in these economies.
[5] See for example Gordon (2000) and Tripplet (1999)
[6] Of course, technological change and investment are interrelated, the latter being the main vehicle by which the former enters the production process.

production activities as having constant returns to scale, which implies that the part of technological change that involves costly R&D is not measured by changes in TFP. In contrast, Barro (1999) uses production functions that allow R&D to generate expanding product variety or quality with increasing returns to the intermediate R&D inputs. In Barro's case, because of the increasing returns to the intermediate R&D input, there is a Hicks neutral, "Mana from heaven" component of technological change that is measured by changes in TFP and a component of the endogenous technological change generated from costly R&D that is not. All of this leaves open the questions about the meta or all encompassing notion of the aggregate production function and about the appropriate formulation of R&D and knowledge production.

Many of the more recent studies analysing the impact of ICT on economic and productivity growth also utilise this aggregate production function framework. (See e.g., Jorgenson and Stiroh, 2000; Oliner and Sichel, 2000; Schreyer 2000; and Skoczylas and Tissot 2004) They also utilise the traditional national income and expenditure accounts industrial classification categories to identify things like the "ICT sector" of the economy, which is typically defined as the ICT producing sector. (see for example, Pilat and Lee, 2001, who provide an overview of uses of OECD data to measure the impact of ICT on productivity change). These studies for the most part find a surge in productivity in the late 1990's particularly in the US but have difficulty attributing this surge to the ICT sector or to ICT using industries such as those in the service sectors of the OECD economies studied. We argue that there are two basic problems with this approach both owing to the fact that ICT is a GPT. First, as has been argued in Carlaw and Lipsey (2003) and Lipsey and Carlaw (2004), new investment in a new technology is not necessarily contemporaneously correlated with increases in output or productivity. According to Carlaw and Lipsey (2003) and Lipsey and Carlaw (2004) TFP change is ideally a measure of the free lunches associated with technological change but with the caveat that it often does not measure even these properly. TFP change is not a measure of technological change. Second, ICT is not contained (and therefore not measurable) in a single sector of an economy. Rather, its general purpose nature implies that it pervades and impacts on all sectors of the economy and may not lead to any kind of measurable productivity increase. We address both of these issues in this paper. To so requires us to summarise the literature on the subject of GPT driven economic growth.

2. GPT-DRIVEN GROWTH THEORY

The GPT concept was invented by Bresnahan and Trajtenberg in (1992), however, the concept of structured technology dates at least back to Nelson and Winter (1980), Freeman and Perez's (1982) and Freeman and Louca's (2001) Technoeconomic Paradigms and Mokyr's (1990) Macro Inventions. Helpman's (1998) volume on long run growth and GPTs is a collection of the first generation of GPT growth models which were inspired by the earlier work of Bresnahan and Trajtenberg. Since Helpman's book very little has been done to advance the state of GPT theorising with the exception of Carlaw and Lipsey (2001, 2006 and 2007) and Chapters14 and 15 of Lipsey, Carlaw and Bekar (2005). Part of the problem for further development has been that need to step out of the traditional mould of modelling economic growth as a stationary equilibrium concept. Carlaw and Lipsey (2001) first introduced a non-stationary equilibrium concept into GPT modelling and have refined the model to include a number of complexities that the first generation of GPT models could not. The model requires computer simulation to solve to the transitional equilibrium in each time period and is also consistent with the stylized growth facts presented in Jones (1998).

Two models: a basic three sector model, and a four sector model that includes structural adjustment costs, both taken from Lipsey, Carlaw and Bekar (2005) are used by Carlaw (2004) and Carlaw and Lipsey (2005) to assess measures of ICT diffusion in Australia relative to measured TFP. In the models, a sequence of GPTs arrive each at uncertain times and with uncertain productive impacts that diffuse according to a logistic process. The models assume behaviour that results in resource allocations such that a non-stationary equilibrium is generated, which is consistent with observed growth through time. The models have the property that in the absence of future GPTs there are diminishing returns

and growth asymptotically approaches zero. But the arrival of new GPTs rejuvenates and sustains the growth process.

Because these models require a numerical solution procedure that is iterated through several periods they provide an opportunity for Monte Carlo analysis of the assumptions that underlay both endogenous growth modelling and TFP growth calculations. Carlaw (2004) and Carlaw and Lipsey (2005) undertake such exercises and confirm the arguments of Carlaw and Lipsey (2003) and Lipsey and Carlaw (2004) that TFP is not a measure of technological change. Under assumptions of zero structural adjustment costs, TFP is positively correlated with direct measures of technological change. However, it persistently under estimates such technological change. Under conditions of positive structural adjustment costs, TPF growth is negatively correlated with measured technological change and persistently underestimates technological change when a new GPT arrives and overestimates technological change as the GPT matures. In both models TFP change fails to detect the arrival of the big technological shocks of GPTs.

In this paper we use the Lipsey, Carlaw and Bekar (2005) models to interpret the empirical evidence for a set of sixteen OECD economies and nine industrial sectors in New Zealand which show that the pattern of ICT diffusion and measured TFP change are either uncorrelated or negatively correlated.

3. THE NEW ZEALAND DATA

The theoretical framework of Carlaw and Lipsey (2001 and 2006) and Lipsey, Carlaw and Bekar (2005) affords the luxury of observing the growth of technological knowledge directly and independently of output and TFP growth. In real data, however, few such direct measures are available. We therefore adopt two alternative, proximate measures of the rate of technological diffusion. Both have the virtue that they are independent of output growth. But, they are also flawed for a number of reasons discussed below.

The contributions of embodied technological change to TFP growth have been studied in the growth accounting literature. Hulten (1992) and Jorgenson (1966) have focused on the measurement of the efficiency of the capital stock and the effects of measurement errors on productivity estimates. These authors argue that quality change (or Investment Specific Technological (IST) change growth) is difficult to observe, and therefore may not be measured accurately in the National Income and Product Accounts (NIPA). In order to obtain an estimate of the size of error associate with the official capital stock estimates, Hulten used quality-corrected data from Gordon (1990). Gordon found that the official deflators for producer durable equipment overstate quality-corrected inflation in capital goods, and therefore understate increases in capital input.

Following Greenwood et al (1997 and 2000), Carlaw and Kosempel (2004) adopt a computable general equilibrium approach to measuring changes in the quality of investment in Canada. They demonstrate that investment specific technological change (IST) made important contributions to Canadian output growth during the 1961-96 period. One of their key findings is that IST is negatively correlated with TFP since 1974.

IST is calculated by making the unrealistic assumption that the economy, sector or industry under examination in is a perfectly competitive general equilibrium which has become characterized as the Ramsey-Cass-Koopmans model following the pioneering work of Ramsey (1928) Cass (1965) and Koopmans (1965). In this framework the microeconomic decisions of consumers determine the saving rates, levels of consumption and stocks of capital in the economy whose aggregate production capacity is characterised by a constant returns to scale production function defined over capital and labour.[7]

[7] It is important to note that the assumption of constant returns to scale is one on which the entire calculation depends. In the absence of constant returns to scale it is not clear that IST is solely a measure of investment quality. We maintain the assumption here and use the measure as being indicative of the point that TFP does not measure changes in technology. Though this likely implies that our independent measure of technological change, IST, is itself flawed.

Within such a framework, constant income share weights but an increasing capital to labour ratio can only be reconciled by an increasing quality of capital, which is the result that Carlaw and Kosempel (2004) verify empirically. In their analysis the measure of residual neutral technological change, which would be equal to TFP in the absence of increases in investment quality, is negative over much of the period from 1974 onward. They interpret this negative measure to potentially indicate a structural adjustment cost associated with the adoption of the new technology implicit in the high quality capital investments of the sort discussed by David (1990) and Lipsey, Bekar and Carlaw (1998b).

We report here some of our follow up analysis of changes in investment quality and changes in TFP in sixteen OECD countries (where comparable data on national accounts, labour and productivity was available from the OECD) reveals that the negative relationship between IST and TFP change appeared in most of the countries in the data set. The data span the period 1970 to 1997, although the times series are not as long for some countries included in the analysis. Correlations and their significance are calculated by linearly regressing TFP growth on IST growth. This simple procedure allows for easy calculation of correlation and the statistical significance of the correlation between the two rates of change, however, it also has some obviously flawed assumptions. For example, it is unlikely that the relationship between TFP and IST growth is linear. We use it because reveals that there is clearly something wrong with TFP as a contemporaneous measure of technological change. We report these results in Table 3.1.

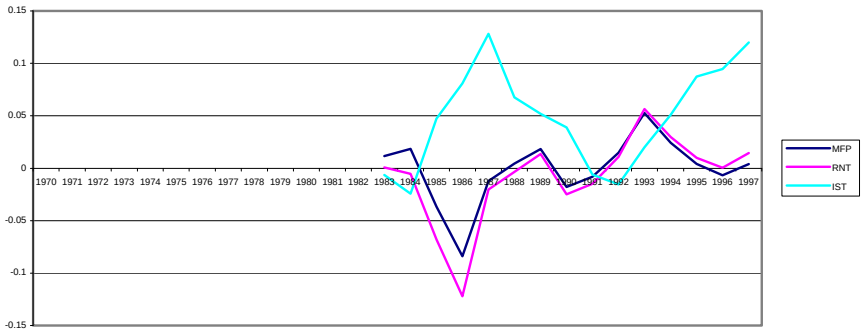
Table 3.1

	Correlation	**Significance (t statistic)**	**Ave. TFP growth**	**Ave. IST growth**
Australia	-0.20	-1.63	0.57%	3.06%
Austria	0.08	0.80	1.62%	1.46%
Canada	-0.04	-0.45	0.49%	6.69%
Germany	-0.90	-1.91	0.24%	1.00%
Denmark	0.06	0.49	0.66%	1.37%
Spain	-0.17	-1.19	0.70%	1.75%
Finland	-0.35	-1.49	0.99%	0.12%
France	0.09	0.66	0.89%	2.23%
United Kingdom	-0.36	-3.45	0.82%	1.11%
Greece	-0.12	-2.57	0.09%	2.57%
Ireland	-0.05	-0.35	1.55%	1.72%
Italy	-0.03	-0.18	0.53%	1.08%
Japan	0.43	2.93	0.96%	3.97%
Netherlands	0.29	2.30	-0.002%	1.75%
New Zealand	-0.22	-1.30	-0.09%	4.90%
Sweden	0.06	0.56	0.40%	2.05%

The results shown in Table 3.1 indicate that the temporal relationship between TFP and IST is not positively correlated. In most cases there is a negative relationship, in two cases a significant one. Only in two cases is there a significant positive relationship. Given the assumptions necessary to make these calculations we do not draw any strong conclusions. But we take this as weak evidence that there is no relationship between our independent measure of technological change and TFP growth. There is possibly a negative relationship over the period examined at least for some economies.

New Zealand is one of the countries which have a negative correlation between TFP and IST change. However, the relationship is not statistically significant for the available data. However, the time series used in the Analysis for New Zealand is quite short (1988-1999) and this no doubt impacts the statistical significance of the coefficient. What is obvious from the Figure 3.1 is that the TFP and IST change appear to move in opposite directions for large parts of the time period.

Figure 3.1: IST and MFP change, New Zealand



What is also obvious from Figure 3.1 is that IST and TFP change has diverged markedly twice in the period, first in beginning in 1984 and continuing until 1992 and again beginning 1993 and continuing until the end of the sample. These dates have obvious significance in New Zealand history. An a possible explanation is that these two periods of regulatory reform sparked investment in higher quality capital which caused productivity to fall but improved the quality of productive technology in New Zealand.

Data produced by Statistics New Zealand and the New Zealand Treasury provide an opportunity to calculate IST growth and compare it to TFP growth at a somewhat disaggregated, industrial level. The data include measures of gross fixed capital formation and their price series by industrial category at approximately the 1 digit ANZSIC level.[8] From these data it is possible to calculate the rate of IST change and compare the diffusion rates to the rate of TFP growth within each industrial sector.

To see the degree of correlation between the rates of IST and TFP change in each industry we again linearly regress TFP growth on IST change for the period 1989-1999 for nine industrial sectors, noting again the flawed assumptions of a linear relationship between TFP growth and the rate of IST change.[9] We use this technique here because it reveals that there is clearly something wrong with TFP as a contemporaneous measure of technological change. Table 3.2 reports the correlation coefficients and *t* statistics for each sector.

Two things are immediately obvious in this analysis. First, in many cases TFP growth is on average below the rate of IST change. Second, the rate of IST change in many sectors appears to be uncorrelated or negatively correlated with the rate of TFP change. Some of the industrial sectors analyzed (all except Mining and quarrying and construction) show a clear increase in IST growth after 1993. In many cases this increase IST growth coincides with a growth rate slow down in TFP. For example Manufacturing, Electricity, gas and water, Business and property services, Personal and community service and Retail and wholesale trade all show a divergence between IST and TFP growth after 1993. In addition, other data on ICT diffusion show the period following from 1993 to be a watershed in New Zealand. Investment in ICT appears to have been at its peak during this period (at least as

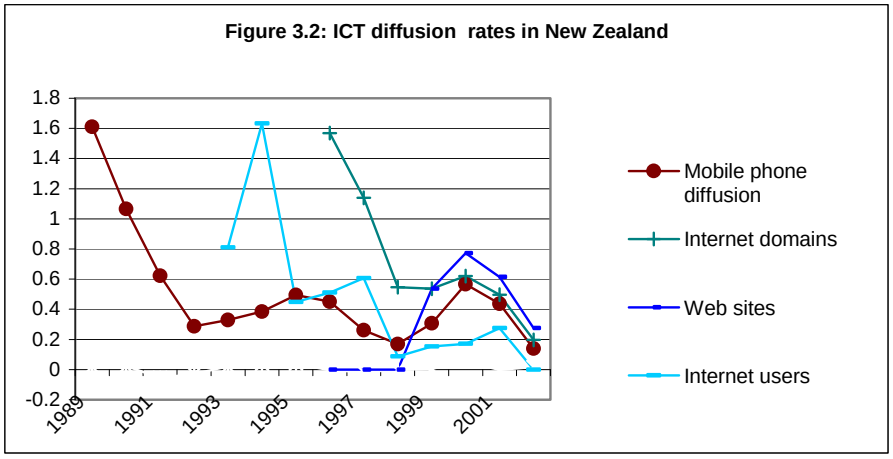
[8] Gross fixed capital formation is calculated by Statistics New Zealand using a chain volume methodology. Treasury has aggregated 31 industrial sectors at the 1 digit ANZSIC level into the 9 industrial sectors. These data are chosen solely because of their consistency of collection, quality and availability. Should more data become available we could broaden our analysis. See Black, Guy and McLellan (2003) for details on this aggregation procedure.

[9] The data are of an annual frequency with the observation taken for the month of June.

far as the available data reveals). Figure 3.2 shows several measures of ICT diffusion rates, which are all high around 1993.[10]

Table 3.2

Industrial Sector	Correlation Coefficient	t-stat.	Ave. TFP Growth Rate	Ave. IST Growth Rate
Primary	0.444755	0.600319	0.013061	0.02585
Mining and quarrying	-0.13522	-0.13445	0.004123	0.019507
Construction	0.545119	0.793985	-0.0184	0.02543
Manufacturing	0.364575	1.007994	-0.00354	0.027529
Electricity, gas and water	-0.09912	-0.2468	0.006498	0.030409
Transport and communications	0.169355	0.541571	0.05626	0.0455
Business and property services	0.091789	0.388129	-0.00964	0.07521
Personal and community services	0.017632	0.054999	0.015462	0.038775
Retail and wholesale trade	0.305878	0.915756	0.003513	0.040282



The diffusion of mobile phones was high from 1989 to 1992. The diffusion of the internet was highest in 1993-94 and the diffusion of other ICT technologies appears to have persisted at a high rate through to 2002.

The data produced by Statistics New Zealand also provide an opportunity to calculate a proximate measure for ICT diffusion rates and compare them to TFP growth rates in the 9 the industrial sectors. The data include measures of gross fixed capital formation of intangible assets which is a category that largely comprises computers and information technology related expenditures by industrial category at approximately the 1 digit ANZSIC level.

The rate of ICT diffusion appears to be uncorrelated or negatively correlated with the rate of TFP change in many industrial sectors. There is a significantly higher volatility in the diffusion rates relative to TFP change as well. After 1993, particularly in the industrial sectors of Business and Property Services, Personal and Community Services and Retail and Wholesale Trade, but through most of the industrial sectors the diffusion of ICT appears to become negatively correlated with TFP change. The time series of data is too short to statistically test sub-periods but this negative relationship between TFP growth and ICT diffusion is consistent with previous analysis of TFP and IST growth.

[10] The data on ICT diffusion comes from the New Zealand Ministry of Economic Development's report titled Statistics on Information Technology in New Zealand: Updated to 2003: URL http://www.med.govt.nz/pbt/infotech/it-stats/it-stats-2003/.

To see the degree of correlation between the two rates of change in each industry we again linearly regress TFP growth on ICT diffusion, noting again the flawed assumptions of a linear relationship between TFP growth and the rate of ICT diffusion. Using this technique again reveals that there is something wrong with TFP as a contemporaneous measure of technological change. Table 3.3 reports the correlation coefficients and *t* statistics for each sector.

Table 3.3

Industrial Sector	Correlation Coefficient	t-stat	Ave. TFP Growth Rate	Ave. ICT Diffusion Rate
Primary	-0.034	-0.892	0.013	0.127
Mining and Quarrying	0.046	1.348	0.004	-0.062
Construction	0.082	2.278	-0.018	0.100
Manufacturing	0.004	0.263	-0.004	0.036
Electricity, gas and water	-0.020	-1.160	0.006	0.106
Transport and communications	0.013	0.957	0.056	0.242
Business and property services	-0.010	-0.408	-0.010	0.078
Personal and community services	-0.001	-0.015	0.015	0.228
Retail and wholesale trade	0.025	1.114	0.004	0.054

The correlations are low in all cases and in some negative. However, none are significant. The high volatility of the ICT diffusion data and short time series of both sets of data preclude making any strong inference. The main conclusion to be drawn from this analysis is that more data is needed to be able to make conduct better statistical analysis.

The analysis of the New Zealand industrial sector data weakly supports the findings of the simulation model and the theoretical predictions of Carlaw and Lipsey (2003) and Lipsey a Carlaw (2004). TFP growth is not correlated with an independent measure of the technological change in the New Zealand economy. For the most part TFP growth and IST change are not significantly correlated.[11] Furthermore, based on available data it appears that New Zealand has been investing in ICT which has manifested in growth of IST at the same time that TFP growth has slowed.

4. CONCLUSION

This paper does two things. First, it adopts the Lipsey, Carlaw and Bekar (2005) and Carlaw and Lipsey (2001 and 2006) models of endogenous growth driven by GPTs in order to begin the development of a theory of TFP. The need for such a theory arrises out of the various, mutually incompatible, interpretations of TFP in the literature. Such a need also arrises out of the inconsistency in the interpretation of TFP growth as a measure of technological change when compared to other independent measures of technological change such as ICT diffusion (Carlaw 2004). The two different measures appear to be uncorrelated or even negatively correlated. Second, the paper utilizes two data sets to measure changes in investment quality which provide a direct measure of technological change that is independent of output growth.

In the adopted models, a sequence of GPTs arrive each at uncertain times and with uncertain productive impacts that diffuse according to a logistic process. The models assume behaviour that results in resource allocations such that a non-stationary equilibrium is generated. The model has the property that in the absence of future GPTs there are diminishing returns and growth asymptotically approaches zero. But the arrival of new GPTs rejuvenates the growth process.

Because this model requires a numerical solution procedure that is iterated through several periods it provides a ready opportunity for Monte Carlo analysis of the assumptions that

[11] These results are to be interpreted with caution since the time period consists of only 10 annual observations in each sector so that any interpretation of statistical significance is suspect.

underlay both endogenous growth modelling and TFP growth calculations. We do such and exercise here and confirm the arguments of Carlaw and Lipsey (2003) and Lipsey and Carlaw (2004) that TFP is not a measure of technological change. We find that while under some conditions TFP is positively correlated with direct and independent measures of technological change it persistently under estimates such technological change. Under other conditions, such as structural adjustment to accommodate a new GPT, TFP growth is negatively correlated with measured technological change and persistently underestimates technological change when a new GPT arrives and overestimates technological change as the GPT matures. In both models, TFP fails to detect the arrival of GPTs appropriately (i.e., as big technological shocks).

The findings in the IST empirical analysis are weak but consistent with the view that ICT is a major new transforming GPT that generates the kind of structural adjustment costs discussed in Lipsey, Bekar and Carlaw (1998b) and Lipsey, Carlaw and Bekar (forthcoming). In ten of the sixteen OECD economies examined TFP showed a negative correlation in a number of cases this correlation was significant. In the New Zealand industrial sector data none of the correlations are significant and two are negative. The interpretation of the sign and significance of the industrial sector data is limited by the small sample of only ten time serries observations.

These empirical findings have to be viewed with a critical eye because there are a number of assumptions necessary to interpret the measures of technological change as being valid. However, they do have the property that they are independent measures of technological change and therefore provide some basis of comparison and testing of the various interpretations of TFP growth. They also point in a common direction. TFP does not measure technological change. Furthermore, it may be negatively correlated with technological change when that change is driven by a transforming GPT such as ICT, which is something that the theory predicts.

What is obvious from the analysis is that more data are necessary to develop a theory of TFP further. A longer time serries of industrial level data would allow for more confident interpretation of the correlation estimates between IST and TFP change. This would also help to identify the characteristics of industrial sectors with respect to emerging general purpose technologies being adopted.

REFERENCES

Abramovitz, M. (1956) 'Resource and Output Trends in the U.S. Since 1870,' *American Economic Review*, 46, 5-23.

Aghion, P. and P. Howitt (1998), "On the Macroeconomic Effects of Major Technological Change", in E. Helpman (ed), *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press), 121-144.

Aghion, P. and P. Howitt (1992), "A Model of Growth Through Creative Destruction", *Econometrica*, 60, 232-396.

Baily, M.N., (2002) "Macroeconomic Implications of the New Economy," Institute for International Economics Working paper, 2002.

Black, M., M. Guy and N. McLellan (2003) "Productivity in New Zealand 1988 to 2002" *New Zealand Treasury Working Paper*, No. 03/06.

Bresnahan, T. and Trajtenberg, E. (1992) "General Purpose Technologies: Engines of Economic Growth?" *NBER Working Paper* No. 4148.

Carlaw, K. I. (2005) "ICT and Australian Economic Performance" in E. Graff (ed.) *Asia Pacific Productivity Conference Research Volume*, Department of Communications, Information Technology and the Arts, 85-120.

Carlaw, K. I. and R. G. Lipsey (2006, January) "GPT-Driven, Endogenous Growth", *Economic Journal* 116, 155-74.

Carlaw, K. I. and R. G. Lipsey (2007) "Sustained Growth Driven by Multiple Co-Existing GPTs" Simon Fraser University Department of Economics Working Paper No. dp 07-17.

Carlaw, K. I. and S. Kosempel (2004) "The sources of total factor productivity growth: Evidence from Canadian data" *Economics of Innovation and New Technology*, 13(4), 299-309.

Carlaw, Kenneth I. and R. G. Lipsey (2003) "Productivity, Technology and Economic Growth: What is the relationship?" *Journal of Economic Surveys*, 17(3).

Carlaw, K. I. and R. G. Lipsey (2002) 'Externalities, Technological Complementarities and Sustained Economic Growth,' *Research Policy*, 31 (8-9), 1305-1315.

Carlaw, Kenneth I. and Richard G. Lipsey (2001), 'Externalities versus Technological Complementarities: a model of GPT-driven, sustained growth', A Paper Presented at The Conference in Honour of the 20th Anniversary of Nelson and Winter's Book *An Evolutionary Theory Of Economic Change*, Aalborg Denmark, 12-15 June 2001.

Cass, D. (1965) "Optimum Growth in an Aggregate Model of Capital Accumulation" *Review of Economic Studies*, 32, 233-240.

David, P. and G. Wright (1999) "General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution", *University of Oxford Discussion papers in Economic and Social History*, Number 31.

David. P. (1990) "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox" American Economic Review, 80(2), 355-61.

Freeman, C. and C. Perez (1988) "Structural Crises of Adjustment, Business Cycles and Investment Behaviour," in Dosi, G. et al (eds.) *Technical change and economic theory*, (London and New York: Pinter; distributed by Columbia University Press, New York), 38-66.

Freeman, Chris and F. Louca (2001) As Time Goes By: From the Industrial Revolution to the Information Revolution, (Oxford: Oxford University Press).

Gordon, R. J. (1990) "The measurement of durable goods prices," National Bureau of Economic Research Monograph series Chicago and London: University of Chicago Press.

Gordon, R.J., (2000) "Does the 'New Economy' Measure up to the Great Inventions of the Past?," *Journal of Economic Perspectives,* 14(4), 49-74.

Greenwood, J., Z. Hercowitz and P. Krusell. (1997). "Long Run Implications of Investment-Specific Technological Change," *American Economic Review* 87(3), 342-62.

Greenwood, J., Z. Hercowitz and P. Krusell. (2000). "The Role of Investment-Specific Technological Change in the Business Cycle," *European Economic Review* 91-115.

Griliches, Z. (1994), "Productivity, R&D, and the Data Constraint", *AER*, 84(1).

Helpman, Elhanan, (ed) (1998) *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press).

Helpman, E. and M. Trajtenberg (1998a), "A Time to Sow and a Time to Reap" in E. Helpman (ed), *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press), 55-84.

Helpman, E. and M. Trajtenberg (1998b), Diffusion of General Purpose Technologies, in E. Helpman (ed), *General Purpose Technologies and Economic Growth*, (Cambridge: MIT Press), 85-119.

Howitt, P. (1998) "Measurement, Obsolescence, and General Purpose Technologies", in E. Helpman ed. (1998) *General Purpose Technologies and Economic Growth*, MIT Press: Massachusetts, 219 – 251.

Hulten, C. R. (1992). "Growth Accounting when Technological Change is Embodied in Capital," *American Economic Review*, 82(4), 964-980.

Hulten, C. R. (2000) 'Total Factor Productivity: A Short Biography,' *NBER Working Paper*, 7471, (Washington: National Bureau of Economic Research).

Jorgenson, D.W. (1966). "The Embodiment Hypothesis," *Journal of Political Economy*, 74(1), 1-17.

Jorgensen, D. and Z. Griliches (1967) 'The Explanation of Productivity Change,' *Review of Economic Studies*, 34, 249-83.

Koopmans, T. C. (1965) "On the Concept of Optimal Economic Growth" in *The Economic Approach to Development Planning* (Amsterdam: Elsevier).

Krugman, Paul (1996) 'The Myth of Asia's Miracle,' in *Pop Internationalism* (Cambridge: MIT Press).

Lipsey, R. G., C. Bekar and K. I. Carlaw (1998a) 'What Requires Explanation,' Chapter 2 in Helpman (1998).

Lipsey, R. G., C. Bekar and K. I. Carlaw (1988b) 'The Consequences of Changes in GPTs,' Chapter 7 in Helpman (1998).

Lipsey, R. G. and K. I. Carlaw (2004) "Total Factor Productivity And The Measurement Of Technological Change", forthcoming in the *Canadian Journal of Economics.*

Lipsey, R.G., K.I. Carlaw and C. Bekar, (forthcoming) *Economic Transformations: General Purpose Technologies and Long Term Economic Growth*, (Oxford UK:Oxford University Press).

Lucas, R. E. (1988) "On the Mechanics of Economic Development," Journal of Monetary Economics, 22(1), 3-42.

Nelson, R. R. (1964) 'Aggregate Production Function and Medium-Range Growth Projections,' *American Economic Review*, 54, 575-606.

Perez, C. (2002) *Technological Revolutions and Financial Capital*, (Cheltenham, UK: Edward Elgar).

Ramsey, F. P. (1928) "A Mathematical Theory of Saving" *Economic Journal*, 38, 543-559.

Romer, P. (1986) "Increasing Returns and Long-run Growth," Journal of Political Economy, 94(5),1002-37.

Rosenberg, N. (1982) *Inside the Black Box: Technology and Economics*, (Cambridge: Cambridge University Press).

Rymes, Thomas K., (1971) *On Concepts of Capital and Technical Change*, (Cambridge: Cambridge University Press).

Solow, R. (1956) 'A Contribution to the Theory of Economic Growth,' *Quarterly Journal of Economics*, 70, 65-94.

Solow, R. (1987) "We'd Better Watch out," *New York Times Book Review*, pp. 36.

Triplett, J.E. 'Economic Statistics, the New Economy, and the Productivity Slowdown' *Business Economics*, 34(2), 1999, pp. 13-17.

Young, A. (1992) 'A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore,' *NBER Macroeconomic Annual*, (Cambridge: MIT press).

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Knowledge Workers and the Knowledge Economy

David Thorns[a] and Hong Wang[a]
[a]School of Sociology and Anthropology, University of Canterbury, Christchurch, New Zealand (david.thorns@canterbury.ac.nz)

Abstract: The paper examines the growth of a new group of workers – knowledge workers seen as critical for the growth of knowledge based economy. The emphasis in much of the literature is upon formal codified knowledge. This focus does not adequately deal with tacit knowledge which is acquired in informal ways from experience and engagement with others. Further when issues of measurement are considered it is around information rather than knowledge that most measures have been based. Currently there is no international agreement on the definition of the ICT sector. The paper reviews definitions adopted by the OECD to show that problem with spillovers from the ICT sector and the difficulty of positioning the ICT sector within existing industry classification persist. The paper concludes by arguing that there is a need for a retheorising of the of the respective roles in the 'new economy of information and knowledge workers and a greater understanding of the interrelationship between explicit formal and tacit informal knowledge

***Keywords*; Knowledge Workers, Tacit Knowledge; Explicit Knowledge; New Economy**

1. INTRODUCTION

In the 1960s and 1970s, a number of scholars began to identify the rise of a new group of workers, which they termed knowledge workers [Machlup, 1962; Drucker, 1969; Bell 1973]. All saw these new workers as essential to the creation of a knowledge economy. More recent work has extended debate about the nature of such workers through looking at the action of knowledge upon knowledge itself as the main source of productivity indicating the significance of abstract thinking for knowledge workers as they need to process symbols as an economic resource [Castells, 1996]. This reflects the greater significance of the symbolic economy in the knowledge age where production has been about the creation of tangible and intangible goods.

Reich [1991] suggested that economic transformation and associated political challenges for the present and next generation required knowledge workers as symbolic analysts who have four basic skills: abstraction, system thinking, experimentation, and collaboration. Pyöriä [2005] held that the symbolic aspects of knowledge work emphasise the abilities of knowledge workers to act as an interface between new technology and human interaction, thus both theoretical and interpersonal knowledge are needed. Finally Stehr [1994] viewed the transmission and application of knowledge as an active process. He emphasised the interaction between theoretical and practical knowledge involved in knowledge-based occupations. A distinction between tacit and explicit forms of knowledge leads to the identification of the range of skills and different forms of knowledge that are now required in knowledge rich economies. This has led to the rapid growth in "knowledge managers" and "change agents" as translators and facilitators of innovation within what has been described as a "knowing capitalism". Thrift sees this as "where circuits of information are embedded in numerous kinds of information technologies" [Thrift, 2005: 1] and the access to data bases, and daily transactional information has increased our capacity to track individual activity and decision making.

2. DEFINING KNOWLEDGE WORKERS BASED ON THEIR KNOWLEDGE AND SKILLS

The emphasis on the codification of theoretical knowledge in the literature has not always been reflected in how knowledge work gets measured. For many there is a retreat back to 'information workers when measurement is required. For example, in their work "The Information Economy", Porat and Rubin introduce the concept of the information sector and define information workers as "those working in occupations whose primary purpose is an output of produced, processed or distributed information, or its infrastructure support" [in Engelbrecht, 2000: 266]. This definition of information workers has some parallels with the idea of a knowledge worker, where the emphasis is on codified knowledge but this does not fully satisfy as it ignores the tacit dimension. The contribution of ICTs to more rapid growth and productivity gains is increasingly recognised albeit with a lagged effect rather than an instant gain [Carlaw et al., 2006]. Although the impact of ICTs differs across countries, studies on the information work force show a steady growth of such workers. A shift from low-skilled information workers (those who handle information in routine ways) to high-skilled information workers indicate that the economy becomes more knowledge-intensive (Engelbrecht, 2000).

The ICT sector

ICT producing industries

Appendix Box 1: Underlying principles

For manufacturing industries, the products of a candidate industry either must be intended to fulfill the function of information processing and communication including transmission and display, or must use electronic processing to detect, measure and/or record physical phenomena or to control a physical process. Components primarily intended for use in such products should also be included.

For services industries, the products of a candidate industry must be intended to enable the function of information processing and communication by electronic means. Furthermore, the service provided should go beyond simply the supply of goods. (OECD, 2003, 1998)

Appendix Box 2: The ICT-producing sector[1]
(ISIC Rev. 3 industries)

Manufacturing:
- 3000: Manufacture of office, accounting and computing machinery
- 3130: Manufacture of insulated wire cable
- 3210: Manufacture of electronic valves and tubes and other electronic components
- 3220: Manufacture of television and radio transmitters and apparatus for line telephony and line telegraphy
- 3230: Manufacture of television and radio receivers, sound or video recording or reproducing apparatus and associated goods
- 3312: Manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes except industrial process control equipment
- 3313: Manufacture of industrial process equipment

Services:
Goods related
- 5150[2]: Wholesale of machinery, equipment and supplies (where available)
- 7123: Renting of office machinery and equipment (including computers)

Intangible
- 6420[3]: Telecommunications
- 7200: Computer related activities

Source: OECD (2003, 1998).

Notes: 1. Definition of the ICT producing sector, approved at the 1998 WPIIS meeting, and based on the ISIC Rev. 3 activity classification.

2. Where available, countries should only include those sub-sectors of this industry that directly provide ICT wholesaling services. This will avoid the inclusion of extraneous wholesaling activities. For example, using the NACE nomenclature, only NACE categories 5143, 5164, and 5165 should be included.

3. In those circumstances where countries include telecommunication activities as part of radio and television activities (ISIC 9213), radio and television activities (9213) should be included in this definition. Otherwise, it should not be included.

There is currently no agreement internationally on the definition of ICT sector. The OECD, for example, has kept changing its definitions. In 1998, OECD countries reached agreement on an industry-based definition of the ICT sector based on International Standard Industry Classification Revision 3. The ICT sector included "manufacturing and services industries whose products capture, transmit or display data and information electronically" [Working Party on Indicators for the Information Society, 2007: 4]. In this definition, ICT is treated as an output of production. Thus, the ICT sector mainly refers to ICT-producing. In 2003, a new definition was constructed that extends its scope to include both ICT producing and using industries. The updated definition showed the increasing influence of ICTs in economic activities requiring the definition to be widened to include both outputs and inputs. However, such an extension risks creating measurement problems as ICT using industries are not defined explicitly, rather the focus is on the use of ICTS which are now found across several industries in both the manufacturing and service sectors. This may

well be a reflection of the diffusion of ICTs into all aspects of economic activity but leaves problematic the measurement of the significance of such workers within the organisation.

ICT-using industries

Appendix Box 3: ICT using industries[1]
(Identified on the basis of data on capital flows and estimates of capital stocks)

Manufacturing:
- Printing and publishing
- Electronic equipment
- Machinery and equipment

Services:
- Communications
- Whole sale and retail trade
- Finance
- Insurance
- Business services

Source: Pilat and Lee (2001).

Note: ICT using industries identified on the basis of capital flow matrices and capital stock estimates in Pilat and Lee (2001).

Figure 1: The OECD definition of the ICT sector in 2003. Source: *New perspectives on ICT skills and employment*. OECD. p.22&23.

A further version in 2006 seems to go back to the 1998 version but makes it more ambiguous in terms of the distinction between ICT producing and using industries. According to the general definition, "the production (goods and services) of a candidate industry must primarily be intended to fulfil or enable the function of information processing and communication by electronic means, including transmission and display" [Working Party on Indicators for the Information Society, 2007: 15]. In this definition, one apparent change is that the term 'products" is replaced by 'production'. It indicates that ICTs might be either tools as inputs or products as outputs or both in different industries. Meanwhile, this definition seems to show the intention of narrowing down the scope of the ICT sector. It regards the ICT sector as being 'primarily' intended to fulfil the function of information processing. This is an attempt to exclude industries where dealing with ICTs is not their primary purpose.

In general, the change of the definition reflects two concerns/difficulties in defining the ICT sector. First, the spillover effect of ICTs means that the definition of the ICT sector has to be updated to respond the changing impact of ICTs in the economy. Second, the effort of positioning the ICT sector in the existing industry classification means that the content of the ICT sector would be different due to the application of different industry classification system in different countries. For example, Statistics New Zealand uses the following definition of ICT industry in its Information and Communications Technology Supply Survey:

The information and Communications Technology (ICT) industry includes production and sales of ICT goods and services. ICT goods include: telecommunications equipment, audio visual equipment, electronic devices, computer and related equipment and software. ICT services include: telecommunications, programme distribution, internet access services, IT support, design, development and infrastructure services; and ICT rental, leasing and training services [Statistics New Zealand, 2007].

This definition shows that the ICT sector is defined more narrowly in the New Zealand context. It is mainly concerned with the production and distribution of ICT goods and services. Industries in the designated categories of the Australian and New Zealand Standard Industrial Classification (ANZSIC) are chosen as follows [Information Technology Policy Group, 2003].

ICT manufacturing industries	
2610	Manufacture of electronic components and boards
2620	Manufacture of computers and peripheral equipment
2630	Manufacture of communication equipment
2640	Manufacture of consumer electronics
2680	Manufacture of magnetic and optical media
ICT trade industries	
4651	Wholesale of computers, computer peripheral equipment and software
4652	Wholesale of electronic and telecommunications equipment and parts
ICT services industries	
5820	Software publishing
61	**Telecommunications**
6110	Wired telecommunications activities
6120	Wireless telecommunications activities
6130	Satellite telecommunications activities
6190	Other telecommunications activities
62	**Computer programming, consultancy and related activities**
6201	Computer programming activities
6202	Computer consultancy and computer facilities management activities
6209	Other information technology and computer service activities
631	**Data processing, hosting and related activities; web portals**
6311	Data processing, hosting and related activities
6312	Web portals
951	**Repair of computers and communication equipment**
9511	Repair of computers and peripheral equipment
9512	Repair of communication equipment

Note: The codes and titles above are those found on the UNSD web site (http://unstats.un.org/unsd/cr/registry/isic-4.asp) on 20 November 2006 plus editing corrections known at 25 January 2007. In case of further changes to ISIC, the final official titles will prevail. Differences in titles or codes from those presented earlier in the paper reflect changes from earlier drafts of ISIC Rev. 4.

Figure 2: The 2006 definition of the ICT sector. Source: Information economy, Sector definitions international standard industry classification (ISIC4).OECD. p.15.

ANZSIC Code	Industry Group
C284100	Computer and Business Machine Manufacturing
C284200	Telecommunication, Broadcasting and Transceiving, Equipment Manufacturing
C284900	Electronic Equipment Manufacturing
F461300	Computer Wholesaling
J712000	Telecommunication Services
L783100	Data Processing Services
L783200	Information Storage and Retrieval Services
L783300	Computer Maintenance Services
L783400	Computer Consultancy Services
P912100	Radio Services
P912200	Television Services

Figure 3: ICT Industry Groups in New Zealand. Source: Statistics on information technology in New Zealand (updated to 2003). Statistics New Zealand. p.14.

Differentiation found amongst information workers not only suggests a linkage between knowledge workers and information workers, but also shows the necessity of considering the skill components of information workforce. Lopez-Bassols [2002] categorises IT skills as three sets: professional IT skills, applied IT skills and basic IT skills or "IT literacy": professional IT skills include ability to use advanced IT tools and/or to develop, repair and create them. Applied IT skills are the ability to apply simple IT tools in general workplace settings (in non-IT jobs). Basic IT skills or "IT literacy" is the ability to use IT for basic

tasks and as a tool for learning. Based on these categories, issues and areas for government action for improving ICT skills are summarised in the OECD table:

Table 8. ICT skills: issues and actions

	Skills formation (which skills?)	Skills acquisition/renewal (when and where?)	Main issues	Areas for government action
Professional ICT skills	Skills required to develop, use or service ICTs professionally	Post-secondary education, IT vendor certification	- Balancing specialist ICT skills with other more generic skills - High mobility - Recognition of non-formal qualifications	- Improve attractiveness of ICT careers - Assist in providing labour market information - Examine options for using foreign labour
Applied ICT skills	Ability to use ICTs in non-ICT jobs	Post-compulsory education, workplace training	- Importance of integrating ICT into a sector/profession	- Help identify emerging ICT skill requirements for non-ICT jobs - Provide incentives for firms to train workers
Basic ICT skills	Strong life-long learning skills: fluency to use ICT for learning, working, recreation	Learning context: schools (children), training (adults) [at work, formal courses, informal exposure]	- Developing common standards - Building core ICT competencies in curricula - Enhancing teacher skills	- Promote ICT skills as important "generic" skills for life-long learning - Facilitate roll-out of ICTs in schools - Promote ICT skills among teachers

Source: OECD based on Multimedia Victoria (2000).

Table 4. Source: ICT skills and employment. OECD 2002. p. 25.

Meanwhile, Lopez-Bassols's analysis of the channels for acquiring skills show that formal education is not the only channel for acquiring ICT skills. According to him, skills are acquired through a variety of ways such as natural abilities, formal education, work experience, on-the-job and external training, and informal learning. Thus, he holds that IT work requires individuals to master codified and tacit knowledge. They need to not only understand technical abstract concepts, which are gained through various formal education but also acquire skills from non-formal ways such as work experiences and interpersonal communications. Surveys from countries such as the United States, Canada, Ireland and the United Kingdom, concur with the finding that IT firms tend to seek a combination of three main types of skills. Technical skills refer to primarily IT, quantitative analysis/data modelling digital media and technical writing. Business/management skills include marketing, strategy and business writing. Personal skills refer to communication, leadership, teamwork, problem-solving ability [Lopez-Bassols, 2002].

In this view, IT jobs as a new occupation, needs workers to have multiple skills including technical and creative skills, business abilities and personal skills to be flexible and creative. What this suggest is that there is a significant transformation from being an IT worker to being a knowledge worker and this is not well represented in the ways that these groups of workers are conventionally measured. The definition of high-skilled information workers as knowledge workers attempts to reflect this view. Pyöriä [2005] holds that the symbolic aspects of knowledge work emphasise the abilities of knowledge workers to act as an interface between new technology and human interaction and thus, both theoretical and interpersonal knowledge are needed. He defines knowledge workers as "wage earners whose jobs meet the following three criteria: the use of IT, at least upper intermediate vocational training and independent design of important aspects of the job" [Pyöriä et al., 2005: 89]. His definition emphasises the significance of both codified knowledge and practical intelligence (tacit knowledge) involved in the knowledge work process.

According to Stehr [1994], the social constitution of experts and expertise creates a social-intellectual division between 'lay' publics and experts. This suggests not that power is changing hands but that the nature, content or substance of exercising power is being modified. Knowledge-based occupations, he defines as "those who consult, provide guidance to others, counsel or give expert advice - as the group of occupations engaged in transmitting and applying knowledge" [Stehr, 1994: 184]. Although this definition seems to suggest that knowledge-based occupations serve as transmitters between knowledge producers and end-users, he explains that he does not mean the flow of knowledge is a passive process thus requires transfer agents of knowledge workers/managers to facilitate.

Here we can see the rationale for the growth of this new group within management education and practices.

3. KNOWLEDGE WORKERS AS HUMAN CAPITAL AND THE MEASUREMENT

To move from information workers to knowledge requires a retheorising of their respect roles within the "new economy" and the respective roles of theoretical and practical knowledge. A number of studies have emphasised that other forms of knowledge and skills, such as practical knowledge and communication skills, are important in knowledge work. These changes point to difficulties with indicators that identify knowledge workers in terms of existing occupational classifications. In these workers' specialisation and educational qualification are major criteria for determining their status in occupational hierarchies. In some occupational classification systems, managers, professionals and technicians are usually identified as major groups of knowledge workers. For example, the Australian Bureau of Statistics [2002] defines knowledge-based workers as "those employed as managers and administrators, and professionals and associate professionals including those in science and engineering, business and information, health and education". According to Statistics Canada, knowledge occupations fall into the three broad groups including professional, management and technical occupations [Baldwin and Beckstead, 2003]. Others have categorised more occupational groups as knowledge workers. For example, Davenport [2005:10] defines knowledge workers as those who "have high degrees of expertise, education, or experience, and the primary purpose of their jobs involves the creation, distribution, or application of knowledge". He places the following categories as knowledge workers: Management; Business and financial operations; Computer and mathematical; Architecture and engineer; Life, physical, and social scientists; Legal; Healthcare practitioners; Community and social services; Education, training, and library; Arts, design, entertainment, sports, media. These definitions and associated measures have two features. First, the level of knowledge and skills, mainly gained through formal education, is the basis for classifying knowledge workers. Second, knowledge workers are understood merely in an economic framework based on the human capital theory.

Treating human beings as capital allows us to explain the role of knowledge and skills individuals contribute to economic growth. In particular, the emergence and rapid spread of knowledge workers in knowledge economies/societies revitalises the explanatory power of human capital theory. For example, the OECD defines human capital as "the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well being" [Keeley, 2007: 29]. They hold that human capital embodied in people's skills, learning talents and attributes is important for individuals to earn a living and countries to widen economic growth. Such knowledge workers have become in high demand internationally and are subject to competition globally. This is then reflected in the way immigration polices have been restructured. However where these have failed to appreciate that "knowledge" transfer is complex and requires the acquisition of both formal and tacit knowledge and the social and cultural capital necessary to integrate into different national contexts they have often not achieved the desired effect and may result in migrants being recruited on the basis of "skills" that are in fact not easily transferable into the receiving economy and thus may not address the perceived or actual shortages [Wang and Thorns, 2008].

The above definition shows that human capital theory is still the dominant discourse with respect to the relationship between human capital, economic development and social wellbeing. Further endorsement of human capital theory can be found in the various frameworks for measuring the knowledge economy initiated by national and international organisations such as the Australian Bureau Statistics and the World Bank [Chen and Dahlman, 2005; Trewin, 2002]. In this case, human capital is always measured with educational inputs and outputs such as literacy rates, school enrolment rate and the number of graduate students. Thus, formal education and training is central to the policy agenda at all levels consistent with most definitions of knowledge workers, which emphasises formal individual educational attainments. Such a way of differentiating workers could create a

new set of inequalities based on a knowledge divide. To provided a more nuanced understanding recent UNESCO reports have emphasised that "learning societies will need to engage in a study of the different forms of knowledge, distinguishing descriptive knowledge, procedural knowledge, explanatory knowledge and behavioural knowledge" [UNESCO, 2005: 60]. Thus, ways of learning and forms of knowledge are closely linked with each other. Knowledge workers should not be seen as only possessing knowledge and skills gained solely through formal ways of learning. As Bell [1973] notes, work in post-industrial societies is a "game between persons" and thus, persons have to learn to live with each other. Taking this view, we need to recognise the role of social capital in economic production.

Social capital is not separate from other forms of capital. Bourdieu's [2001] analysis of capital highlights the fact that different forms of capital exist. Social capital as a durable network of mutual acquaintance and recognition reflects how effectively a given agent can mobilise other types of capital possessed by him/her and others. However, he notes that:

Economic capital is at the root of all the other types of capital and that these transformed, disguised forms of economic capital, never entirely reducible to that definition, produce their most specific effects only to the extent that they conceal the fact that economic is at their root… [Bourdieu, 2001: 106]

For knowledge workers their contribution to economic growth lies not only in their human capital embodied in their formal knowledge but also in the social capital they possess. Their work involves more person-to-person relationships than labour-machine relationship in industrial societies. Both Stehr's view of knowledge workers as mediators and Pyöriä's emphasis on the significance of teamwork indicate that social capital and tacit knowledge are crucial variables for measuring the contribution of knowledge work. Countries and organisations have put considerable effort into identifying and measuring social capital. For example, Statistics New Zealand developed a framework of measuring social capital in 1997. It defines social capital as "relationships among actors (individuals, groups, and/or organisations) that create a capacity to act for mutual benefit or a common purpose" [Spellerberg and Statistics New Zealand, 2000: 9]. The framework focuses on four themes: behaviours (what people do), attitudes and values (what people feel), population groups (what people are) and organisations. In the UK the Office for National Statistics identifies five constructs of social capital: views about the local area, civic engagement, reciprocity and trust, social networks and social support. The Australian Bureau of Statistics (ABS) suggests the following data items for the measurement of social capital: social networks and support structures, social and community participation, civic and political involvement and empowerment, trust in people and social institutions, tolerance of diversity and altruism, philanthropy and voluntary work [Statistics New Zealand, 2001]. These frameworks for measuring social capital have been criticised as not being comprehensive as each just grasps some characteristics of social capital. Despite this, it is necessary to bring the concept of social capital into consideration when defining knowledge workers in order to fully understand their role in knowledge economies.

4. CONCLUSIONS

The difficulties of identifying knowledge workers in existing occupational classifications have now been explored. The final question is whether we need more radical changes to the way we define and measure knowledge workers within the "new economy" that is emerging if we are to effectively document the nature of the transformation currently taking place in many economies. To fully explore the transnational movement of skills and the policy platforms required to build a workforce to create a knowledge based economy we need improved understanding of the role of different types of knowledge and different forms of knowledge work.

ACKNOWLEDGEMENTS

We wish to acknowledge support from the NZ Marsden Fund and NZ Foundation for Research, Science and Technology which aided completion of this research.

REFERENCES

Australian Bureau of Statistics, *1370.0-Measuring Australia's Progress, 2002*.

Baldwin, J.R. and D. Beckstead, *Knowledge Workers in Canada's Economy* 1971-2001, Statistics Canada, Ottawa, 2002.

Bell, D., *The Coming of Post-industrial Society: A Venture in Social Forecasting*, Basic Books, New York, 1973.

Bourdieu, P., The forms of capital. In M. Granovetter and R. Swedberg (Eds.), *The Sociology of Economic Life*, 96-111, Westview Press, Colorado and Oxford, 2001.

Carlaw, K., L. Oxley, P. Walker, D. Thorns and M. Nuth, Beyond the hype: Intellectual property and the knowledge society/knowledge economy, *Journal of Economic Surveys*, 20(4) September, 633-690, 2006.

Castells, M., *The Rise of the Network Society*, Blackwell Publishers, Cambridge, 1996.

Chen, D. H. C., and C. J. Dahlman, *The Knowledge Economy, the KAM methodology and World Bank Operations*, The World Bank, Washington, 2005.

Davenport, T.H., *Thinking for Living: How to Get Better Performance and Results from Knowledge Worker*, Harvard Business School, Press Boston, MA, 2005.

Drucker, P. F., *The Age of Discontinuity: Guidelines to Our Changing Society*, Heinemann, London, 1969.

Engelbrecht, H.-J., Towards a knowledge economy? Changes in New Zealand's Information Work Force 1976-1996, *Prometheus, 18*(3), 265-282, 2000.

Keeley, B., *Human capital: How What You Know Shapes Your Life*, OECD, Paris, 2007.

Information Technology Policy Group, *Statistics on Information Technology in New Zealand [Updated to 2003]*, Ministry of Economic Development, Wellington, 2003.

Lopez-Bassols, V., *ICT Skills and Employment*, OECD, Paris, 2002.

Machlup, F., *The Production and Distribution of Knowledge in the United States*, Princeton University Press, Princeton, N.J, 1962.

Pyöriä, P., The concept of knowledge work revisited**,** *Journal of Knowledge Management*, 9 (3), 116-127, 2005.

Pyöriä, P., H. Melin., and R. Blom, *Knowledge Workers in the Information Society*, Tampere, Finland: Tampere University Press, Tampere, Finland, 2005.

Reich, R. B., and R. B. Reich, *The Work of Nations: Preparing Ourselves for the 21st Ccentury Capitalism*, Alfred A Knopf, New York, 1991.

Spellerberg, A., and Statistics New Zealand, *Framework for the Measurement of Social Capital in New Zealand*, Statistics New Zealand, 2001.

Statistics New Zealand, *Information and Communications Technology Supply Survey*. Statistics New Zealand, Wellington, 2007.

Stehr, N., *Knowledge Societies*, Sage, London, 1994.

Thrift, N., *Knowing Capitalism*, Sage, London, 2005.

Trewin, D., *Measuring a Knowledge-Based Economy and Society: An Australian Framework*, The Australian Bureau of Statistics, Canberra, 2002.

Working Party on Indicators for the Information Society, *Information Economy—Sector Definitions Based on the International Standard Industry Classification (ISIC4)*, OECD, Paris, 2007.

UNESCO, *UNESCO World Report: Towards Knowledge Societies*, UNSCO Publishing, Paris, 2005.

Wang, H., and D. C. Thorns, Different Forms of Knowledge and New Chinese Skilled Immigrants' Adaptation to New Zealand's Knowledge Society ISSJ 2008.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Information and Communication Technologies and the Income Distribution: A General Equilibrium Simulation

T. Fukiharu

Faculty of Economics, Hiroshima University, Japan (fukito@hiroshima-u.ac.jp)

Abstract: The special feature of the innovation in the information and communication technologies (ICTs) implies the improvement in the quality of consumption of goods and services including leisure hours, as well as the modification in the production process. The traditional analysis has emphasized solely the modification in the production process. This paper incorporates this special feature in analysing the relation between the innovation and the income distribution. It has been argued that the innovation of ICTs expands the unfairness of income distribution. In the traditional expression, this may be rephrased that the wealthy capitalist class becomes relatively better off, while the poor working class becomes relatively worse off. Assuming three social classes: the entrepreneurs, the capitalists, and the workers, this paper examines how the relative shares of these classes in the national income changes due to the innovation of ICTs. For three types of production functions: Cobb-Douglas type, CES type with negative substitution, and CES type with positive substitution, this paper conducts the comparative statics analysis by actually computing general equilibrium prices for the specified different parameters of ICTs in the production function. This paper derives the completely opposite conclusions between the CES type with negative substitution, and the one with positive substitution. Thus, it is shown by these simulations that theoretically, no definite relation holds between the innovation of ICTs and income distribution (or utility level). In order to derive the definite relation, one must examine what type of production function the economy has.
Keywords: ICT, general equilibrium, income distribution, simulation.

1. INTRODUCTION

Since the invention of computer in the 1930s, the information processing technology was mainly innovated by IBM until 1970s. AT&T was the main innovator of communication technology until 1970s. After the anti-trust lawsuits against these giant corporations, the barrier between communication and information industries, set by the US administration of justice (antitrust division), was removed in 1983 by the administration. In the 1990s, the US enjoyed the historic prosperity stemming from the fusion of information and communication technologies (ICTs). The innovation of ICTs has benefited not only the US but also the whole world. Furthermore, it has benefited not only the production and research facilities but also the general public by enriching their daily life. It has been argued, however, that the innovation of ICTs has expanded the unfairness of income distribution. In the traditional expression, this may be rephrased that the wealthy capitalist's class becomes relatively better off, while the poor working class becomes relatively worse off. The aim of this paper is to examine whether this unfairness emerges in the purely theoretical model. While in the traditional economics, the innovation was characterized by the modification of the production process: *e.g.* the shifts in the production functions, the special feature of the innovation in the ICTs implies the improvement in the quality of consumption of goods and services including leisure hours, as well as the modification of

the production process. As an example of the analysis on the traditional innovation, we may refer to Fukiharu [2007a], which examined the relationship between the Green Revolution (GR), one of the innovations, and the profit. It is not certain whether the farmers' profit rises due to the GR, since it raises supply of grain, while reducing production cost. In Fukiharu [2007a] it was shown that when the production function is of CES (Constant Elasticity of Substitution) Type, the farmers' profit might fall with the assumption of positive substitution. This paper adopts the simulation approach to examine if the unfairness emerges under the assumption of different production functions.
In this paper, assuming three social classes; the entrepreneurs, the capitalists, and the workers, how the relative shares of these classes in the national income changes due to the innovation of ICTs. Following the Classicals' framework, the capitalists have capital goods and do not work, while the workers have no capital goods and earn income by providing his initial endowment of leisure for the other members. If the production function is assumed to be under constant returns to scale, the payment of rent for the capital goods and the one of the wage for the labor supply occupies the whole revenue of products. In this paper, however, the production function is assumed to be under decreasing returns to scale, following Fukiharu [2007a]. This assumption guarantees the positive profit: surplus. Thus, in this paper, the entrepreneur class exists. By computing the General Equilibrium (GE) prices, the comparative statics analysis is conducted. First, examining three cases on the assumption on the production function, how the shares of three social classes change due to the innovation of the ICTs is analyzed. Second, the utility change of three social classes due to the innovation of the ICTs is analyzed. We start with the Cobb-Douglas function case.

2. COBB-DOUGLAS PRODUCTION FUNCTION CASE I

In this paper, it is assumed that there is one (representative) firm, producing consumption good, q, utilizing the capital goods, K, and the labor input, L. In this section, the production function, $f[K, L]$, is assumed as in what follows:

$$q=f[K, L]=(c_K K)^{a1}(c_L L)^{a2} \quad a_1+a_2<1 \qquad (1\text{-}1)$$

where c_K is the level of ICTs with respect to the capital input in the production process, c_L is the level of ICTs with respect to the labour input. The improvement in this section is called the *additive type*. The production is conducted under decreasing returns to scale, so that the positive profit is guaranteed. The production function, defined in (1-1) is called Cobb-Douglas type.

2.1 The Behaviour of the Firm

The (representative) firm maximizes profit, π:

$$\max \pi=pq-rK-wL$$

where p is the price of the commodity, r is the rental price of capital goods, and w is the wage rate. By the profit maximization of the firm, the capital demand function, K_d, the labour demand function, L_d, the supply function of the consumption good, q_s, and the profit, π_0, are computed.

2.2 The Behaviour of the Workers

It is assumed that there is the (aggregate) household, who has the initial endowment of leisure hours, L_0=100, with no profit distribution from the firm. This class aims at utility maximization subject to income constraint:

$$\max u_L[x, l]=x^{c1b1} l^{c2b2} \text{ s.t. } px=w(L_0-l) \qquad (2)$$

where $u_L[x, l]$ is the worker's utility function, x is the quantity of consumption good, and l is the leisure consumption. Note that the innovation of the ICTs causes the modification of

the utility function. The level of ITCs with respect to the consumption good is expressed by c_1, while the one with respect to leisure time is expressed by c_2. From the assumption of utility function as Cobb-Douglas type in (2) the demand function of consumption good, x_{dW}, and the labour supply function, L_s, are computed.

2.3 The Consumption Behaviour of Entrepreneurs

It is assumed that the (representative) entrepreneur, who receives the whole profit, owns the firm. This class aims at the utility maximization subject to the income receipt, π_0, consuming the commodity, x, and hiring a part of the workers, L, for this class, without providing itself as workers inside and outside the production process.

$$\max u_E[x, L]= x^{c1b1} L^{c2b2} \quad \text{s.t } px+wL=\pi_0 \qquad (3).$$

where $u_E[x,L]$ is the entrepreneur's utility function, x is the quantity of consumption good, and L is the consumption of workers' leisure hours. Note that the innovation of the ICTs causes the modification of the utility function in exactly the same way as in the case of household. The level of ICTs with respect to the consumption good is expressed by c_1, while the one with respect to leisure time is expressed by c_2. From the assumption of utility function as Cobb-Douglas type in (3), the entrepreneur's demand function of consumption good, x_{dE}, and the entrepreneur's demand function for labour, L_{dE}, are computed.

2.4 The Consumption Behaviour of Capitalists

It is assumed that there is the (aggregate) capitalist, who has the initial endowment of capital good, K_0=100, with no profit distribution from the firm, consuming the consumption good, x, and hiring a part of the workers, L, for this class without providing itself as workers inside and outside the production process. This class aims at the utility maximization subject to income constraint:

$$\max u_K[x, L]= x^{c1b1} L^{c2b2} \quad \text{s.t.} px+w\text{L}=rK_0 \qquad (4)$$

where $u_K[x,L]$ is the capitalist's utility function. From the assumption of utility function as Cobb-Douglas type in (4), the capitalist's demand function of consumption good, x_{dK}, and the capitalist's demand function for labour, L_{dK}, are computed.

2.5 General Equilibrium

There are three markets in this model: commodity market, capital market, and the labour market. The general equilibrium analysis computes equilibrium prices, p^*, r^*, and w^*, which simultaneously equates demand supply for the three markets. Equilibrium conditions for the three markets are stipulated as in what follows.

$$x_{dW}+x_{dE}+x_{dK}=q_s \quad \text{(commodity market)} \qquad (5)$$
$$K_d=K_0 \quad \text{(capital market)} \qquad (6)$$
$$L_d+L_{dE}+L_{dK}=L_s \quad \text{(labour market)} \qquad (7)$$

Assuming w^*=1, we can compute other equilibrium prices:

$$p^*=100^{1-a1-a2}a_2^{-a2}(a_2b_1+b_2c_2/c_1)^{a2-1}c_K^{-a1}c_L^{-a2}$$
$$r^*=a_1b_1/(a_2b_1+b_2c_2/c_1) \qquad (8\text{-}1)$$

As the improvement of ICTs emerges in production part, commodity price, p^*, declines, with no effect on the rental price, r^*, as shown in (8-1). As the improvement of ICTs emerges in daily life part, p^* and r^* declines if c_2/c_1 increases, as shown in (8-1).

2.6 Income Distribution and Real Income (Utility) Changes

These improvements of ICTs cannot influence the income distribution, since we have

$$w^*Ls^*/p^*qs^*= a_2, \quad (9\text{-}1)$$
$$\pi_0^*/ p^*qs^*= 1-a_1-a_2, \quad (10\text{-}1)$$
$$r^*K_0/p^*qs^*=a_1. \quad (11\text{-}1)$$

The collection of (9-1)~(11-1) is a classical one. The income distribution change is one way of evaluating the economic situation of each class. In this subsection, we examine how the real incomes for the three classes change when the improvements of ICTs emerge. The utility change for each class may be another way of evaluating the economic situation of each class. In order to do this, suppose that

$$a_1=a_2=b_1=b_2=1/3 \quad (12\text{-}1)$$

By the improvement of ITCs in the production part: c_K and c_L, raises the utility levels of all the class. For example, when $c_1=c_2=1$, the utility level of the workers' class increases from 7.67671 to

$$7.67671 c_K^{1/9} c_L^{1/9} \quad (13\text{-}1)$$

as c_K and c_L change from 1. When $c_1=c_2=1$, the utility level of the entrepreneur's class increases from 3.0465 to

$$3.0465\ c_K^{1/9} c_L^{1/9} \quad (13\text{-}2)$$

as c_K and c_L change from 1. When $c_1=c_2=1$, the utility level of the capitalist' class increases from 3.0465 to

$$3.0465\ c_K^{1/9} c_L^{1/9} \quad (13\text{-}3)$$

as c_K and c_L change from 1. From the simulation we can show that by the improvement of ICTs in the daily life part: c_1 and c_2, raises the utility levels of all the class.
A remark is in order. As shown in (8), p^* declines if c_2/c_1 increases. We can show that the profit also declines in this case, since $\pi_0^*=100c_1/(c_1+3c_2)\}$ holds when $c_K=c_L=1$.

3. COBB-DOUGLAS PRODUCTION FUNCTION CASE II

In this section, the production function, $f[K, L]$, is assumed as in what follows:

$$q=f[K, L]=K^{cka1} L^{cla2} \quad c_k a_1+ c_l a_2<1 \quad (1\text{-}2)$$

where c_k is the level of ICTs with respect to the capital input in the production process, c_l is the level of ICTs with respect to the labour input. The improvement in this section is called the *structural type*. The production is conducted under decreasing returns to scale, so that the positive profit is guaranteed. These improvements of ICTs can influence the income distribution, since we have

$$w^*Ls^*/p^*qs^*= c_l a_2, \quad (9\text{-}2)$$
$$\pi_0^*/ p^*qs^*= 1- c_k a_1- c_l a_2, \quad (10\text{-}2)$$
$$r^*K_0/p^*qs^*= c_k a_1. \quad (11\text{-}2)$$

When the innovation of ICTs improves the efficiency in production, it raises the shares in national income, as is clear from (9-2) and (11-2). This, however, reduces the share of the profit in national income. In this situation, we examine the utility change, starting from the case of workers class. In addition to (12-1), it is assumed

$$c_1=c_2=c_k=c_l=1. \quad (12\text{-}2)$$

The utility level of the workers class, u_{L0}, and the one of capitalists class, u_{K0}, are 7.67671 and 3.0465, respectively at GE when (12-1) and (12-2) are assumed. If c_l rises to 5/4, then the share of the workers rises as shown in (9-2). In this case, not only the utility level of the workers' class but also the one of the capitalists class rise to u_{L1}=8.76165 and u_{K1}=3.33934, respectively. The results for each improvement of ICTs are summarized in Table 1.
As shown in Table 1 workers and capitalists become better off by any improvement of ICTs. If c_l rises to 5/4, then the share of the entrepreneurs falls as shown in (10-2). The utility level of this class, u_{E0}, is 3.0465 at GE when (12-1) and (12-2) are assumed. If c_l rises to 5/4, then the utility level of this class, u_{E1} falls to u_{E1}= 2.75656. This is due to the declined profit from π_{00}=25 to π_{01}=17.6471. The effects for each improvement of ICTs on the utility level and profit are summarized in Table 2. Note that as shown in Table 2, when c_2 rises to 5/4, then the utility level of the entrepreneur's class rises in spite of the declined profit.

	u_{L0}= 7.67671	u_{K0}= 3.0465
c_1=5/4	u_{L1}= 8.85721	u_{K1}= 3.53743
c_2=5/4	u_{L1}= 11.1008	u_{K1}= 3.45013
c_k=5/4	u_{L1}= 8.7243	u_{K1}= 4.01758
c_l=5/4	u_{L1}= 8.76165	u_{K1}= 3.33934

Table 1.Utility Changes for Workers and Capitalists: Cobb-Douglas Case

	u_{E0}=3.0465	π_{00}= 25
c_1=5/4	u_{E1}= 3.53743	π_{01}= 29.4118
c_2=5/4	u_{E1}= 3.45013	π_{01}= 21.0526
c_k=5/4	u_{E1}= 2.85802	π_{01}= 18.75
c_l=5/4	u_{E1}= 2.75656	π_{01}= 17.6471

Table 2. Utility and Profit Changes for Entrepreneurs: Cobb-Douglas Case

4. CONSTANT ELASTICITY OF SUBSTITUTION (CES) PRODUCTION FUNCTION CASE I: NEGATIVE PARAMETER

In this section, the CES type production function is assumed as in what follows:

$$q=f[K, L]=((c_K K)^{-t}+(c_L L)^{-t})^{-n/t} \quad n=1/2,\ t=-1/2 \qquad (1\text{-}3)$$

where c_K is the level of ICTs with respect to the capital input in the production process, and c_L is the level of ICTs with respect to the labour input. It is assumed that the production is conducted under decreasing returns to scale, so that the positive profit is guaranteed, as in the previous sections.
With the sole modification of production function, the general equilibrium prices, p^*, r^*, and w^*, are computed by solving (5), (6), (7). It is not easy to compute these prices with such a plain expression as in (8-1) when no specification of parameters is made. Thus, suppose that

$$c_1=c_2=c_K=c_L=1. \qquad (12\text{-}3)$$

When (12-3) is satisfied, we can compute equilibrium prices:

$$p^*=20/3,\ r^*=1/3. \qquad (8\text{-}2)$$

It is easy to ascertain that this equilibrium is a stable one, by computing the eigenvalues of the Jacobian matrix. The *negative* eigenvalues are derived as {–601.504, –1.49625}. The Walrasian tatonnement process is defined by the following differential equations.

$$dp[\tau]/d\tau= x_{dW}[\tau]+x_{dE}[\tau]+x_{dK}[\tau]-q_s[\tau]$$
$$dr[\tau]/d\tau=K_d[\tau]-K_0$$

where τ stands for time. The price trajectories on the Walrasian tatonnement process, starting from $p[0]=1$ and $r[0]=1$, are provided by Figure 1 and Figure 2.

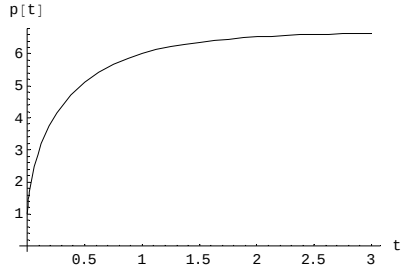


Figure 1. Price Trajectory of $p[\tau]$

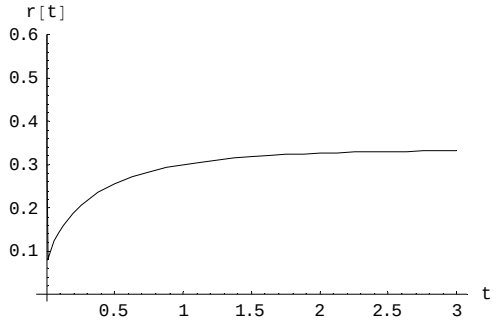


Figure 2: Price Trajectory of $r[\tau]$

4.1 Income Distribution

With no specification such as in (12-3), we can compute the income distributions, w^*Ls^*/p^*qs^*, π_0^*/p^*qs^*, and r^*K_0/p^*qs^*, as in what follows.

$$w^*Ls^*/p^*qs^*=[-c_2c_K+\{c_K(c_2^2c_K+c_1^2c_L+2c_1c_2c_L)\}^{1/2}]/ 2[c_1c_K+c_2c_K+\{c_K(c_2^2c_K+c_1^2c_L+2c_1c_2c_L)\}^{1/2}], \quad (9\text{-}3)$$

$$\pi_0^*/p^*qs^*=1/2, \quad (10\text{-}3)$$

$$r^*K_0/p^*qs^*=(c_1+2c_2)c_K/2[c_1c_K+c_2c_K+\{c_K(c_2^2c_K+c_1^2c_L+2c_1c_2c_L)\}^{1/2}]. \quad (11\text{-}3)$$

First, from these computations, the following holds.

w^*Ls^*/p^*qs^*rises as c_L rises, (9-3A)
r^*K_0/p^*qs^* falls as c_L rises. (11-3A)
w^*Ls^*/p^*qs^* falls as c_K rises, (9-3B)
r^*K_0/p^*qs^* rises as c_K rises. (11-3B)

Note that (9-3A) corresponds with (9-2), while (11-3B) corresponds with (11-2).
Next, with respect to the effect of c_1 and c_2 on the income distribution, we must assume that

$$c_K=c_L=1. \quad (12\text{-}4)$$

When (12-4) is satisfied, the following holds.

w^*Ls^*/p^*qs^* rises as c_1 rises, (13-1)
r^*K_0/p^*qs^* falls as c_1 rises. (14-1)
w^*Ls^*/p^*qs^* falls as c_2 rises, (15-1)
r^*K_0/p^*qs^* rises as c_2 rises. (16-1)

4.2 Real Income (Utility) Change

Following the previous analysis, the simulation analysis of utility change through the innovation of ICT is conducted. Starting from (12-3), the utility of the workers' class is computed as u_{L0}=7.21125. When c_1 rises to 5/4, the utility rises to u_{L1}= 8.07332. The utility changes for the workers' and capitalists' classes through the innovation of ICTs are summarized in Table3.

	u_{L0}= 7.21125	u_{K0}= 3.46681
c_1=5/4	u_{L1}= 8.07332	u_{K1}= 3.94296
c_2=5/4	u_{L1}=10.5663	u_{K1}= 4.12924
c_K=5/4	u_{L1}=7.35236	u_{K1}=3.66283
c_L=5/4	u_{L1}=7.35226	u_{K1}= 3.40032

Table 3. Utility Changes for Workers and Capitalists: CES Case I

	u_{E0}=4.19974	π_{00}=44.4444
c_1=5/4	u_{E1}=5.03291	π_{01}=53.2544
c_2=5/4	u_{E1}=4.98573	π_{01}=36.7347
c_K=5/4	u_{E1}= 4.321	π_{01}=45.0543
c_L=5/4	u_{E1}=4.24148	π_{01}= 43.8173

Table 4. Utility and Profit Changes for Entrepreneurs: CES Case I

Note that as c_K rises r^*K_0/p^*qs^* falls, as shown in (11-3A), while the utility rises in this

parameter change, as shown in Table3. Furthermore, as c_L rises, r^*K_0/p^*qs^* falls as shown in (9-3B), while the utility falls in this parameter change. Thus, symmetry is not guaranteed in this result.
The utility and profit changes for the entrepreneurs are summarized in Table 4. As shown in (10-3) the profit share of entrepreneurs' class is 1/2. The profit itself, however, falls when c_2 or c_L rises as shown in Table 4. The consequent price decline to maintain the profit share at 1/2 raises the utility level in these parameter changes as shown in Table 4. Thus, for all the parameter changes the utility of the entrepreneurs' class rises, as shown in Table 4.

5. CONSTANT ELASTICITY OF SUBSTITUTION (CES) PRODUCTION FUNCTION CASE II: POSITIVE PARAMETER

The aim of this section is to examine whether (9-3A), (11-3A), (9-3B), and (11-3B) hold when the parameter changes from t=–1/2 to t=1/2. When t=1/2, however, the computation of GE is not easy. Thus, the Newton method is utilized in the computation of general equilibrium, and the comparison is made between the case for (12-3) and the cases in which (a) only c_1 rises, (b) only c_2 rises, (c) only c_K rises, (d) only c_L rises from (12-3). We start from the computation of general equilibrium.
When (12-3) is satisfied, we can compute equilibrium prices by the Newton method:

p^*=22.5, r^*=0.125. (8-3)

It is easy to ascertain that this equilibrium is a stable one, by computing the eigenvalues of the Jacobian matrix. The *negative* eigenvalues are derived as {–888.978, –0.13332}. We cannot conduct a graphical simulation of this stability due to the slow convergence.

5.1 Income Distribution

First, from the simulation in terms of the Newton method, the following holds.

w^*Ls^*/p^*qs^* falls from 0.333333 to 0.322185 as c_L rises, (9-4A)
r^*K_0/p^*qs^* rises from 0.166667 to 0.177815 as c_L rises. (11-4A)
w^*Ls^*/p^*qs^* rises from 0.333333 to 0.344208 as c_K rises, (9-4B)
r^*K_0/p^*qs^* falls from 0.166667 to 0.155792 as c_K rises. (11-4B)

Note that (9-4A)~(11-4B) are opposite to (9-3A) ~ (11-3B).
Next, with respect to the effect of c_1 and c_2 on the income distribution, the following holds.

w^*Ls^*/p^*qs^* falls from 0.333333 to 0.325184 as c_1 rises, (13-2)
r^*K_0/p^*qs^* rises from 0.166667 to 0.174816 as c_1 rises, (14-2)
w^*Ls^*/p^*qs^* rises from 0.333333 to 0.341688 as c_2 rises, (15-2)
r^*K_0/p^*qs^* falls from 0.166667 to 0.158312 as c_2 rises. (16-2)

Note that (13-2)~(16-2) are completely opposite to (13-1) ~ (16-1).

	u_{L0}=4.8075	u_{K0}=1.20187
c_1=5/4	u_{L1}=4.90633	u_{K1}= 1.21416
c_2=5/4	u_{L1}=6.97766	u_{K1}=1.23581
c_K=5/4	u_{L1}=4.89191	u_{K1}=1.16287
c_l=5/4	u_{L1}=4.89865	u_{K1}=1.28586

Table 5. Utility Changes for Workers and Capitalists: CES Case II

	u_{E0}=2.5	π_{00}=37.5
c_1=5/4	u_{E1}=2.67034	π_{01}=53.2544
c_2=5/4	u_{E1}=2.92782	π_{01}=31.4132
c_K=5/4	u_{E1}= 2.53016	π_{01}=37.1966
c_L=5/4	u_{E1}=2.5617	π_{01}= 37.8162

Table 6. Utility and Profit Changes for Entrepreneurs: CES Case II

5.2 Real Income (Utility) Change

The utility changes for the workers' and capitalists' classes through the innovation of ICTs are summarized in Table5. The tendency of rise and fall in Table 3 and Table 5 is not the

same.
The utility and profit changes for the entrepreneurs are summarized in Table 6. The tendency of rise and fall in Table 4 and Table 6 is not the same.

6. CONCLUSIONS

This paper examined the effects of four elements in innovation on the income distribution. Traditional elements in the innovation argument were restricted on the production process. In this paper, the production function has two factors of production: labour input and capital input. Thus, in this paper, the 1st element is the efficiency improvement of the labour input through ICTs, and the 2nd element the one of capital input. The special feature of ICTs consists in the fact that quality of life improves through ICTs. Thus, in this paper, since the utility function has two variables: consumption good and leisure hours, the 3rd element is the improvement of the quality of consumption good through ICTs, and the 4th is the one of leisure hours through ICTs.
In Section 1, the production function is specified by Cobb-Douglas type where the efficiency improvement implies the *additive type* as in the traditional argument: *i.e.* one unit of input becomes c unit (c>1) after the innovation. Under this assumption, the innovation of ICTs cannot change the income distribution at all. When measured in terms of utility change, however, every element contributes to the enhancement of utility for all the social classes. It was shown that the profit declines due to the 4th element.
In Section 2, the production function is specified by Cobb-Douglas type where the efficiency improvement implies the *structural type* in the sense that the improvement implies the increased parameter of the Cobb-Douglas production function. It was shown that the 3rd and 4th elements have no effect on the income distribution structure. The 1st and 2nd elements can influence the income distribution structure. When measured in terms of utility change, the workers' class and the capitalists' class become better of due to any element. It was shown that the utility of entrepreneurs' class declines due to the first or second element. In this case, the profit was also shown to decline.
In Section 3, the production function was specified by CES type with negative substitution parameter. It was shown that all the four elements influence the income distribution structure, while the income share of the entrepreneurs is constant. For example, first, the 1st element contributes to the rise of workers' income share, while the 2nd element contributes to the rise of capitalists' income share. Second, the 3rd element contributes to the rise of workers' income share, while the 4th element contributes to the rise of capitalists' income share. As for the utility change, except for the 3rd element, all the social classes become better off. The 3rd element reduces the utility of capitalists' class. It was shown that the profit falls due to the 1st or 3rd element.
In Section 4, the production function was specified by CES type with positive substitution parameter. It was shown that all the four elements influence the income distribution structure, while the income share of the entrepreneurs is constant. The results, however, completely contradict those in the previous section. It was shown that the profit falls due to the 2nd or 4th element.
The local stability was guaranteed for CES type production functions. Thus, it was shown by these simulations, that theoretically, no definite relation holds between the innovation of ICTs and income distribution (or utility level). In order to derive the definite relation, one must examine what type of production function the economy has.

REFERENCES

Fukiharu, T, The Green Revolution and the Water Crisis in India: an Economic Analysis, *Farming Systems Design 2007, Int. Symposium on Integrated Analysis on Farm Production Systems*, M. Donatelli, J. Hatfield, A. Rizzoli, Eds. Catania (Italy), 10-12, September 2007, Book1, pp.153-4, 2007a.
Fukiharu, T, Information and Communication Technologies and the Income Distribution: A General Equilibrium Simulation, *http://home.hiroshima-u.ac.jp/fukito/index.htm*, 2007b.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Simulation study on spatial and temporal changes of factories in an industry district

Hiroyasu Nagata*, Kei-ichi Tainaka and Jin Yoshimura
Department of Systems Engineering, Shizuoka University, Hamamatsu 432-8561, Japan
**Corresponding author (f5645024@ipc.shizuoka.ac.jp)*

Abstract: Factories or plants are built or dismantled for various complex reasons. Understanding the construction and dismantlement processes of factories is one of complicated economic problems. The present article has two purposes: one is to explore the spatial and temporal changes of factories at Hamamatsu city in Japan, and the other is to represent the spatial and temporal changes by a simple simulation model. The time dependence of factory number can be classified into two periods. Before the collapse of Japanese stock market (1990), the number of factories increases smoothly. On the contrary, after the collapse, it rapidly decreases. The spatial structures of factories are statistically analyzed by auto-correlation. It is found that the factories more or less form aggregated or crowded structures. When the industrialization is not yet fully developed, the aggregation is very strong. In contrast, just before the collapse (in "bubble period"), the factories nearly form a random distribution rather than aggregation. After the collapse, they form slightly aggregated pattern. Such a profile is similar to that of biological population dynamics. We illustrate that the construction and dismantlement processes of factories can be qualitatively explained by the contact process which is a simple ecological model. Moreover, we discuss the relation between the distribution of factories and the Just-in-Time system in Hamamatsu industry.

***Keywords**:* spatial and temporal changes of factories; degree of factory aggregation; contact process; Just-in-Time system.

1. INTRODUCTION

Factories are built by various social reasons, such as economic growth and increase of consumer's wealth (Weber, 1929; Harris, 1954; Ullman, 1956). Geographical location of factory is also determined by many factors; examples are transportation cost, situations of consumers and laborers, supplying power, and some political or historical reasons. Therefore, it is difficult to explore the causal mechanism of factory distributions. Instead of reasoning, we explore the phenomenal features of factory distribution with the use of spatial simulation models. The present article has two purposes: 1) we report spatial and temporal changes of factories in Hamamatsu, and 2) we explain such a process by a simple lattice model.

Hamamatsu is one of major industrial cities in Japan (Shizuoka Tiiki-gakkai, 1996; Hamamatsu Kikakushitsu, 1954), and located near the middle of Japan (Fig. 1). There are three major companies; Honda and Suzuki motor companies and Yamaha Co. They are very famous as motor companies in the world. In Hamamatsu, new factory tends to be built near existing factories. This is convenient for cutting the costs in transportation and various managements. In particular, Japanese motor makers tend to form aggregation, since major companies introduced the so-called "Just-in-Time (JIT)" system which is first introduced by Toyota Corporation (Sheard, 1983; Estall, 1985). A parent company orders the parts of products (cars or motor cycle) to many subcontractors. According to various demands, the parent company adjusts the production. Many parts are collected and put together, only when they are needed. Hence, parent companies have very few warehouses. By the aggregation of parts subcontractors; physical distribution cost can be reduced. Recently, JIT

system becomes more flexible, so that the aggregation of factories seems somewhat reduced (Nojiri, 2005; Mair, 1992).

Our research area is a rectangle of 18km 24km as displayed in Fig. 1 (b). Real data on the spatial temporal changes are displayed in Fig. 2 (Hamamatsu Chamber of Commerce and Industry, 2006; Hamamatsu Kikakusitsu, 2006). Here each plot represents the center location of each factory. From Fig. 2, we notice the crucial effect of the collapse of Japanese stock market. Near 1990 the stock market crashed, and the values of real estates declined sharply. The time dependence of factory number can be classified into two periods. Before 1990, the number of factory increased smoothly. On the contrary, after the collapse of stock market, it decreased rapidly.

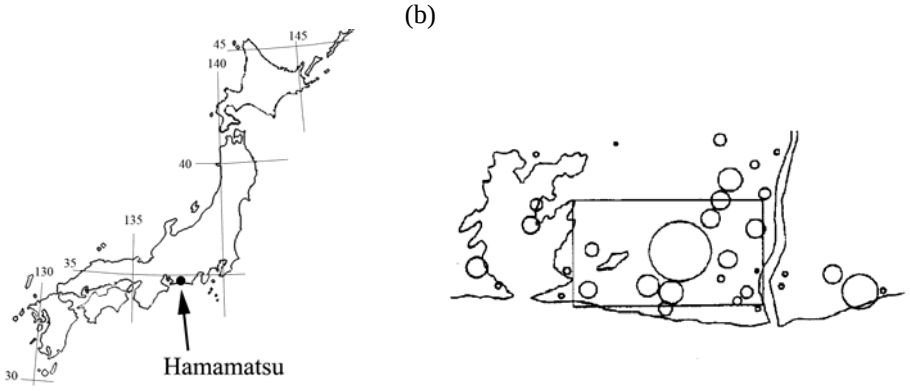


Fig. 1. (a) Location of Hamamatsu. (b) Research area (Hamamatsu Chamber of Commerce and Industry, 1951).

Geography of factories has been studied intensively by many authors (Dacey, 1962; Harvey, 1969; Scott, 1988; Sugiura, 2003). The clumping behavior in factory structure or the formation of "factory district" has been pointed out in relation to ecological distribution (King, 1962; Burton, 1963; Haggett, 1965). In these studies, auto-correlation is adopted as a statistical approach (Schaefer, 1953; Clark and Evans, 1954). In the present paper, we also apply the auto-correlation. However, we use a lattice square: All data of Fig. 2 are once redrawn on the lattice space. After such a replacement, we measure the auto-correlation. This is because our purpose is to represent the dynamics of factories by a simple lattice model.

2. CONTACT PROCESS (CP)

In order to represent the dynamics of factories, we apply "contact process" on a two-dimensional lattice. The contact process (CP) was first introduced by Harris (1974) as a model of infectious disease. CP has been extensively studied from mathematical (Harris, 1974; Liggett, 1985), physical (Katori & Konno, 1991; Marro & Dickman, 1999) and ecological (Miyazaki et al, 2006; Itoh et al, 2004) aspects. Each lattice site takes one of two states: empty (O) or occupied (X) by factory. Interactions for CP are defined by

$$\mathrm{X} + \mathrm{O} \xrightarrow{@} 2\mathrm{X} \quad (\text{rate } c), \qquad (1)$$

$$\mathrm{X} \xrightarrow{@} \mathrm{O} \quad (\text{rate } d), \qquad (2)$$

The reaction (1) means the construction process of factories; a new factory is build at the neighboring site of existing factory. The parameter c thus denotes the construction rate of factory. On the other hand, the reaction (2) denotes the dismantlement process, and the parameter d is a dismantlement rate. In the original model (Harris, 1974), the site X (or O) meant the infectious (or healthy) person; the reactions (1) and (2) respectively represented the infection and healing of disease.

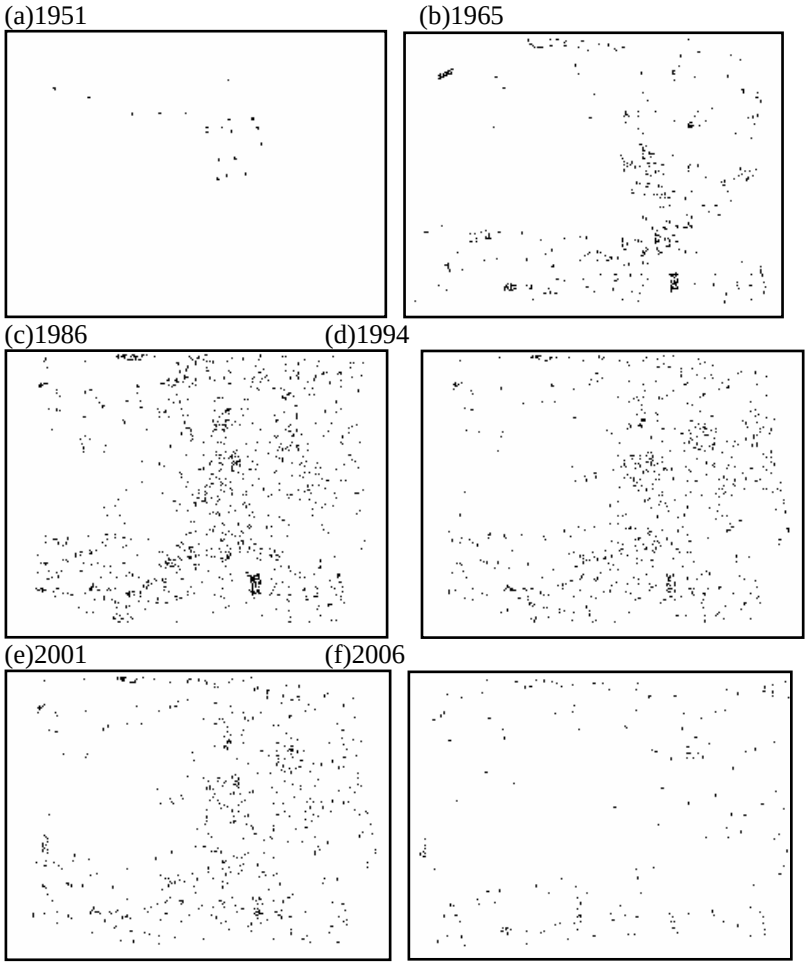


Fig. 2 Spatial and temporal changes of factories in Hamamatsu. Each plot represents the centre of factory which has more than three workers

In the field of ecology, CP is also used; in this case, the site X means an individual or the occupied site by biospecies. The reactions (1) and (2) respectively represent the birth and death processes. The site O denotes empty. The parameter *c* and *d* are birth and death rates, respectively.

Simulation method for CP is as follows:
1) Initially, we distribute factories on the lattice. Each lattice site is one of two states; namely occupied (X) and empty (O) sites.
2) The reactions (2) are performed in the following two steps:
(i) We perform two-body reaction (1). Choose one lattice site randomly, and then randomly specify one of neighboring sites. If the pair of sites are (X, O), then O is changed into X.
(ii) We perform one-body reaction (2). Choose one lattice point randomly; if the site is occupied by X, the site will become O by the rate *d*.
3) Repeat the above step 2) T times, where T is the total number of lattice points. This is called the Monte Carlo step (MCS) (Tainaka, 1988). In other words, time in simulation is measured by the unit of MCS.
4) Continue the step 3) until an adequate time.

The initial structure is not so important, since the system evolves into a stationary state. The density (number) of X in the final (equilibrium) state depends on both parameters *c* and *d*. When the value *d* (or *c*) is increased (or decreased), then the final density decreases. According to the previous studies, the final density is uniquely determined by the ratio *d/c* (Harris, 1974; Marro & Dickman, 1999). Namely, CP can be represented by a single parameter. For this reason, in the present paper, we fix the value of *c* (*c*=1), and change parameter *d*.

3. METHODS

Lattice representation

We first represent the actual data in Fig. 2 by the use of a lattice. The rectangle of 18km× 24km is divided into the mesh of 180×240. The total number of lattice sites is thus given by 180× 240 sites, and each site corresponds to a 100-meter square. The state of a single lattice site takes one of two states; either occupied (X) or empty (O). The site X means there is at least one factory in the square. In contrast, the site O means the absence of factory. Overall profile in the lattice representation is almost the same as spatial pattern displayed in Fig. 2.

Statistical method: Auto-Correlation

We analyse the spatial pattern by the use of correlation function. Let F(r) be the auto-correlation, where r is the distance between a pair of lattice sites. The distance r takes discrete values as shown in Fig. 3. The shortest distance (r=1) means the nearest (one-step) neighbour, and r=2 means the two-step distance. The distance r=1 is most important, since the correlation function is usually a decreasing function of distance. The auto-correlation F(r) means the conditional probability. Assume that there is X at a lattice site A (centre site in Fig. 3), then F(r) is defined by the probability finding X at another site B, where B locates at the distance r apart from the site A. If the distribution is completely random, F(r) is equal to the overall density (the probability finding X at a site). Hereafter, the auto-correlation is scaled (divided) by the overall density; in this way, we can set F(r) =1 for the random distribution. According to the value of auto-correlation, the spatial structure of factory is classified as follows (Tainaka and Fukazawa, 1992):

(a) F(r) >1: aggregated (clustering) pattern
(b) F(r) =1: random distribution
(c) F(r) <1: uniform (dispersed) pattern

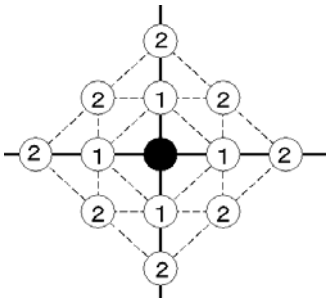


Fig. 3
Schematic illustration of distance r. The numerals in circles denote the distance (r). The nearest (one-step) neighbour is represented by r=1, and the two-step neighbour corresponds to r=2.

4. RESULTS

4.1 Statistical Results

From the spatial and temporal data of factories at Hamamatsu, we obtain statistical results. In Fig. 4, the time dependences of both density (number) and degree of clumping are displayed. It is found from Fig. 4 that the degree of clumping was crucially influenced by the collapse of stock market (near 1990). Just before the collapse (in “bubble period”), the factories nearly followed a random distribution rather than aggregation. In contrast, in the early stage of industrialization (1951), the distribution of factories showed an extreme aggregation tendency. Almost all factories were clustered in local area (“factory district”). However, as the number of factory had increased during the bubble economic stage, the degree of clumping had decreased almost to be random. After the collapse of bubble, the clumping tendency returned. The value of clumping degree in 2006 was much smaller than that in 1951. The present-day factories are not so aggregated compared with the early stage. This may be associated with the fact that JIT system became more flexible than the early stage (Nojiri, 2005; Linge, 1991). In summary, when the industrialization is not yet fully

developed, the aggregation is very strong. In contrast, when the density of factories is high (bubble period), the factories nearly distribute randomly. After the bubble, the degree of clumping slightly recovered.

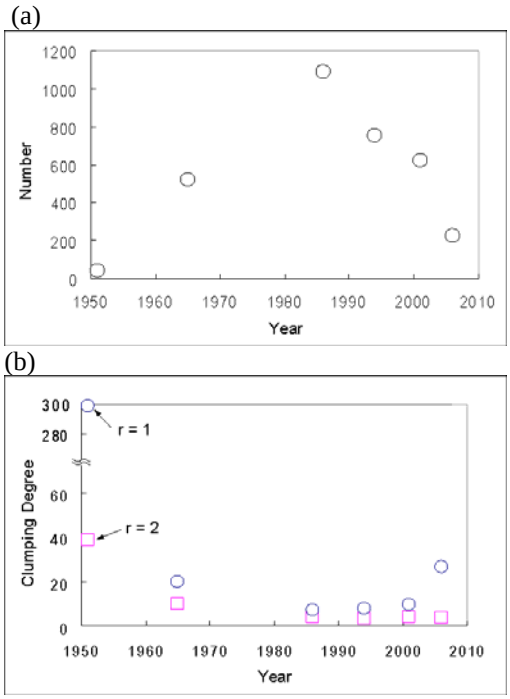


Fig. 4
The statistical results of factories in Hamamatsu. (a) Number of factories, and (b) degree of clumping (auto-correlations) F(*r*), where *r*=1, 2. If the distribution of individuals is random, then F(*r*) =1. In the early stage of industrialization (1951), the aggregation is found to be very strong. Just before the collapse (in "bubble period"), the factories nearly form a random distribution rather than aggregation

4.2 Simulation Results

The clustering profile of factories is very similar to that of biological individuals. We represent the construction and dismantlement processes by the contact process. The parameter value of *d* is changed as shown in Fig. 5. Here we assume the linear time dependence (straight lines); the slope of the line is positive before the collapse of stock market (1990), while it is negative after 1990. The value of *d* is determined as the density profile agrees with actual data in Fig. 4(a).

In Fig. 6, the simulation results are shown. It is apparent that the time dependence of density in Fig. 6(a) is almost the same as that in Fig. 4(a). The important point is the outcome in Fig. 6(b). The profile of clumping degree in Fig. 6(b) is similar to the real data in Fig. 4(b). When the total number of factories is small, the aggregation is strong. Figure 7 illustrate a typical spatial pattern in the case of low density. It is found that factories are densely aggregated. In contrast, when the density of factories is high, the factories nearly form a random distribution. In particular, at the bubble period, the factories nearly form a random distribution. These characteristics can be represented by the simulation of contact process.

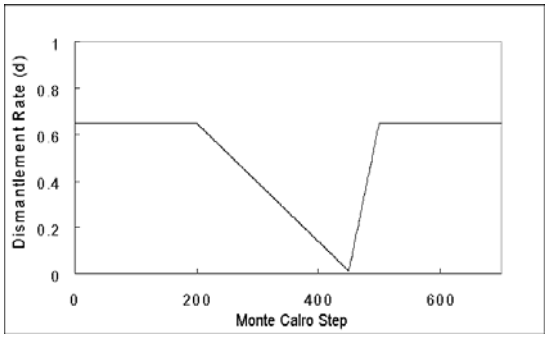


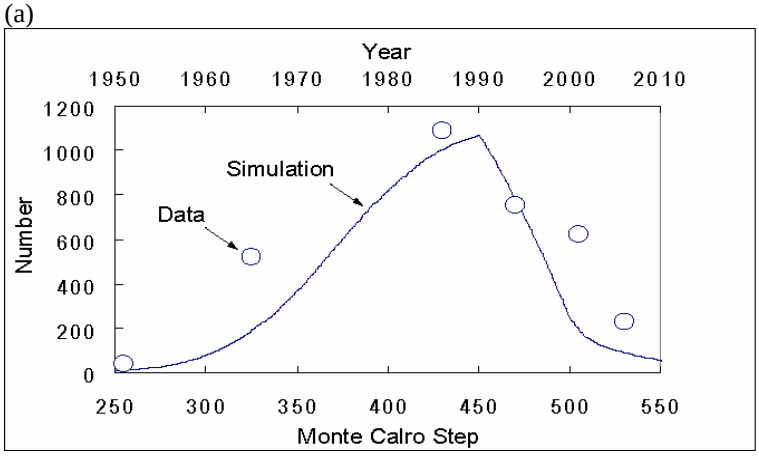
Fig. 5
The assumption of value of parameter *d*. It is known for the contact process that we can put *c*=1 without the loss of generality.

5. DISCUSSION

We study dynamics process of factories at Hamamatsu city in Japan. The time dependence of factory number can be classified into two periods. Before the collapse of stock market (1990), the number of factories increases smoothly. On the contrary, after the collapse, it rapidly decreases. The spatial structures of factories are statistically analyzed by auto-correlation (clumping degree). It is found that the factories more or less form aggregated or crowded distribution. When the total density of factories is low, the aggregation is found to be strong. In contrast, just before the collapse (bubble period), the spatial pattern of factories is almost random.

We represent the construction and dismantlement processes of factories by the contact process which is a simple simulation model. The time dependence of clumping degree (Fig. 6(b)) has the similar profile to the real data in Fig. 4(b). Both characteristics of density and auto-correlation can be explained by the use of a single parameter (ratio or dismantlement rate). The values of auto-correlation obtained from simulations (Fig. 6(b)) agree well qualitatively with the actual data (Fig. 4(b)). Note that the fact that the absolute values does not match with the simulated values is not important, because the absolute density of factories depends on the sizes of cells and study areas. If we consider larger area than that in Fig. 1(b), then the density (clumping degree) should be decreased (increased). This is because the factory density outside Hamamatsu is much less than that in Hamamatsu. The important point is the relative changes of clumping degree or density in time

(a)



(b)

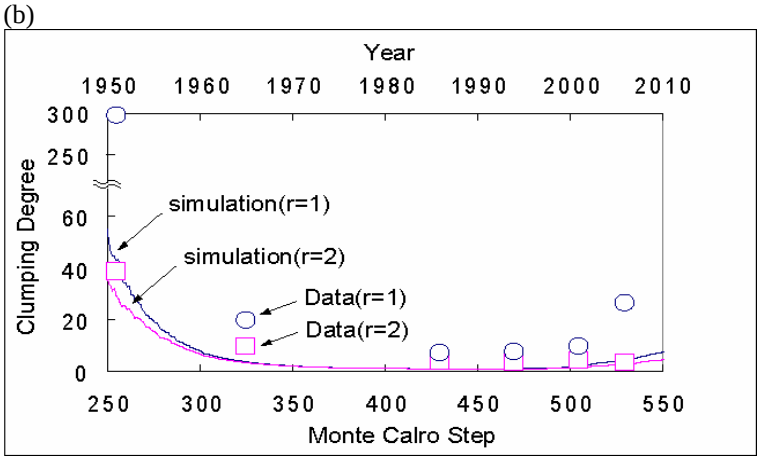


Fig. 6
The results of simulation. (a) Number of factories, and (b) degree of clumping. The actual data are represented by plots which are the same as in Fig. 4.

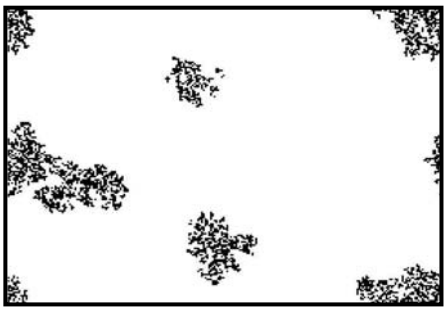


Fig. 7
Snapshots of typical spatial patterns at t = 540. Black and white sites denote occupied and empty site of factory.

We discuss the reason why the aggregation of factories is naturally formed. In Hamamatsu city, there are three major companies: Honda and Suzuki motor companies and Yamaha Co. They are all very famous as a motor company in the world. When the industrialization is not yet fully developed (1951), the aggregation is found to be very strong. Such an aggregated formation of factories (factory district) may be associated with the construction process of factory. Namely new factory tend to be built near existing factories [the reaction (1)]. This is convenient for cutting the costs in transportation (Estall, 1985) and information exchange (Dyer, 1996).

Finally, we discuss the relation between degree of clumping and just-in-time (JIT) system. In 1950's, Toyota motor companies introduced JIT system. A decade later, JIT had been adopted by the motor companies in Hamamatsu (Sheard, 1983). Due to JIT system, physical distribution costs could have been reduced. Our results indicate that the degree of clumping is strongest, when the industrialization is not yet fully developed (1951). During the bubble period, the needs for transportation cost-cuts were reduced, due to high demands for products. After the collapse of markets in 1990, the distribution of factories became aggregated (clumped) again, but a much lesser degree. This might be due to the fact that the characteristics of JIT have been diversified due to the severe evaluation by consumers after the bubble stage (Nojiri, 2005). Thus the temporal changes in the distribution of factories in Hamamatsu might be explained by the temporal economic changes and the characteristics in JIT systems. Our results may suggest that the causal mechanisms of aggregation tendencies in industry may be well simulated by a contact process in the lattice model.

ACKNOWLEDGMENTS
The authors wish to thank Mr. Taro Hayashi for his help in simulations.

REFERENCES

Burton, I., The Quantitative Revolution and Theoretical Geography, *Canadian Geogr*., 7, 151-162, 1963.

Clark, P.J. and F.C. Evans, Distance to nearest neighbor as a measure of spatial relationship in populations, *Ecology*, 35, 445-453, 1954.

Dacey, M.F., "Analysis of Central Place and Point Patterns by a Nearest Neighbor Method" in Norborg, K. ed. "Proceedings of the I.G.U. Symposium in Urban Geography. Lund. 1960", *Lund Studies in Geography*, B. Vol. 24, CWK (Gleerup, Lund) 55-75, 1962.

Dyer, J.H., Specialized supplier networks as a source of competitive advantage: evidence from the auto industry, *Strategic Management Journal*, 17, 271-291, 1996.

Estall, R.C., Stock control in manufacturing: the Just-in-Time systems and its locational implications, *Area*, 17, 129-133, 1985.

Haggett, P., Location Analysis in Human Geography, London, Edward Arnold, 1965.

Hamamatsu Kikakushitsu, Hamamatsu-City Statistics Book (Hamamatsu-City Group Handbook before 1951), Hamamatsu (in Japanese), 2006.

Hamamatsu Kikakushitsu, Hamamatsu Developing History, Hamamatsu (in Japanese), 1954.

Hamamatsu Chamber of Commerce and Industry, Hamamatsu Shoko Meikan (in Japanese), 1951.

Hamamatsu Chamber of Commerce and Industry, Hamamatsu Shoko Meikan (in Japanese), 2006.

Harris, C.D., "The Market as a Factor in the Localization of Industry in the United States", *Annals of the Association of America Geographers*, 44, 315-348, 1954.

Harris, T.E., Contact interaction on a lattice, *Ann Prob.*, 2, 969-988, 1974.

Harvey, D., Explanation in Geography, London, Edward Arnold, 1969.

Itoh, Y. Tainaka, K. Sakata, T. Tao, T. and Nakagiri, N., Spatial Enhancement of Population Uncertainty near the Extinction Threshold, *Ecological Modelling*, 174, 191-201, 2004.

Katori, M. and Konno, N., Upper bounds for survival probability of the contact process, *J. Stat. Phys.*, 63, 115-130, 1991.

King, L.J., A Quantitative Expression of the Patterns of Urban Settlement in Selected Areas of the United States, *Tijdschrift voor Economische en Sociale Geografie*, 53, 1-7, 1962.

Liggett, T. M., Interacting particle systems, New York, Springer-Verlag, 1985.

Linge, G.J.R., Just-in-time: more or less flexible?, *Economic Geography*, 67, 316-332, 1991.

Mair, A., Just-in-Time manufacturing and the spatial structure of the automobile industry: lessons from Japan, *Tijdschrift voor Economische en Sociale Geografie*, 83, 82-92, 1992.

Marro, J. and Dickman, R., Nonequilibrium phase transition in lattice models, Cambridge, Cambridge Univ., Press, 1999.

Miyazaki, T. Tainaka, K. Togashi, T. Suzuki, T. and Yoshimura, J., Spatial coexistence of phytoplankton species in ecological timescale, *Popul. Ecol.*, 48, 107-112, 2006.

Nojiri, W., Review of studies on Just-in-Time in economics and economic geography, *Memoirs of Momoyama Gakuin University*, 31, 21-32, 2005.

Scott, A.J., Metropolis: From the Division of Labor to Urban Form. Berkeley and Los Angeles, University of California Press, 1988.

Schaefer, F. K., Exceptionalism in Geography: A Methodological Examination, *Ann. Assoc. Amer. Geogr*., 43, 226-249, 1953.

Sheard, P., Auto production systems in Japan; organizational and locational features, *Australian Geographical Studies*, 21, 49-68, 1983.

Shizuoka Tiiki-gakkai, Shizuoka no tiikisangyou (Area's Industries of Shizuoka), Tokyo, Kyoritu (in Japanese), 1996.

Sugiura, Yosio, Tiri Kukan Bunseki (Spatial Analysis in Geography), Tokyo, Asakura (in Japanese), 2003.

Tainaka, K., Lattice model for the Lotka-Volterra system, *J. Phys Soc. Jpn*, 57, 2588-2590, 1988.

Tainaka, K. and Fukazawa, S., Spatial pattern in a chemical reaction system: prey and predator in the position-fixed limit, *J. Phys. Soc. Jpn*, 61, 1891-1894, 1992.

Ullman, E.L., "The Role of Transportation and the Bases of Interaction" in Thomas, W.L. (ed.), Man's Role in Changing the Face of the Earth, Chicago, 1956.

Weber, A. and Friedrich, C.J. (ed.), Alfred Weber's Theory of the Location of Industry, Chicago, 1929.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

An approach to formal modeling of environmental knowledge via concept maps and ontologies

A. Simón-Cuevas[1], L. Ceccaroni[2] and A. Rosete-Suárez[1]

[1] Centro de Estudios de Ingeniería de Sistemas (CEIS), Facultad de Ingeniería Informática, Instituto Superior Politécnico "José Antonio Echeverría", 114, No. 11901, Marianao, C. Habana, Cuba ({asimon, rosete}@ceis.cujae.edu.cu)
[2] Departament de Llenguatges i Sistemes Informàtics (LSI), Universitat Politècnica de Catalunya (UPC), Campus Nord, Edif. Omega, C. Jordi Girona, 1-3, 08034 Barcelona, Spain (luigi@lsi.upc.edu)

Abstract. In order to obtain semantic interoperability in open systems, the systems involved need to agree on commonly understandable knowledge representations. The case considered in this paper is one of a system needing to communicate and share knowledge using concept maps and ontologies, in the context of the environmental sciences and the semantic Web. We present how to formally obtain ontologies codified in the OWL language from concept maps. Concept maps are a flexible and informal form of knowledge representation, while OWL is a language oriented to processing carried out by machines. The mapping between concept maps and ontologies is a formal transformation, which semantically analyzes the relations linking the concepts in the map. The proposed method includes a concept-sense disambiguation procedure and uses the WordNet lexical knowledge base. It also includes automatic learning of the semantics of the relations between concepts. The proposed method has been applied to the environmental-knowledge domain, through tests of several concept maps with labels in Spanish.

***Keywords**:* Semantic analysis; knowledge sharing; concept map; ontology; OWL.

1. INTRODUCTION

In the environmental sciences, as in most scientific domains, information needs sometimes to be analyzed and processed by machines. In the knowledge representation oriented to the semantic analysis and processing by machines, context in which a certain degree of formalization is required, the development and use of *ontologies* is increasingly common. However, the processes of designing and creating ontologies, the tools available, and the specification languages are still complex for non-experts in this subject. This complexity represents a difficulty in environments requiring the collaboration of humans for the development and processing of ontologies. This suggests that a form of representation that can be used naturally by humans and integrated with ontologies (in such a way that the latter can be automatically obtained) should be useful.

Concept maps (CMs) [Novak and Gowin, 1984] are one of these human-friendly knowledge-representations. The integration between CMs and ontologies, specifically in the case of OWL [Smith et al. 2004] ontologies, is pursued through the incorporation of more formalization in CMs and through the analysis of the relations among concepts. Parts of the proposed method are based on a concept-sense disambiguation (CSD) algorithm defined by Simón et al. [2007a], and on WordNet [Miller et al. 1990]. The CSD algorithm

tries to assign the most rational sense of a given concept in the CM, using WordNet, *contextual analysis* and *domains information*. The algorithm explores the context in which a given concept appear in the CM and try to determine a corresponding, similar context in WordNet, using the *synsets* of the concept at issue. A similar contextual analysis is carried out with the gloss.

Along the paper, to represent the English translation of the Spanish terms used, the following notation will be used: *español* ("Spanish").

2. CONCEPT MAPS

CMs are a tool especially defined for application in the learning process; they are easy to be created, flexible and intuitive for people. They are a graphically rich technique for organizing and representing knowledge proposed by Novak and Gowin [1984]. They include *concepts*, *linking-words* (relationships between concepts) and *propositions*. CMs' propositions contain two or more concepts connected using linking-words and sometimes are called semantic units, or units of meaning. The concepts are represented in a hierarchical fashion with the most inclusive, most general concepts identified as *root-concept*, at the top of the CM and the more specific, less general concepts arranged hierarchically below. Figure 2 shows an example of a CM, about the nitrogen cycle as a part of the ecosystem in which *Nitrógeno* ("nitrogen") is the *root-concept*. The CM shown is an enriched interpretation of the picture in Figure 1. CMs, being easily created, flexible and intuitive for people, are especially useful for knowledge management.

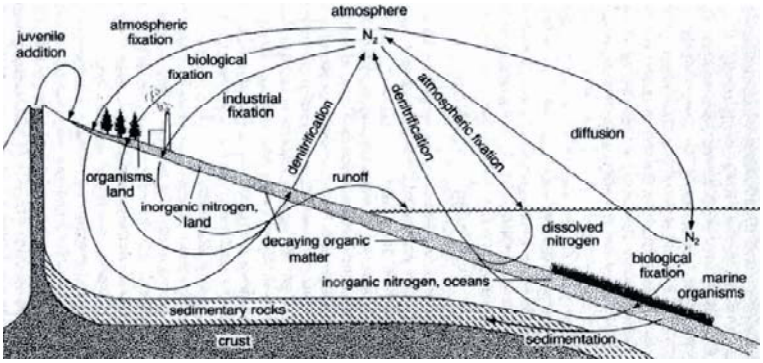


Figure 1. Representation of nitrogen cycle obtained from Jones et al. [1992].

3. ONTOLOGIES AND OWL DL

In artificial intelligence, ontologies were introduced to share and reuse knowledge. They provide the common reference frame for communication languages in distributed environments (such as multi-agent systems or the semantic Web) and a formal description for automatic knowledge processing. Several languages have been defined to implement them; OWL [Smith et al. 2004] is the latest, standardized ontology language. OWL is based on XML, RDF and RDFS, and includes three specifications, with different expressiveness levels: OWL Lite, OWL DL and OWL Full. The code obtained by the method described in this paper is generated according to OWL DL specifications. OWL DL is so named due to its correspondence with *description logics*. *Description logic* (DL) is the name for a family of knowledge representation formalisms that represent the knowledge of a domain by first defining the relevant concepts of the domain (its terminology), and then using these concepts to specify properties of objects and individuals occurring in the domain [Baader and Nutt, 2003]. The terminology specifies the vocabulary of a domain, which consists of concepts and roles, where the concepts denote individuals while roles denote binary relationship between individuals.

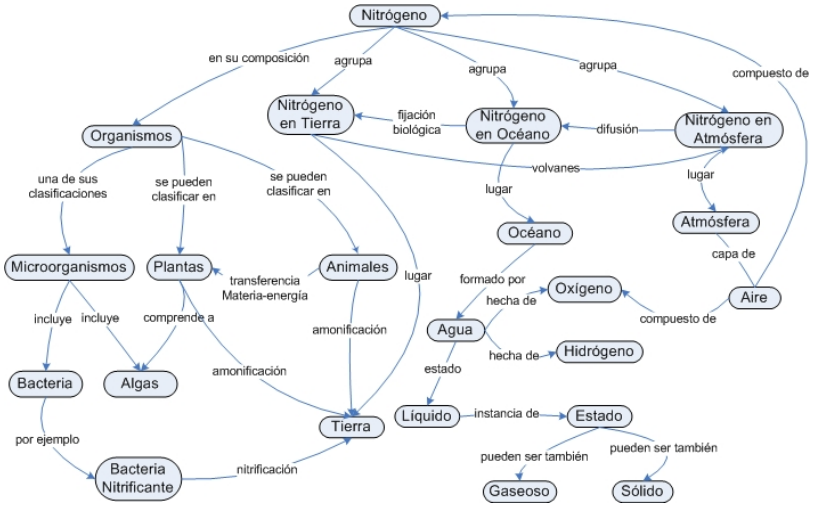


Figure 2. Concept map of the nitrogen cycle (in Spanish).

4. BASIC ASPECTS OF CM-OWL MAPPING

The knowledge in OWL ontologies is expressed as *classes*, *properties* and *instances* [Smith et al. 2004], while in CMs much of this formal and explicit specification does not exist, and has to be inferred. Nonetheless, some initial structural mapping between CMs and OWL can be easily established [Simón et al. 2007b]:

- *concepts* correspond to: *classes* and *instances*;
- *linking-words* correspond to *properties* (A property is a binary relation between instances of classes in OWL [Smith et al. 2004].);
- *propositions* correspond to *classes and properties' restrictions* and other OWL constructs.

Some type of semantic relation, such as *class-subclass*, *class-property*, *class-property-value*, *class-instance*, can be inferred from certain *linking-words* used in CMs, in accordance with Brilhante et al. [2006].

In addition to the CM to be formalized, two external knowledge sources will be used in this work: WordNet [Miller et al. 1990] and a CM repository. WordNet is a lexical knowledge base, whose basic structure is the *synset. Synsets* form a semantic network and are interconnected among themselves by several types of relations, some of which are used in the proposed algorithm, such as *hypernymy-hyponymy* (class/subclass) and *meronymy-holonymy* (part/whole). The *synset* define the meanings of a word, which, in the case of polysemy, can be found in various *synsets*. WordNet can be used as an ontology if its links are associated to a formal semantics. The CM repository used here is ServiMap [Simón et al. 2006], which stores several CMs of different domains (including the environmental domain).

The mapping and semantic inference leading to OWL coding in this paper, is carried out combining the analysis of:

- the syntax of the *propositions;*
- the occurrence of similar relations in WordNet and the external CM repository.

Initially, some frequently used *linking-words* are defined and organized in four **categories**, according to the semantics that can be associated to them and their correspondence with the semantic relations in WordNet. They are:

- **Classification** (CC), for *linking-words* that may indicate *(super class-class)*relations between concepts in a *proposition* (e.g., *comprende a* ("comprise") in the proposition *(Plantas, comprende a, Algas)* in Figure 2); corresponding to *hypernymy* and *hyponymy* relations in WordNet;
- **Instance** (IC), for *linking-words* that may indicate *(class-instance)* relations between concepts in a *proposition* (e.g., *por ejemplo* ("for example") in the proposition *(Bacteria, por ejemplo, Bacteria Nitrificante)* in Figure 2);
- **Property** (PC), for *linking-words* that may indicate *(class-property)* relations between concepts in a *proposition* (e.g., *tiene* ("has")); corresponding to *has_meronym* and *has_holonym* relations in WordNet;
- **Property-Value** (PVC) for *linking-words* that may indicate *(class-property-value)* relations between concepts in a *proposition*, such as nouns (e.g., *lugar* ("place") in the proposition *(Nitrógeno en Océano, lugar, Océano)* in Figure 2); corresponding to basic *meronymy* and *holonymy* WordNet's relations, and different from the more specific *has_meronym* and *has_holonym* relations (e.g. *has_mero_partOf* and *has_holo_madeOf*).

This method allows everyday natural language to be used at CM construction time. Lexemes are used to avoid duplications due to verb forms' variability, e.g.: *ten-* (from *tener*) instead of *tiene* ("has") and *tienen* ("have"). The *linking-words* are defined in Spanish, which is the natural language used for this study, and they are continually and automatically enriched.

In the mapping method, the CM under consideration is analyzed as a structured text. A concept sense-disambiguation algorithm [Simón et al., 2007a] is used to infer the most rational sense (in terms of WordNet's *synsets*) for all concepts in the CM. Once inferred a *synset* for each concept in a *proposition*, the semantics of the CM relation among them can be inferred from the relation in WordNet (if one exists).

5. OBTAINING OWL-DL ONTOLOGIES

In this section, the process of obtaining OWL-DL ontologies form CMs is presented. It begins with a CM expressed as an XML file, in which the elements are ordered using a breadth-first strategy (beginning from the *root-concept*) to explore the content of the CM. The content structure (concepts, *linking-words* and propositions) of the CM is maintained in the XML file. The process is organized in three phases: preprocess, mapping and codification. Four modules are defined for the implementation of the system: parser, disambiguator, semantic interpreter and OWL codifier.

5.1 Preprocess phase

The ***parser*** analyzes the CM to be translated to OWL, identifying propositions and their parts (*concepts* and *linking-words*):

- It extracts all propositions from the CM's XML file and creates a proposition set (PS) with (C_o, *linking-word*, C_d) as basic structure, where C_o is the origin concept, and C_d is the destination concept.
- It creates the concepts set (CS) which includes all concepts in PS.
- It extracts from the CM repository all propositions with at least one concept included in CS and creates, with these, a second proposition set (PS-CMR), which is used to infer symmetric properties.

The ***disambiguator*** infers the most rational *sense* (in terms of WordNet *synsets*), using the algorithm defined by Simón et al. [2007a], and analyzes this sense according to the relations between *synsets* found in WordNet:

- A **PS_WN**$_{hype\text{-}hypo}$ set is created with the pair (C, C'), if the *synsets* of two concepts *C* and *C'* are directly related by a *hyperonymy* or *hyponymy* relation in WordNet.
- A **PS_WN**$_{mero\text{-}holo}$ set is created with:
  the pair (C, C') if the *synsets* of two concepts *C* and *C'* are directly related by a *has_meronym* WordNet's relation and;
  the pair (C', C) if the *synsets* of two concepts *C* and *C'* are directly related by a *has_holonym* WordNet's relation.
- A **PS_WN**$_{mero\text{-}holo\text{-}type}$ set is created with:
  the triple (C, C', type of relation) if the *synsets* of two concepts *C* and *C'* are directly related by some type of *meronymy* WordNet's relation (e.g. *has_mero_madeOf*), different from *has_meronym* and;
  the triple (C', C, type of relation) if the *synsets* of two concepts *C* and *C'* are directly related by some type of *holonymy* WordNet's relation (e.g. *has_holo_madeOf*), different to *has_holonym*.

5.2 Mapping phase

Several heuristic rules (not presented here) are defined for mapping between the propositions included in PS to OWL constructs. These rules use the set of *linking-words* (*lw*) included in the categories presented in section 4, the sets generated in the *preprocess phase* (see section 5.1) and the sets generated as follows:

- **S_CS** is a set of pairs of concepts *(C, C')* which are part of a proposition of type (*superclass*, *lw*, *subclass*), where *C* is the superclass and *C'* is the subclass; the *lw* included in CC and the PS_WN$_{hype\text{-}hypo}$ set are used.
- **S_CI** is a set of pairs of concepts *(C, I)* which are part of a proposition of type (*class*, *lw*, *instance*), where *C* is the class and *I* is an instance of *C*; the *lw* included in *IC* are used.
- **S_CP** is a set of pairs of concepts *(C, Pr)* which are part of a proposition of type (*class*, *lw*, *property*), where *C* is the class and *Pr* is a property of *C*; the *lw* included in *PC* and the PS_WN$_{mero\text{-}holo}$ set are used.
- **S_CPV** is a set of propositions of type (*class*, *property*, *range of value*); the *lw* included in *PVC* and the PS_WN$_{mero\text{-}holo\text{-}type}$ set are used.
- **S_CPV**$_{hasValue}$ is a set of propositions of type (*class*, *property*, *instance*).
- **Pr**$_{symmetric}$ is a set of symmetric properties.
- **Pr**$_{functional}$ is a set of functional properties.
- **Intersection**$_{classes}$ is a set of pairs *(C, {…C'$_i$…})*, where *C* is subclass of each *C'$_i$*;
- **Intersection**$_{classes\text{-}property}$ is a set of triples *(C, C', Pr)*, where *C* is subclass of *C'* and *Pr* is one property of *C*;
- **Union** is a set of pairs *(C, {…C'$_i$…})*, where *C* is superclass of each *C'$_i$*.

The ***semantic interpreter*** applies the set of heuristic rules to the propositions obtained by the *parser*.

5.3 Codification phase

The ***OWL codifier*** uses the sets generated by the *semantic interpreter* and writes out the corresponding OWL constructs according to W3C Recommendation [Smith et al. 2004], considering the mapping conventions shown in Table 1.

6. APPLICATION IN THE ENVIRONMENTAL DOMAIN

The theoretical modeling method presented has been applied to several CMs about the environmental domain. These CMs have been constructed from different source (texts and

figures) with the assistance of environment experts (e.g., the Figure 2 is constructed using the Figure 1 as source). In summary, the modeling method starts with the CM constructed using Macosoft CM editor [Simón et al. 2006] and expressed as an XML file. In the preprocess phase, all propositions are extracted by the *parser* and stored in *PS*; propositions in the CM repository whose concepts appear in *PS* are also obtained by the *parser*.

Table 1. Conventions for mapping between inferred sets and OWL constructs.

Inferred sets	**Basic structure**	**OWL constructs**
S_CS	(C, C')	*C* is coded as *owl:class*. *C'* as *owl:class* and also as *rdf:subClassOf* (of *C*).
S_CI	(C, I)	*C* is coded as *owl:class*. *I* as an instance of *C*: *<I rdf:ID = "C" />*.
S_CP	(C, Pr)	An *owl:ObjectProperty* with the name formed by the concatenation of labels *tiene* ("has") and the one used for *Pr* is coded. *C* is coded as *owl:class* and also as *rdfs:domain* (of *owl:ObjectProperty*). *Pr* is coded as *owl:class* and also as *rdfs:range* (of *owl:ObjectProperty*).
S_CPV	(C, Pr, C')	*C* is coded as *owl:class*. *Pr* as *owl:onProperty* (of *C*). *C'* as restriction *owl:someValueFrom* (of *Pr*) in the code of *C*.
$S_CPV_{hasValue}$	(C, Pr, C')	*C* and *C'* are coded as *owl:class*. *Pr* is code as *owl:onProperty* (of *C*) with restriction of *owl:hasValue* being *C'* the value.
$Pr_{symmetric}$	$Pr_1 \ldots Pr_n$	The *owl:SymmetricProperty* construct is incorporated to the code of each Pr_i.
$Pr_{functional}$	$Pr_1 \ldots Pr_n$	The *owl:FunctionalProperty* construct is incorporated to the code of each Pr_i.
$Intersection_{classes}$	(C, {…C'$_i$…})	*C* and C'_i are coded as *owl:class*. The collection of C'_i is codes as *owl:intersectionOf*.
$Intersection_{class\text{-}property}$	(C, C', Pr)	*C* is coded as *owl:class*. C' is coded as owl:class. *Pr* as *owl:onProperty* (of *C*), with restriction *owl:hasValue* of *C'*. The collection of *C'* and *Pr* is coded as *owl:intersectionOf*.
Union	(C, {…C'$_i$…})	*C* is coded as *owl:class*. C'_i as *owl:class*. Collection of C'_i as *owl:unionOf*.

Next, the sense of each concept included in *PS* is inferred, and the *hyperonymy* and *meronymy* relations between *synsets* in a Spanish version of WordNet (the version developed by the Natural Language Processing Group (TALP) of the Software Department (LSI) of the Technical University of Catalonia (UPC)) are identified by the *disambiguator*. We use a portion of the CM shown in Figure 2 as the *experimental set* (ES), which includes the propositions for the following modeling example. All propositions included in ES are sequentially analyzed by the *semantic interpreter*:

- In the case of proposition *(Nitrógeno, agrupa, Nitrógeno en Tierra)*, the rule *if* $P = (C_o, lw, C_d) \in PS \wedge lw \in CC$ *then* $PS = PS - \{P\}$, $S_CS = S_CS \cup \{(C_o, C_d)\}$ is executed. The *lw agrupa* ("groups") is included in Classification category.
- In the case of proposition *(Organismos, una de sus clasificaciones, Microorganismos)*, the rule *if* $P = (C_o, lw, C_d) \in PS \wedge lw \notin CC \wedge (C_o, C_d) \in PS_WN_{hype\text{-}hypo}$ *then* $PS = PS - \{P\}$, $S_CS = S_CS \cup \{(C_o, C_d)\}$, $CC = CC \cup \{lw\}$ is executed. The *synsets {ser_vivo#1, ser#1, organismo#1}* of concept *Organismo* ("organism") and *{microorganismo#1}* of concept *Microorganismo* ("microorganism") are inferred by the *disambiguator* and the *has_hyponym* relation between them is found in WordNet; consequently, the *lw una de sus clasificaciones* ("one of their classification") is added to Classification category.

- In the case of proposition *(Aire, compuesto de, Oxígeno)*, the rule *if* $P = (C_o, lw, C_d) \in PS \wedge (C_o, C_d, type) \in PS_WN_{mero\text{-}holo\text{-}type}$ *then* $PS = PS - \{P\}$, $S_CPV = S_CPV \cup \{(C_o, property, C_d)\}$, $PVC = PVC \cup \{lw\}$ is executed. In this rule, *property*, is the property corresponding to type in PVC, for example, *hecho de* ("made of") in case of *has_mero_madeof* relation. The *synsets {aire#1}* of concept *Aire* ("air") and *{oxígeno#1, O#1, número_atómico_8#1}* of concept *Oxígeno* ("oxygen") are inferred by the *disambiguator* and the *has_mero_madeof* relation between them is found in WordNet; consequently, the *lw compuesto de* ("compose of") is added to the Property-Value category.
- In the case of proposition *(Líquido, instancia de, Estado)*, the rule *if* $P = (C_o, lw, C_d) \in PS \wedge lw \in IC$ *then* $PS = PS - \{P\}$, $S_CI = S_CI \cup \{(C_o, C_d)\}$ is executed. The *lw instancia de* ("instance of") is included in Instance category.
- In the case of proposition *(Agua, estado, Líquido)*, the rule *if* $(C, Pr, I) \in S_CPV \wedge \exists (C', I) \in S_CI$ *then* $S_CPV_{hasValue} = S_CPV_{hasValue} \cup \{(C, Pr, I)\}$ is executed. The *lw estado* ("state") is a noun according to WordNet, therefore *(Agua, estado, Líquido)* is included in S_CPV as result of the rule *if* $P = (C_o, lw, C_d) \in PS \wedge lw \in PVC$ *then* $PS = PS - \{P\}$, $S_CPV = S_CPV \cup \{(C_o, lw, C_d)\}$ previously executed, and the pair *(Líquido, Estado)* is included in S_CI.

Finally, the code obtained by *OWL codifier* is:

```
<owl:Class rdf:ID="Nitrógeno"/>
<owl:Class rdf:ID="Oxígeno"/>
<owl:Class rdf:ID="Estado"/>
<owl:Class rdf:ID="Organismos"/>
<owl:Class rdf:ID="Nitrógeno en Tierra">
    <rdfs:subClassOf rdf:resource="#Nitrógeno"/>
</owl:Class>
<owl:Class rdf:ID="Microorganismos">
    <rdfs:subClassOf rdf:resource="#Organismos"/>
</owl:Class>
<owl:Class rdf:ID="Aire">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hecho de" />
        <owl:someValuesFrom rdf:resource="#Oxígeno" />
      </owl:Restriction>
   </rdfs:subClassOf>
</owl:Class>
<Líquido rdf:ID="Estado"/>
<owl:Class rdf:ID="Agua">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#estado" />
        <owl:hasValue rdf:resource="#Líquido" />
      </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>
```

7. CONCLUSIONS

To realize a semantic Web, tools are required that allow users with little technical background to generate their own ontologies and collaborate in the construction of distributed knowledge bases. The work presented here is a contribution to the creation of these tools: a method to formally obtain ontologies codified in the OWL-DL language from an informal knowledge representation, such as concept maps. In this paper, we combine the semantic analysis of some predefined *linking-words* (continuously automatically enriched), mechanisms of natural language processing based on a concept-sense-disambiguation

algorithm, and two knowledge bases (WordNet and a concept maps repository) in a novel way, which significantly distinguishes from the existing literature. The mapping between concept maps and OWL ontologies creates the bases for the collaborative development of ontologies in a more intuitive, friendlier manner for humans. In its current state, the method presented advances the state of the art through the use of tools and techniques from natural language processing, such as the integration of WordNet and the definition of a concept sense disambiguation algorithm. This, combined with the topological analysis of concept maps, allows maintaining a greater flexibility and more independence during concept map construction. These aspects are important for less-expert users in ontology construction, and are not considered by Gómez et al. [2004] and Hayes et al. [2005]. These aspects, combined with the increase in the formalizations level of the *linking-words* and the use of external knowledge representations, allow to augment the semantic inference in concept maps and to obtain more expressiveness in the resulting OWL than the ones reported by Gómez et al. [2004] and Brilhante et al. [2006]. The automatic enrichment of the *linking-words* repository is another contribution of the work presented in this paper. The method presented is mainly applicable to shallow domains, due to the fact that the terms in WordNet are about general knowledge; and this is an important restriction to be taken into account. Finally, the use of domain ontologies as alternative knowledge sources (to be used in a way similar to WordNet) and the use of concept-map repositories will increase the applicability of the method proposed; we are considering both issues as future work.

REFERENCES

Baader, F., and W. Nutt, Basic Description Logics, The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, 47-100, 2003.

Brilhante, V., G. Macedo, and S. Macedo, Heuristic Transformation of Well-Constructed Conceptual Maps into OWL Preliminary Domain Ontologies, paper presented at Workshop on Ontologies and their Applications, WONTO'2006, Brazil, 2006.

Gómez, H., B. Díaz, and A. González, Two layered approach to knowledge representation using conceptual maps description logic, paper presented at the first International Conference on Concept Mapping, Pamplona, Spain, 2004.

Hayes, P. J., T. C. Eskridge, R. Saavedra, T. Reichherzer, M. Mehrotra, and D. Bobrovnikoff, Collaborative knowledge capture in ontologies, paper presented at the third International Conference on Knowledge Capture, Canada, 99-106, 2005.

Jones, G., A. Robertson, J. Forbes, and G. Hollier, The HarperCollins Dictionary of Environmental Science, HarperCollins, 453 pp., 1992.

Miller, G. A., R. Beckwidth, C. Fellbaum, D. Gross, and K. J. Miller, Introduction to WordNet: An On-line Lexical Database, *International Journal of Lexicography*, *3*(4), 235-244, 1990.

Novak, J. D., and D. B. Gowin, Learning how to learn, Cambridge University Press, 1984.

Simón, A., V. Estrada, A. Rosete, and V. Lara, GECOSOFT: Un Entrono Colaborativo para la Gestión del Conocimiento con Mapas Conceptuales, paper presented at the second International Conference on Concept Mapping, Vol. 2, San José, Costa Rica, 114-117, 2006.

Simón, A., L. Ceccaroni, and A. Rosete, Concept sense disambiguation in concept maps using WordNet, Technical Report LSI-07-44-R, Departament de Llenguatges i Sistemes Informàtics (LSI), Universitat Politècnica de Catalunya (UPC), Campus Nord, Edif. Omega, c. Jordi Girona, 1-3, 08034 Barcelona, Spain, 2007a.

Simón, A., L. Ceccaroni, and A. Rosete, Generation of OWL Ontologies from Concept Maps in Shallow Domains, LNAI, Vol. 4788, Springer-Verlag, 259-267, 2007b.

Smith, M., Ch. Welty, and D. McGuinness, OWL Web Ontology Language Guide, W3C, 2004.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Managing Insular Tropical Environment through Data and Knowledge Bases by using Web Services: a case study on Corals and Herbarium of La Réunion Island

N. Conruyt, D. Sebastien, D. Payet, Y. Geynet, D. Caron, and D. Grosser

Laboratory of Computer Science and Mathematics, IREMIA, University of Reunion Island, Saint-Denis, 97715, France
(conruyt/didier.sebastien/denis.payet/geynet/caron/grosser@univ-reunion.fr)

Abstract: Decision-makers who want to manage Insular Tropical Environment more efficiently need to narrow the gap between the production of scientific knowledge in universities or other labs and its pragmatic use by the general public. One key environmental problem is to preserve biodiversity of ecosystems that are under human pressure. As we only protect what we know, a solution is to share expert knowledge about habitats and species on the Web for educating the public about their richness and beauty. Data and knowledge bases are part of a biodiversity information system that we have built to deal with this research enhancement problem, through the ETIC program. They were conceived using Web Services in order that each module communicates its functionalities and information one another and with external systems.

Keywords: Biodiversity Management; Knowledge Bases; Web Services; Information System; ETIC.

1. INTRODUCTION

The biological diversity of the islands in the South West Indian Ocean (Madagascar, Comoros, Mauritius, La Réunion, etc…) is still rich despite the important anthropic pressures, which are increasing from year to year. La Réunion Island is considered as one of the 20 richest world's biodiversity hotspot [Myers *et al.*, 2000]. But its population (also very diverse) is expected to reach the one million mark in just thirty years from the present 700,000. This will be a real problem for the management of the territory, as well as for the protection of fragile species. To face this environmental sustainable development problem, two natural parks have been created (one for the sea[1] and one for the mountains and cirques[2]). Their missions are to promote the natural and cultural heritage by increasing and sharing knowledge about its biodiversity, and to protect it against destruction. On the other side, the University of La Réunion has accumulated plenty of observations, data, information and knowledge on ecosystems over the past forty years. This information may be found in laboratory checklists, collections, museums, literatures, charts, maps, images, movies and sound files, individual databases, yet is hardly exploited by anyone except the authors themselves. There exists a gap between the missions of these institutions, because the first have to *manage biodiversity objects* whereas the second have to *produce data and knowledge about them* from their research activities. In order to propose new, sustainable practices of biodiversity management, research must not only lead to new knowledge and understanding of the interactions between the functioning of ecosystems and human activities, but also lead to pedagogical transmission of this knowledge through research en-

[1] [http://www.reunion.ecologie.gouv.fr/rubrique.php3?id_rubrique=2, visited on 11/04/08]

[2] [http://www.parc-national-reunion.prd.fr/, visited on 11/04/08]

hancement. Governments and European Commission that are aware of sustainable development make environmental education a priority[3], and reef ecosystems and tropical forests may be preserved against damage by creating awareness, sensitivity, and skills while fostering participatory action [Stepath, 2002].

2. ENHANCING BIODIVERSITY RESEARCH WITH ICT

As part of the natural environment, biodiversity has been defined as "the variety of life in all its forms, levels and interactions. It includes ecosystem, species and genetic diversity" [Hunter, 1996]. For our research, we focus on populations of specimens between the taxa[4] and organ levels of biodiversity research (Figure 1), i.e. *Systematics*. It is the scientific discipline that deals with listing, describing, naming, classifying and identifying living organisms. The originality of our insular tropical biodiversity management method is that we concentrate on natural *objects* that are specimens in the fields (living specimens) and in museums (collection specimens). Experts in biology at the University have studied them intimately for years and are the only persons able to recognize their names that give access to more information. They build their personal or tacit knowledge [Polanyi, 1962] by observing them in the fields and in their laboratories under the microscope, then interpreting them with descriptions. These described objects form the development basis of their formal or explicit knowledge in monographs that makes authority in their domain of speciality. But this expertise that is based on interpretation of objects through a description process is becoming extremely rare because experts are retiring without being replaced. For future biodiversity studies relying on species identification, decision-makers, environmental technicians and the general public will only be left with monographic descriptions and collections in museums. Nevertheless, with all the possibilities that the Internet has to offer, research is no longer confined to books dedicated to specialists. New innovative research can easily reach a wide audience, even in developing countries - at the click of a mouse, by offering online biodiversity information. In this context of knowledge sharing, the role of Information and Communication Technologies (ICT), and in particular Artificial Intelligence Techniques, is to give access to different kinds of online information that are structured through *data and knowledge bases* and can exchange information using *Web Services*. Our research work contributes to the vast effort of environmental conservation by introducing *knowledge management with ICT* at the beginning of the process.

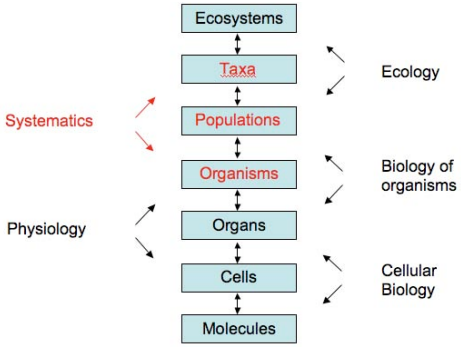


Figure 1. Different levels of biodiversity research (taken from [Lebbe, 1995])

3. THE ETIC PROGRAM

ETIC[5] is a publicly funded project, based on La Réunion Island in the South-West of Indian Ocean, whose goal is to develop innovative ideas and ICT solutions for the management of biodiversity research contents. The program was created in 2004 at University of La Réunion Island for research and knowledge enhancement of Insular Tropical Environments, by using AI techniques such as *Knowledge Engineering* for building expert systems and *Collective Intelligence* for building multi agent systems, and *Information and Communication* tools such as content management systems [Conruyt *et al*, 2006a]. Indeed, the first step to protect our insular tropical environment is to better educate citizens about its richness because we can only protect what we know! ETIC is based on several thematic projects and

[3] [http://ec.europa.eu/research/leaflets/biodiversity/index_en.html, visited on 11/04/08]

[4] Taxa are the names of ranks in the scientific classification: Species, Genus, Family, Order, Class …

[5] [http://etic.univ-reunion.fr, visited on 11/04/08]

a collaborative methodology, stressing partnerships between researchers, educators, decision-makers, enterprises, associations and end-users who wish to share and communicate their environmental data and knowledge off and on line. With the help of computer scientists, web designers, programmers and graphics experts, the common goal is to participate in the construction of an *Information System* (IS) for environmental management on the Internet. Content include terrestrial (i.e. Herbarium) and marine (i.e. Corals) biodiversity descriptions about specimens, their geography, ecology, photography, taxonomy and bibliography contextual information in La Réunion Island in the South-West Indian Ocean.

4. THE BIODIVERSITY MANAGEMENT MODULES

The core of the ETIC IS platform is the integration of knowledge bases and databases about biodiversity knowledge and data by using Web Services.

4. 1 Knowledge bases

The former applications are instances of a Knowledge Based Management Tool called IKBS. This Iterative Knowledge Base System [Conruyt and Grosser, 2003] lets specialists define an Object-Attribute-Value *descriptive model* of the domain knowledge (input) and describe *cases* (output) based on this ontology (Figure 2). This knowledge acquisition phase can be repetitive because IKBS applies the scientific method of Popper (1973) in biology (conjecture and test) with an iterative process of knowledge management: 1) Observe and familiarize oneself, 2) Represent observations, i.e. make a descriptive model and related descriptions (cases), 3) Build hypotheses from pre-classified descriptions, i.e. generate identification keys (supervised classification), 4) Test and use them with new observations, i.e. identify new specimens, 5) Refine the initial knowledge (new characters, cases and classifications) (Figure 3).

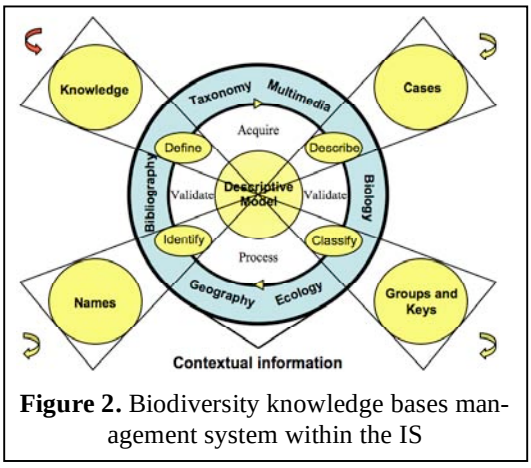


Figure 2. Biodiversity knowledge bases management system within the IS

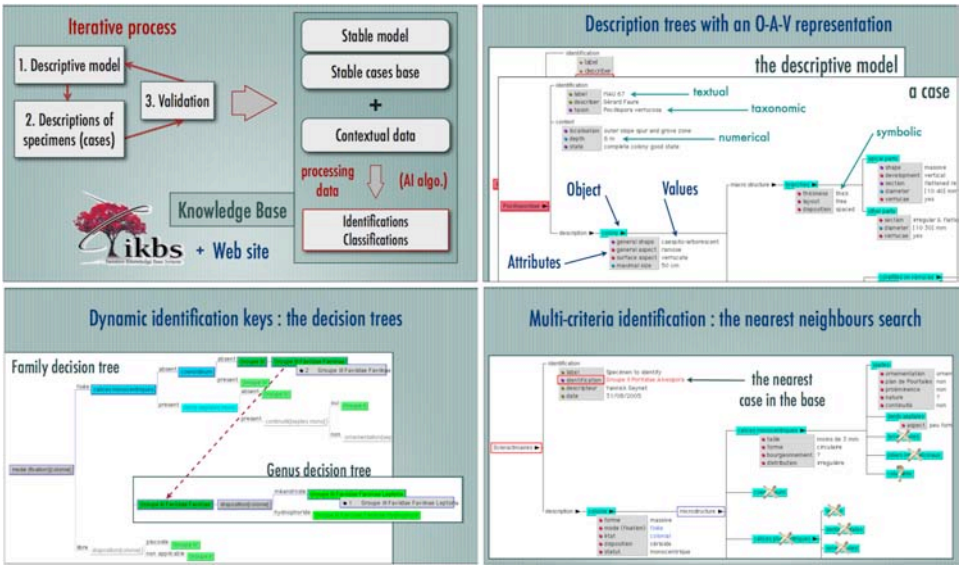


Figure 3. IKBS methodology of Systematics knowledge management

After an automatic classification process based on tree induction of pre-classified cases, end users are able to identify new descriptions with a questionnaire. End-users proceed by photo-interpretation of specimens to obtain a genus name, or by observing microscopic specimen elements under the binocular to identify a species name [Conruyt *et al*, 2006a].

4. 2 Databases

The latter is a *biodiversity module*, i.e. a database of *objects* that stores, organizes and presents scientific data about field observations, collected specimens (samples) and taxa descriptions. Other database modules (i.e. directory, multimedia, thesaurus and cartography) complement this central module: 1) the directory gives access to the *subjects*, i.e. the individual and community researchers with their profiles in a card index, 2) the multimedia database manages all types of *documents* (photo, video, sound, etc.) that can be indexed to specimen objects, 3) the thesaurus will be an illustrated glossary that stores the meaning of Insular Tropical Environment *vocabulary*, 4) the cartography is a tool for *georeferencing* data on a map (made with the GoogleMaps API). All of these modules are linked by Web Services so as to constitute a modular, interoperable and integrated biodiversity specimen and species database management system (Figure 4). The data entry process in the biodiversity module is organized around the memorization of specimens' information, which is collected in notebooks by biologists when they are inventorying biodiversity. It has been structured in five edition tasks (actions) that fit with the daily work of monitoring specimens in the fields: 1) *Origin* of specimen, where was it found? 2) Short *description*, what was observed? 3) Taxon *identification*, what is it? 4) *Status* of specimen, i.e. sex, nature, state, fertility, development stage, 5) *Label* of specimen if it is to be put in collection. In addition to this *internal specimen information* in the biodiversity module, the surrounding modules manage *external contextual data* such as the location of the specimen (space identification with the Geolocation system), the identity of the subject (who is the observer in the Directory), the image or video of the specimen with associated metadata in the Multimedia Database (Figure 5).

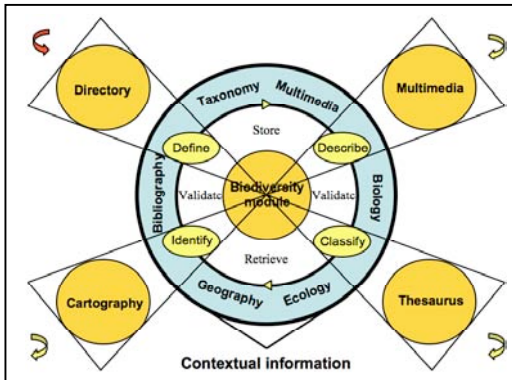


Figure 4. Biodiversity database management system within the IS

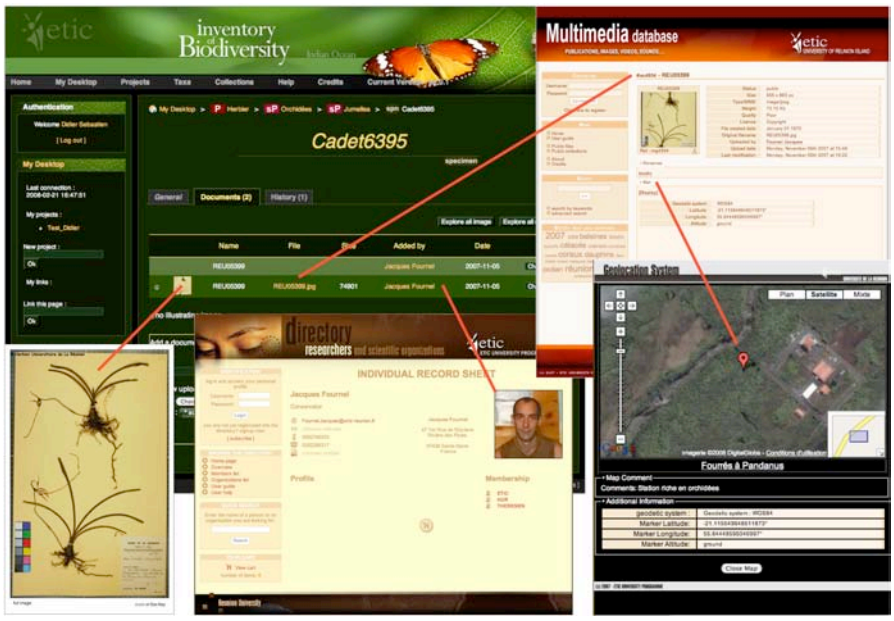


Figure 5. ETIC process of biodiversity data management

5. WEB SERVICES AS SUPPORT FOR INTEGRATION AND OPENESS

In order to make our IS modular and interoperable, we have chosen to design it according to Services Oriented Architecture (SOA). As it is a web platform, we chose the Web Services approach [Curbera *et al.*, 2001], which provides a number of features and benefits.

Among the technologies to implement Web Services, we chose the WS-* technology whose specifications are based on SOAP and WSDL standards:

- SOAP (Simple Object Access Protocol) for exchanging messages is a RPC (Remote Procedure Call) protocol built on object-oriented XML.

- WSDL (Web Service Description Language) for describing: Web Services, their operations, messages used, the types of used data, and the used protocols. The WSDL describes a public interface for access to a Web Service. This description is written in XML and indicates "how to use the service."

5. 1 Advantages of using Web Services in ETIC IS

Using Web Services structured the way we developed ETIC IS:

- Firstly, we have chosen this technology because it allows other IS to use our modules independently [Pires *et al.*, 2002]. This way, information stored in ETIC IS is fully open and exploitable by other institutions [Yang *et al.*, 2002].
- Secondly, it modified the way we imagine the connections in the IS itself. It let us develop each module in a heterogeneous way, using the appropriate technology (PHP/JAVA/Flash Action Script). For example, the ETIC Directory is developed in PHP whereas the first version of our Multimedia Database [Sebastien and Conruyt, 2008] was relying on the JAVA technology. Although we used different programming languages, these two modules were able to exchange secured information, thanks to Web Services.
- Thirdly, from a technical point of view, Web Services impose a rigour that permits to update our systems and services without losing the compatibility with older versions. Each new version of a service is indexed with a number, and the WSDL provides the way to implement the new client, but it does not impact on the older versions that remain fully functional.
- Fourthly, from a management point of view, using Web Services was a good choice. Indeed, the ETIC development team often changed. Because of this important turn-over, it was difficult to transmit the development key points from an engineer to his successor. Thanks to the Web Services (and particularly the WSDL declaration), it was not necessary for the new team to fully understand a service source code to use it as a client. It saved a lot of development time.

Thus, using Web Services was of course important to open our IS, but it also helped us to improve its inner structure.

5. 2 What are the ETIC services?

Because each module is dedicated to accomplish a precise task, each of them provides some specific functionalities and dedicated Web Services. This way, any other IS can reach ours and decide to use one or the whole of ETIC modules for its own purpose. But contrary to data consultation that is completely open, these clients need to be referenced in ETIC directory if they want to add information in ETIC IS. This security allows us to ensure the traceability of information injected in our system.

The implemented connections between ETIC modules are shown in Table 1. On the five modules deployed now (Directory, Multimedia, Cartography, Biodiversity, IKBS) all use Web Services as client, but only three provide services (Table 2).

The work is still in progress. For example, the development of the Thesaurus (one of the database's modules, see Figure 4) is not presented here. Furthermore, the current version of the Cartographic Browser does not implement a connection to the ETIC directory. In fact, it cannot be used as a standalone module yet. Other modules asking for its services ensure

themselves the client identification process. Moreover, several functionalities supported by the modules don't have a Web Service allowing a distant interrogation of it yet (e.g. to use remotely the multi criteria search of the MDB, see Table 2 below). Only main services are already released. Of course, this will change with the development of our modular IS.

Table 1. Connections between ETIC modules

Web Services		**Client**				
		Directory	Multimedia	Cartography	Biodiversity	IKBS
Provider	Directory		✓	○	✓	○
	Multimedia	○		✗	✓	✓
	Cartography	○	✓		✓	○
	Biodiversity	✗	✗	✗		✗
	IKBS	✗	✗	✗	✗	

Legend: ✓: already implemented, ○: in progress, ✗: not relevant now

Table 2. Services already functional in ETIC IS

Module	**Tasks provided by Web Service**
Directory	▪ Find a user by his name ▪ Get the list of members' id ▪ Get a minimum of information on each member of the directory ▪ Get a minimum of information on a specified member ▪ Get all information about a specified member ▪ Authenticate a member with its login/password (secured service)
Multimedia Database (MDB)	▪ Return two links (documents' thumbnail and mini-thumbnail) corresponding to the input array of document id ▪ Return all information (metadata, download link) corresponding to the input array of documents' id ▪ Return two links (to the information card on the MDB and to the file) corresponding to the input array of documents' id ▪ Allow the distant upload of a document on the MDB ▪ Find a document on the MDB (Quick Search)
Cartographic Browser	▪ Verify if a map exists for a specified id ▪ Get all information about specified map ▪ Find a map from its description ▪ Delete maps from the given list

Independently of the fact that we have to develop new Web Services for each module, we also have to increase the integration of the existing services *between* modules. This integration must be the result of focus groups realised with biologists in order to provide them the right information when they need it. This tuning is a Web Design concern because, for end-users, Web Services must be transparent.

6. INTEGRATION OF MODULES

The usefulness of integrating both data and knowledge bases by using Web Services is to make them interoperable for better information (data + knowledge) management between research scientists before Web publication (authentication with e-research or on-line collaborative research), and then facilitate the sharing of their data and knowledge with more general public on the Internet (education with e-learning) [Conruyt *et al.*, 2002]. Our ETIC platform gathers tools for helping decision makers to manage the biodiversity of natural parks (see Figure 6 for the Marine park example). The process starts with Knowledge Bases that are used for the process of describing specimens, classifying them and identifying a name for unknown specimens. Databases give then access to different types of ecological information on known taxa. Indeed, the first question is to recognize the name of what has been observed to make a biodiversity diagnosis. Then, in order to answer questions about the evolution of ecosystems, the next step will be to monitor the dynamics of populations with temporal databases on coral reef (or forest) and simulate their behaviour with Multi-Agent Systems when facing damages such as pollution, pests, over-fishing, etc.

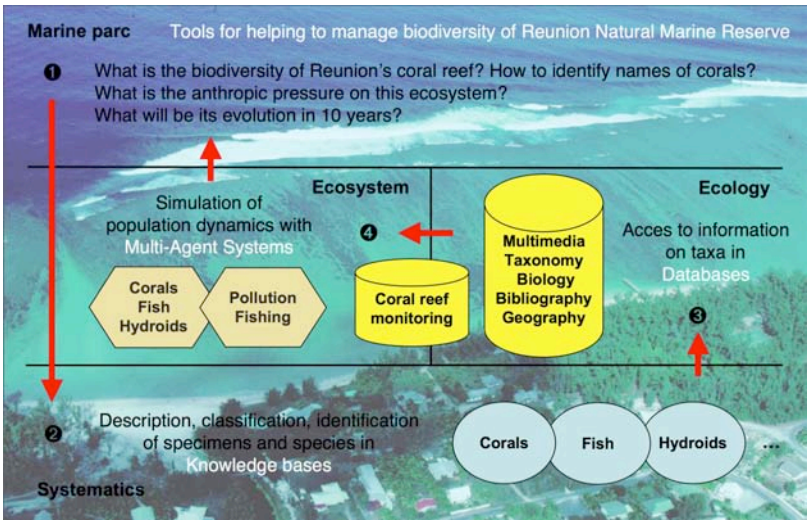


Figure 6. Coral reef biodiversity management process

Web Services have been designed currently to collect the information about actors of the ETIC community (the Directory[6]). Some persons are identified as project leaders in their scientific domains. As they want to publish their data and knowledge on biodiversity with their colleagues, they create their proper organization with authorized members in the directory. They can then share their documents in the multimedia database[7], edit them together, and also localize them with a geo-referenced system using Google Maps. But their objective is to store information about taxa and specimens in the fields before putting them in collections. The Biodiversity module[8] is thus used as the core Web Service to inventory species they are working on.

On a technological level, our approach of integration has been clearly made in the *delegation* meaning. When a functional need is identified during the module design, we find out if the missing functionality must be implemented by adding code in the module or as an interaction with another existing module or a new module. This way allows us to easily identify all modules to be deployed, and interactions between modules are explained as the system is built. It is an efficient process when such project is started from an empty module set.

But the next challenge is to apply the opposite approach: *data aggregation*. We have to ask ourselves about the potential benefits that can emerge from new composition and fusion of knowledge from different modules. This is precisely the topic of our future work, through which we hope to formalize a cross ontology that will serve as a basis for analytical tools (including Multi-Agent Systems simulations) in order to improve our understanding of data, and perhaps produce new knowledge.

7. CONCLUSION

In the domain of life sciences, engineers and designers who build information systems have to model expert knowledge at two skill levels. On the one hand, they store their explicit knowledge in databases, because it is made of contextual data (eco-biological, biogeographical, photographical, biographical, taxonomical, etc.) that is found in books. This information can be retrieved more easily in databases with a multi-criteria research. On the other hand, they acquire expert implicit or tacit knowledge in knowledge bases, because this know-how is more complex to elicit: this experimental and intuitive expertise is grounded in their daily work throughout a life practised in the fields. It is better to process

6 [http://etic.univ-reunion.fr/annuaire/, visited on 11/04/08]

7 [http://etic.univ-reunion.fr/bdm/, visited on 11/04/08]

8 [http://etic.univ-reunion.fr/biodiv/, visited on 11/04/08]

case descriptions with induction techniques than to tell the expert to explain what are his rules of thumb. All these interoperable modules are prototypes that are still under development. They have been tested using the Herbarium boards and monographs of Reunion University [Cadet, 1980] and the coral specimens collection and taxa descriptions of the Mascarene Islands [Faure, 1982]. In order to improve the usability of these tools and anticipate biologists' needs, we have used a co-design platform methodology with several focus groups [Conruyt et al., 2006b]. The next step is to deliver a fully operational IS to naturalists who are working in natural parks in Reunion Island and to be interoperable with the "Système d'Information sur la Nature et les Paysages" (www.naturefrance.fr) at a national level and the Global Biodiversity Information Facility Network (www.gbif.org) at an international level.

ACKNOWLEDGEMENTS

The ETIC program is financed by the French National Government, the Regional Council of Reunion and the European Union (http://www.reunioneurope.org/docup/A9-04-CI.html).

REFERENCES

Cadet, T., La végétation de l'île de La Réunion. *Thèse es sciences*, Université de La Réunion. 312 p, 1980.

Conruyt, N., Geynet, Y., Grosser, D., Santally, M., Senteni, A., Sonah, S., The IKBS-Darwin Research & Learning System for Sharing Biodiversity Knowledge of Tropical Islands. *Proc. of the International Conference on Computers in Education*, pp. 760-764, Auckland, New Zealand, 2002.

Conruyt, N., Grosser, D., Knowledge engineering in environmental sciences with IKBS. *AI Communications*, 16(4), IOS Press, 267-278, 2003.

Conruyt, N. Grosser, D., Geynet, Y., Enhancing Insular Tropical Environment contents through ICT and a Co-design methodology: the case of the ETIC program, *Proc. of the 5th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence* (BESAI'06), 1-1–1-7, Riva del Garda, Italy, 2006a.

Conruyt, N., Conruyt, P., Sebastien, O., A methodology for Designing E-services from a Co-Design Platform, *Proc. of the 6th International Conference on Knowledge Management* (IKNOW'06), Graz, Austria, September 6-8, 2006, K. Tochtermann / H. Maurer (Eds.), pp. 226-230, ISSN 0948-695x, J.UCS, Graz, Austria, 2006b.

Curbera, F., Mukhi, M., Weerawarana, S., On the Emergence of a Web Services Component Model, *Proceedings of the 6th International Workshop on Component Oriented Programming*, (WCOP'01), 2001.

Faure, G., Recherche sur les peuplements de scléractiniaires des récifs coralliens des Mascareignes. *Thèse es sciences*, Université Aix-Marseille II, 1982.

Lebbe, J., Systématique et informatique. Systématique et biodiversité, Bourgoin T. (Ed), *Biosystema*, Paris, 13:71-79, 1995.

Hunter, M.L., Jr. Fundamentals of conservation biology. Blackwell Science, Cambridge, Massachusetts, 482pp., 1996.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B., Kent, J. Biodiversity Hotspots for Conservation Priorities, *Nature*, 403:853–858, 2000.

Pires, P. F., Benevides, M. R. F., Mattoso, M., Building Reliable Web Services Compositions, *Proc. of the International Conference on Web Databases and Web Services*, LNCS, vol. 2593, pages 59-72, Springer-Verlag, 2002.

Polanyi, M., Personal Knowledge: Towards a Post-Critical Philosophy; Chicago: The University of Chicago Press, 1962.

Popper, K.R., La logique de la découverte scientifique. Payot (Eds.) Press, Paris, 1973.

Stepath, C.M. Awareness and community-based monitoring, *Proc. of the 9th International Coral Reef Symposium*, Vol:2. Bali, 807-812, 2002.

Sebastien, D., Conruyt, N., Online multimedia database for communities of practice in Biology: a real use challenge, *Proc. of the 3rd International Conference on Internet and Web Applications and Services*, 2008 (to appear).

Yang, J.,. Papazoglou, M.P., Web Component: A Substrate for Web Service Reuse and Composition , *Proc. of the 14th International Conference on Advanced Information Systems Engineering* (CAISE'02), Toronto, Canada, 2002.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

An Intelligent Decision Support System for Environmental Impact Assessment of Industrial Projects

Shady Aly[a], **Ivan Vrana**[b]

[a]*Assistant Professor, Helwan University, Cairo, Egypt* (*shadyref@yahoo.com*)

[b]*Professor, Czech University of Life Sciences, Prague, Czech Republic(vrana@pef.czu.cz)*

Abstract: Environmental Impact Assessment (EIA) of industrial projects is a typically ill-structured, multi-aspect, multi-criterion decision making problem, in which usually not all input relevant factors are known, and the relationships between these factors and the direct impact on environmental elements can not be stated or formulated exactly. The assessment can have several dimensions like air pollution impact, water pollution impact, etc. Each dimension has its own set of homogenous variables. This paper presents a modular decision making solution for treating the complex ill-structured EIA decision problem. The proposed solution is an intelligent environmental decision support system (EDSS) that makes use of human expertise's and fuzzy logics in assessing the environmental impacts of the proposed or existing industrial projects. It consists of multiple fuzzy expert systems (FESs), each of which contains the homogenous knowledge's and expertise's relevant to a one environmental dimension. Simple and efficient heuristics are introduced for integrating FESs within the proposed EDSS.

Keywords: Environmental Impact Assessment; Environmental Decision Support; Fuzzy Expert Systems; Analytical Hierarchy Process; Binary Group Decision-Making.

1 INTRODUCTION

All industrial and development projects can affect their surroundings. An EIA is an assessment of the likely positive and/or negative influence a project proposal may have on the environment. The problem typically involves: huge quantities of data to manipulate, low quality of data (uncertainty, measurement errors, missing data), different spatial and temporal scales, dynamic and stochastic behavior, and being at the crossroad among many disciplines/domains, and so many qualitative or subjective factors. In fact, the review of literature has demonstrated that the development of environmental decision support systems (EDSS) is rapidly progressing [Matthies, 2007]. Some of the reported literature concentrated on the role of Artificial Intelligence like [Cortès et al., 2000; Sokolova & Fernandez, 2008]. Recent literatures has revealed the focus on Group Decision Making and Consensus [Liu and Wirtz, 2006; Turon et al., 2007], and the utilization of fuzzy logics [Liu and Wirtz, 2006 ; Nasiri and Huang, 2008] as appropriate to the nature of EIA decision problems. Matthies et al. [Matthies et al., 2007] discussed the background of recent developments in EDSS and summarized a selected set of papers that were presented at the 2nd Biennial Conference of the International Society of Environmental Modeling and Software (IEMSS 2004). They pointed out that there is a general tendency toward better utilization of interdisciplinary data, integration and visualization of temporal and spatial results.

The most common approach of EIA is Leopold matrix [Leopold et al., 1971]. The system consists of a matrix with columns representing the various activities of the project, and rows representing the various environmental factors to be considered. The intersections are filled in to indicate the magnitude (from -10 to +10) and the importance (from 1 to 10) of the impact of each activity on each environmental factor. However, Gilberte White [White, 1972] pinpointed several limitations of the matix. He stated that Leopold matrix contains no provision for indicating uncertainty resulting from inadequate data or knowledge. He also pointed out that the Leopold matrix does not provide explicit criteria for assigning numerical values to weights of impacts. Moreover, the synthesis of the predictions into aggregate indices is not possible. Therefore, given the limitations of Leopold Matrix, the EDSS proposed in this paper will enable elimination of such limitations, first through employing fuzzy logics for handling uncertainty. Second, it consistently enables for inferring the environmental impacts of the project's activities with respect to various environmental factors through decision logics representing the knowledge and expertise's human environmental experts, and finally enables synthesizing the main impact magnitudes corresponding to the main environmental factors into an overall impact, which was not possible with Leopold matrix approach. It should be noted that there is a basic difference between EIA and the LCA (ISO 14042) methodologies in that the former is a framework for identifying, predicting, evaluating, and mitigating the biophysical, social, and other effects of proposed projects or plans and physical activities, whereas as the later is a method of accounting the environmental impacts of a product, service, or process over the course of its life cycle from extraction of materials to disposal or reuse of the final product.

This paper is organized as follows. Section 2 presents the proposed modular and intelligent EDSS. In section 3, the basic numerical scale for judging environmental impact magnitudes is established, and the Analytical Hierarchy Process (AHP) is proposed as a method for computing weights of environmental factors and systems. Section 4 explains the decision procedure using the proposed EDSS and gives illustrative examples.

2 THE PROPOSED EDSS FOR EIA

2.1 The Rationale for Modular and Collective EDSS

Turon et al. [Turon et al., 2007] stated that the judgments of many will usually prove superior to the judgment of one. One kind of critical decision making problems is the binary or Yes/No decision making. The classification of environmental impact of an industrial project into either "Positive" or "Negative" is one typical instance of such critical decision problems. Integrating multiple intelligent decision support or expert systems is considered particularly useful in obtaining high quality, comprehensible and reliable decision solution. Such integration is justified when a complex problem in hand has multiple aspects (e.g., the EIA) and requires multiple expertise's to cope with its ill-structuredness and ambiguity. Sometimes the integration is accomplished for sake of reliability. However, still some other reasons behind independence amongst expert systems are: cohesion of knowledge units, control and decision responsibility, avoidance of knowledge interaction and mutual influence, modularity in decision analysis and explanation, sensitivity of aggregate knowledge (sometimes), consistency in handling relationships and reasoning. Also, besides the above pragmatic reasons, building small and separate expert systems can justify for operational inefficiency usually exhibited with huge and cumbersome large-scale expert system. Now, based on the above reasons, our proposed systems will be composed of multiple fuzzy expert systems (FESs), as appropriate to the nature of EIA decision making context. Next section introduces the architecture of the proposed EDSS.

2.2 The Architecture of the Proposed EDSS

The proposed EDSS consists of multiple environmental decision making units, the FESs, as shown in figure 1. In such configuration, FESs are connected in parallel. Each FES provides its

output on whether the corresponding environmental impact is positive or negative. Each individual system accepts different sets of relevant environmental input factors. Although the idea of integrating multiple decision making systems is not new and classic, combining the outputs of multiple FESs is a new and has been first introduced in [Aly & Vrana, 2006]. Multiple criteria could be processed within the decision logic of each, or each FES by itself could represent a one judgment criterion. The proposed decision making system is well suited for two different decision making situations. In the first situation different knowledge sources (i.e., FESs) each of which corresponds to different environmental dimensions (e.g., air pollution, water pollution, etc.). The second situation involves similar knowledge sources all specialized in a one environmental issue, like air pollution, each of which holds different skills, tools, or techniques. In the first situation, the final consolidated output of the given FESs is obtained through aggregation of their final impact values, since all knowledge sources are necessary to judge the overall environmental impact of the proposal, whereas, in the second case, a simple consensus heuristic is to be used for combining FESs' outputs, because the inclusion of multiple knowledge sources aims to enhance the reliability, specially for such ill-structured and ambiguous decision problems. Next section explains in brief the internal components of each FESs.

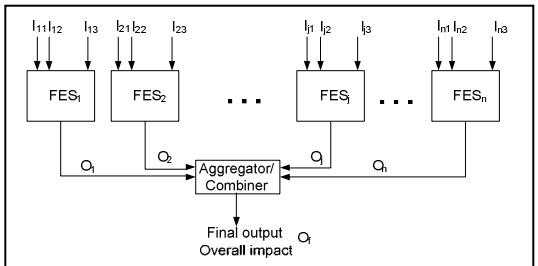


Figure 1. The architecture of the proposed EDSS.

2.3 The Internal Structure and Processing of Modular Decision Making Units

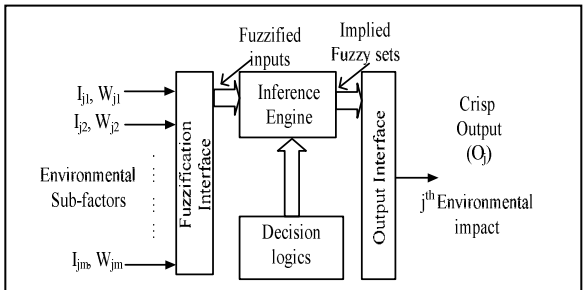


Figure 2. The internal structure of environmental FESs.

As described previously, the proposed EDSS consists of multiple FESs. Every FES processes the influence of a subgroup of input environmental factors, with respect to one main environmental issue like Air pollution, or soil contamination, on environmental impact. The standard FES (see figure 2) consists of four components: a fuzzification subsystem, a knowledge-base, an inference mechanism, and a defuzzification subsystem. The fuzzification subsystem converts the values of the quantitative and qualitative environmental sub-factors into fuzzy sets. The knowledge base contains the decision making logics containing the experts' knowledge about how to convert the values of sub-factors into an environmental impact regarding one environmental issue of main factor. However, it could contain an empirical knowledge as well. The inference mechanism match the current values of input factors with the set of applicable rules, and then infer the implied fuzzy set of each rule. Finally, the defuzzification subsystem converts the implied fuzzy sets into crisp value expressing the Positive/Negative alternatives. The special concern put on FESs in general is attributed to its wide applicability and use due to its capability to treat vagueness, and subjectivity. More description of the FES can be found in [Kilagiz et al., 2004]. Next section explains how to quantify both environmental impacts' magnitudes and importance's of FESs.

3 THE JUDGMENT SCALE AND THE IMPORTANCE'S OF ENVIRONMENTAL FACTORS AND IMPACTS

The proposed scale (figure 3) for quantifying the environmetal impact outputs of the indivdual FESs is same like that of the Leopold matrix. The magnitudes of environmetal impacts is quantified within the range [-10, +10], where -10 indicates sharp "Negative" impact, +10 indicates sharp "Positive" impact, and 0 means "Non-biased" value. Intermediate values reflect the bias toward either two options. The same scale is to be used as the universe of discourse for the implied or output fuzzy sets within each FESs. Generally, positive values of impact indicate or signify benefits caused by the assessed project, whereas negative values indicate harm caused by the project. On the other hand, in order to reflect the relative importance of every FES in judging the environmental impact, and the importance's of the environmental factors as well, the most commonly used decision-aiding tool, the Analytical Hierarchy Process (AHP) [Saaty, 1980] is to be used to compute the absolute weights of different environmental factor, and the weights of the individual FESs, too. In AHP, the decision maker carries out simple pair-wise comparative judgments, which are then used to develop overall priorities for ranking alternatives, factors or criteria. These priorities or weights are normalized within the computations of the AHP. The reader may refer to [Saaty, 1980] for more details about AHP and its computational procedure. Next section will describe the basic mechanism for integrating the FESS within the proposed intelligent EDSS.

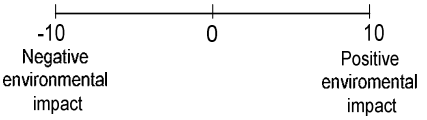


Figure 3. The numerical scale for quantifying environmental impacts.

4 THE DECISION MAKING PROCEDURE

There are two decision making situations to be treated in this article. In the first decision making situation, several different knowledge sources or expertise's (FESs) each of which holds different knowledge about different environmental dimensions, integrated for sake of comprehensibility. Consequently their individual environmental output impacts should be aggregated. On the other hand, in the second situation, we have several knowledge sources (FESs) having same or equal knowledge but different tools, skills or techniques, integrated for sake of enhancing the reliability of the final decision solution or overall impact. For more explanation of difference between knowledge's combination and aggregation, the reader may refer to [Aly & Vrana, 2006]. It should be noted that the utilization of evidence theory combination rule could be exploited but we well confront a problem of estimating the aggregate belief in the knowledge contained in each FESs. Next two subsections will introduce a decision making procedures for both cases.

4.1 A Heuristic for Aggregating the Impacts of Different Environmental Knowledge Sources

Here, final overall environmental impact is obtained through aggregating the environmental impact of the individual FESs corresponding or pertaining to each environmental issue using the weighted average, and taking into account the weights of each FESs computed via the AHP. The aggregation is formally stated as follows:

Let O_j : the output impact for the j^{th} FES.

W_j : the weight of the j^{th} FES.

O_t^- : a threshold value for negative environmental impact. This threshold value is optional and could be set according to the opinion of field experts and decision analysts.

Step 1: Compute the weights of FESs using AHP

Step 2: Check the negative output impacts against the threshold (O_t^-):

$\forall j$ IF $O_j \leq O_t^-$ THEN $O_f = O_t^-$; reject project; otherwise go to step 3.

This means that if the output impact of any FES exceeds the negative threshold, then the project should be rejected without any concern to other outputs. Other decision policy could be established based on the opinions of field experts and decision analysts.

Step 3: Establish a total numerical scale from within an arbitrary range from –S to +S value, to represent the decisive degree between "Positive" and "Negative" decisions. The value -S corresponds to "Negative", and the value +S corresponds to "Positive". The middle value of such total scale is zero.

Step 4: Apportion the total numerical scale established in the previous step into smaller numerical scales allocated to every FES in proportion to its computed weights, as follows:-

$$S_j = w_j * 2S \qquad \forall j \tag{1}$$

Where,

S_j: is the total range for output scale of the j^{th} FES.

Then, the crisp output of each j^{th} FES should be produced within the allocated numerical scale, from 0 to S_j. Proportionality could be used to convert outputs within the range [-10, +10] into the partitioned scales, as another options.

Step 5: Given the crisp output of each FES, aggregate expertise's by summing all crisp outputs to given the finally aggregated group output, O_f (eq. 2):

$$O_f = \sum_{i=1}^{n} Oj \tag{2}$$

Then, the final output is judged as either "Positive" or "Negative" as follows:-

If $O_f > 0$, then the final group decision is "Positive"; accept (continue) project.

If $O_f < 0$, then the final group decision is "Negative"; reject (stop) project.

If $O_f = 0$, then the final group decision is "Non-biased"; optional, up to experts and decision analysts opinion. Stop.

Example 1:

Suppose that given a proposed industrial project involving some mechanical industrial operations and chemical process. The project outputs its chemical waste to a river, whose water is used in irrigation. Before actual construction and operation phases, it is required to assess the overall environmental impact of the project's industrial activities on the surrounding environment on several dimensions that leads finally to a decision of whether to undertake the project or not. There are four environmental aspects to consider: air pollution, noise/vibration, water pollution and soil contamination impacts. Every environmental aspect includes several input factors, may be quantitative or qualitative. Some of these factors could have uncertain or vague values. Moreover, due to ill-structuredness, the relationship between these inputs factors and the direct impact on the elements of the environment is not fully grasped. Accordingly, our proposed decision making systems here is appropriate. Therefore, four FESs corresponding to the four environmental aspects could be constructed; namely, air pollution expert system, noise/vibration expert system, water pollution expert system and soil contamination expert system. Each FES processes a homogenous group of input factors or variables as relevant to one environmental aspect. Every system contains the expertise's, and knowledge of several experts specialized in certain environmental aspect. Let us suppose that the output impacts and weights of each FES were as follows:

- Air Pollution Expert System (FES_1) $O_1 = -8$, $W_1 = 0.51$
- Noise/vibration Expert System (FES_2) $O_2 = 0$, $W_2 = 0.22$
- Water Pollution Expert System (FES_3) $O_3 = -2$, $W_3 = 0.26$
- Soil Contamination Expert System (FES_4) $O_4 = +6$, $W_4 = 0.01$

Now applying the above procedure:

Threshold check in step 2 fails. Then, we proceed to step 3. Now, we have the outputs values of impacts all evaluated within [-10, +10]. Applying step 3 we establish a total scale arbitrarily selected as the sum of indivdual scale, [-40, +40]. Here, S equal 40, then we partition this total scales into four smaller ones to determine the judgmental quota or space for each environmental dimension according to its weight. Using proportioality, the total ranges for allocated scales are: 40.8, 17.6, 20.8 and 0.8, for FES_1, FES_2 , FES_3 and FES_4 respectively, and converted outputs becomes as follows: -16.32, 0, -2.08, +0.24. The aggregation of these output using equation 2 gives O_f = - 18.16 < 0, thus the final decision is "Negative"; reject (stop) project. It should be noted that a threshold could be established also on the value of the O_f, and thus if the negative impacts is less than that threshold, only the project is rejected.

4.2 A Consensus Heuristic for Combining the Impacts of Knowledge-equal Sources

Sometimes the outcomes and performance of the proposed project is affected by the occurrence of some uncertain future events. In such situation, it is adequate to construct multiple FESs, each of which specialized in same areas, but utilize different tools or techniques. The final overall impact is based on combining the decision outputs of the participating expert systems rather than aggregating them, since each FES could solely perform the decision making task, but they are accumulated for reliability. Below, we propose a simple consensus-based heuristic that makes the final decision. In addition to the notations of the outputs and weights used in previous heuristic, Let two simple indicators defined as follows:-

- The Percentage of Class Voting's: the fraction of voting's given to certain class (Positive, Negative or Non-biased) = Number of Class voting's / Total number voting's.
- Sum of weights of voting's: the sum of weights of voting's given to a class

Based on the above definitions, we have six magnitudes: Percentage "Positive" voting's (*%PV*), Percentage "Negative" voting *(%NV*), Percentage "Non-biased" voting (*%NBV*), Sum of weights of "Positive" voting's (*SWPV*), Sum of weights of "Negative" voting's (*SWNV*) and Sum of weights of "Non-biased" voting's (*SWNBV*). Thus, High value of (*%PV*) means that there is a considerably high degree of consensus or voting's dominance assigned to "Positive Impact" decision option, and so on. Also, similarly, high value of SWPV means that there is a considerably high degree of weight dominance level that imposes undertaking the "Positive" direction, and so on.

Now, the steps of the aggregation heuristic are as follows:

Step 1: Compute weights of Individual FESs using AHP (W_i)

Step 2: Check the extreme values of outputs impact:
$\forall j$ IF O_j =10 THEN O_f = 10; accept project; Stop.
$\forall j$ IF O_j =- 10 THEN O_f = -10; reject project; Stop.

The rationale for rules above is that any expert system whose expertise exhibits complete bias to any one of the two alternatives, and then we should adhere to such emphasized opinion of such recognized expertise.

Step 3: Attributing outputs:

Every FES's impact judgment, O_i, is attributed to one of three classes or consensus sub-group depending on whether or not this value is above, below or at the zero:

Condition 3.1: IF $O_i > 0$, THEN O_i is attributed to "Positive" class.
Condition 3.2: IF $O_i < 0$, THEN O_i is attributed to "Negative" class.
Condition 3.3: IF $O_i = 0$, THEN O_i is attributed to "Non-biased" class.

Step 4: Preliminary checking**:** the heuristic is to be terminated because of either high voting's or weights dominance levels under the following two conditions (otherwise, go to step 3):-

Condition 4.1: IF max{*%PV, %NV, %NBV*} ≥ 75 %, THEN there is a high degree of voting's dominance level and is given by the class argument of max{*%PV, %NV, %NBV*}(arg max{*%PV, %PV, %NBV*}). Stop. Otherwise Go to step 5

Condition 4.2: IF max {*SWPV, SWNV, SWNBV*} ≥ 0.75, THEN a high degree of weighting dominance level and the DCA is given by the class argument of max{*SWPV, SWNV, SWNBV*}(arg max {*SWPV, SWNV, SWNBV*}). Stop. Otherwise Go to step 5

Note: arg max {*SWPV, SWNV, SWNBV*}or arg max{*%PV, %NV, %NBV*}gives the decision class which has either maximum voting's or weights dominance levels respectively.

Step 5: Compute the Weighted Arithmetic Mean of Impacts (WAMI) for the given output impacts:

$$WAM\ I = \sum O_j\ W_j \qquad (3)$$

Step 6: Interpret the value of AM into an overall positive or negative impact as follows:

Condition 4.1: IF the *WAMI* < 0 (negative overall impact), then project must be rejected or stopped Stop.

Condition 4.2: IF the *WAMI* ≥ 0, then project can be undertaken or continued. Stop.

Example 2

Suppose that it is required to evaluate whether an existing nuclear power plant will have a positive or negative impact on the Ozone depletion after 25 years from now. It is obvious that the decision answer is influenced by some uncertain future events and outcomes, pertaining to both elements of environments and the operational performance of the plant. Here, again it is difficult to exactly state all input factors or formulate their relationships with impacts after 25 years. Therefore, one effective solution is to rely on the multiple expertise's in order to cope with such ambiguity to enhance reliability. Again, our proposed system of multiple FESs fits here. Suppose that relevant to the industrial context and considered environmental dimension, a set of five expertise's factors and techniques specialized in Ozone depletion environmental issue. So, we construct five corresponding FESs. Every FES includes the expertise's and tools of subset of environmental experts or scientists, and manipulate its own set of relevant input factors. Then, every FES outputs its judgment, within -10 to +10, about whether the impact of the project after 25 years will be positive or negative. Now, suppose that the weights computed via AHP, and the output judgments of the systems were as follows:-

- FES_1 $O_1 = 0$, $W_1 = 0.2$
- FES_2 $O_2 = -2$, $W_2 = 0.22$
- FES_3 $O_3 = 0$, $W_3 = 0.32$
- FES_4 $O_4 = 6$, $W_4 = 0.01$
- FES_5 $O_5 = -9$, $W_5 = 0.25$

Now, the steps of the consensus heuristic are as follows: Extreme values check in step 2 fails since $O_j \neq \pm 10$ for $\forall\ j$. step 3 attributes output impacts as follows:
Positive class = $\{O_4, O_5\}$, Negative class = $\{O_2\}$, Non-biased class = $\{O_1, O_3\}$.
Application of step 4 does not reveal any voting's or weights dominance levels. Then, consensus could not be detected, and the last option is averaging using equation 3:

$$WAMI = \sum O_j\ W_j = 0\ (0.2) + -2\ (0.22) + 0\ (0.32) + 6\ (0.01) + -9\ (0.25) = -2.18$$

Since the *WAMI* gives negative, the impact of the project after 25 years on the Ozone depletion is considered harmful.

5 CONCLUSION

The article has presented a modular and intelligent decision making systems for coping with ill-structuredness and uncertainty confronted in EIA. The presented approach improves decision process in this area and overcome some of the limitations of most common Leopold Matrix approach, through its ability to manipulate vagueness and uncertainty, and the ability to synthesis the impacts of the large number of factors into an overall impact in a logic and simple way. The proposed system also relies on the efficient human expertise's control, which is an unequaled solution to complex ill- structured decision problems. The system also avoids the operational limitations of having a single, huge and slow expert system. Finally, the proposed research could be exploited similarly in approximating vagueness and treating ill-structuredness within the LCA decision making.

AKNOWLEDGEMENTS

Authors wish to thank the Czech Ministry of Education for support of this work by the Grant No. MSM 6046070904 "Information and knowledge support of strategic control" and Grant No. 2C06004 "Intelligent tools for content assessment of relevance of general and specialized data and knowledge resources".

REFERENCES

Aly, S., & Vrana, I.;; Integrating Multiple Fuzzy Expert System under Restricting Requirements, Agriculture Economics, 52 (4), 2006.

Cortés U., M. Sànchez-Marrè, L. Ceccaroni, I. R-Roda and M. Poch. Artificial Intelligence and Environmental Decision Support Systems. Applied Intelligence 13(1), 77-91, 2000.

Kilagiz, Y., Baran, A., Yildiz, Z., and Cetin, M.; "A Fuzzy Diagnosis and Advice System for Optimization of Emissions and Fuel Consumption"; Expert Systems with Applications"; Vol. xx, No. xxxx, PP. 1-7, 2004.

Leopold, L. B., Clarke, F. E., Hanshaw, B. B. and Balsley, J. R., A procedure for evaluating environmental impact. Geological Survey Circular 645, Government Printing Office, Washington, D.C. 13, 1971.

Liu, X. and Wirtz, K. W., Consensus oriented fuzzified decision support for oil spill contingency management, Journal of Hazardous Materials, 134(1-3), 30 Pages 27-35, 2006.

Matthies, M., Giupponi, C. and Ostendorf, B., Environmental decision support systems: Current issues, methods and tools, Environmental Modeling & Software, 22(2), 123-127, 2007.

Nasiri, F. and Huang, G., A fuzzy decision aid model for environmental performance assessment in waste recycling, Environmental Modeling & Software, 23(6), 677-689, 2008.

Saaty, T.L, The Analytical Hierarchy Process, New York: McGraw-Hill, 1980.

Sokolova, M. V. and Fernandez, A., Modeling and implementing an agent-based environmental health impact decision support system, Expert Systems with Applications, In Press, Accepted Manuscript, 2008.

Turon, C., Comas, J., Alemany, J., Cortés, U. and Poch, Manel, Environmental decision support systems: A new approach to support the operation and maintenance of horizontal subsurface flow constructed wetlands, Ecological Engineering, 30(4), 362-372, 2007.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Intelligent Environmental Decision Support System for integrated operation of Membrane Bioreactors

I. Rodríguez-Roda[1], J. Comas[1], H. Monclús[1], J. Sipma[1], M Poch[1]
[1] *Laboratori d'Enginyeria Química i Ambiental, University de Girona. Campus Montilivi s/n, E-17071 Girona, Spain (E-mail: ignasi@lequia.udg.edu; quim@lequia.udg.edu).*

Abstract: The target of this paper is to present a prototype of Intelligent Environmental Decision Support System (IEDSS) that guarantees robust and independent monitoring and real-time control of membrane bioreactors, and at the same time allows remote supervision of the integrated operation of the biological process and the physical separation by the membranes. The rule-based and model-based systems include empirical and tacit knowledge to relate parameters of biological control to the membrane unit operation. At the same time, the IEDSS is meant to be equipped with an automatic learning module of the membranes life cycle that records and analyses accumulated experiences and allows acquired knowledge to be reused in the future whenever similar situations arise.

Keywords: IEDSS, wastewater, membrane bioreactors, fouling, integrated operation.

1. INTRODUCTION

Membrane bioreactor technology is a fast-development concept that is gaining ground in the field of wastewater treatment, mainly in small facilities that have high quality requirements for water reuse (e.g. [Judd 2005; Lesjean *et al.*, 2004; Cho *et al.*, 2004; Jiang *et al.*, 2004]). Membrane bioreactors offers several advantages over conventional activated sludge plants, e.g. operation at high biomass concentrations, reduced excess sludge production, extremely low suspended solid concentrations in the treated effluent, drastically enhanced elimination of pathogens and viruses, and a superior effluent quality. Furthermore, the high biomass concentration and long sludge retention times in membrane bioreactors positively affects the overall activity of slow growing microorganisms acting in e.g. nitrification or degradation of specific refractory pollutants and lead to a higher stability and persistence to shock loads. Although membrane bioreactors present many advantages, as in most membrane filtration processes, the permeate flux declines during filtration due to membrane fouling [Judd, 2005], which is significantly influenced by the hydrodynamic conditions, membrane type and module configuration and the presence of higher molecular weight compounds, either produced by microbial metabolism or introduced into the process. Controlling membrane fouling is the key issue in membrane bioreactors operation. Nevertheless, general guidelines for the integrated operation and control of membrane bioreactors, which take into account not only type and washing frequency of the membranes, but also the main biological parameters, i.e. especially microbiological imbalances, have not been published yet.

Intelligent Environmental Decision Support Systems (IEDSS) are multi-level, knowledge-based computer systems that not only reduce decision-making time, but also improve the consistency and quality of the decisions. DSS are able to deal with complex problems by integrating AI techniques with statistical/numerical methods under a common architecture [Poch *et al.*, 2004]. IEDSS have demonstrated potential to improve the management of Wastewater Treatment Plants (see for example [R-Roda *et al.*, 1999; Sànchez-Marrè *et al.*,

1999; R-Roda *et al.*, 2001; Rodríguez-Roda *et al.*, 2002; Comas *et al.*, 2001, Comas *et al.*, 2003; Martínez *et al.*, 2006a; Martínez *et al.*, 2006b]).

This paper summarizes and highlights the preliminary results of a demonstration research project. The main objective of the project is to design, develop and implement an IEDSS to control remotely and supervise in real-time the integrated operation of a membrane bioreactor pilot plant for wastewater treatment and reuse. Special emphasis is placed on handling problematic situations whose correct and quick detection minimizes the risk of biofouling. The DSS also establishes a hierarchy in fault detection [Genovesi *et al.*, 2000], distinguishing three main levels: sensor failure (SF), sub-process failure (SPF, for example in control loops) and operational problems (PF, process failure). The main purposes of the IEDSS are:

- to increase operational reliability; for this it is of great importance to integrate the biological process for organic matter and nutrient removal with the physical effect of membrane filtration and backwashing/relaxation,
- to reduce operational and energy costs,
- to maintain or improve the effluent quality for water reuse (higher efficiency),
- to learn automatically and reuse the acquired knowledge.

The paper is organized as follows: First of all, the environmental system under study is described detailing the design and operating characteristics. Then, the architecture and development of the IEDSS, especially the reasoning core (rule-based expert system, model-based system, case-base reasoning system, control algorithms and mathematical model), are described. Finally, some conclusions are drawn.

2. ENVIRONMENTAL SYSTEM UNDER STUDY

The experimental environmental system under study consists of a pilot plant for biological nitrogen removal and tertiary treatment of real wastewater (Figure 1). The pilot plant comprises a pre-screening system, bioreactor (total volume 2260 L) with UCT configuration (anaeobic, anoxic and aerobic tanks) followed by a compartment with hollow fiber membranes (Microza from Asahi Kasei Chemicals Corporation, a total area of 12 m^2 of membrane of pore size 0.1 μm.). Finally, a reverse osmosis system treats the membrane bioreactor permeate producing water of good quality for any water reuse. Raw wastewater from the full-scale Castell d'Aro WWTP (175000 P.E.), located in Catalonia (NE of Spain), is pumped continuously into the pilot plant to maintain a fixed permeate flux (between 7.5 and 24 L/m^2h).

The pilot plant is fully equipped and automated with online sensors (conductivity, pH, RedOx potential, ammonium, dissolved oxygen, suspended solids, temperature and scum detector), pressure transducers (for transmembrane pressure -TMP- and permeabilty calculation), flow meters and level transmitters, and the use of a PLC and SCADA system that acquires digital and analogical data and controls all the automatic control loops of the plant, mainly the control of aeration, permeate and backwashing fluxes, hydraulic retention time, sludge retention time or mixed liquor suspended solids (biomass) concentration and recycles.

In addition to online data, the routinely monitoring protocol of the pilot plant performance also includes the measurement (of influent and effluent), three times per week, of the rest of the parameters of interest for the supervision of the organic matter and nutrient removal as well as the physical effect of membrane filtration and backwashing: organic matter (TC, TOC, COD and BOD), nitrogen (TKN, ammonium, nitrite and nitrates), phosphate, inorganic carbon, alkalinity, extracellular polymeric substances and suspended solids. Moreover, biological activity and diversity of the biomass is regularly checked.

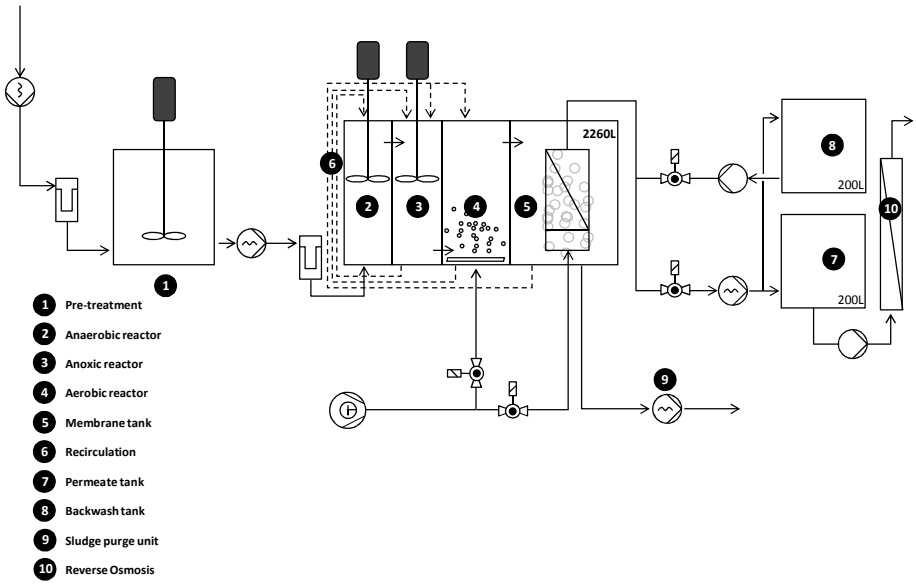


Figure 1. Schematic of the pilot plant under study.

3. IEDSS FOR MEMBRANE BIOREACTORS

The prototype of the intelligent environmental decision support system for membrane bioreactors should guarantee monitoring and real-time robust control, as well as fault detection and remote supervision of the integrated operation of the biological nutrient removal processes and physical filtration and backwashing processes that take place in membrane bioreactors. Figure 2 illustrates the hierarchical control architecture proposed for the intelligent control of membrane bioreactors:

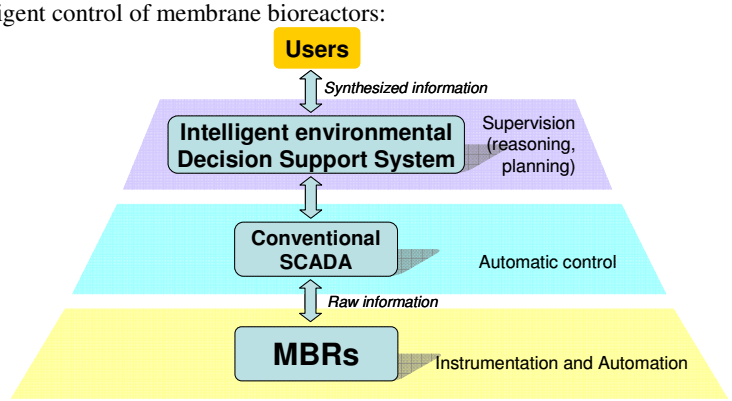


Figure 2. Hierarchical control arquitecture proposed for the integrated control of membrane bioreactors (adapted from [Ayesa *et al.*, 2006]).

The proposed architecture locates the intelligent decision support system hierarchically on top of the conventional supervisory control and data acquisition systems (SCADAs) typically existing in most of the full-scale plants. SCADA systems carry out the acquisition and validation tasks of the data to be used later in the automatic control loops that are implemented (as minimum control of dissolved oxygen in aerobic reactor and, in some cases, control of biomass concentration or sludge age, control of nitrate, control of permeate flux, etc.). The supervisory tasks carried out by the IEDSS allow to establish automatically and in real time the control loop set points needed to optimise the membranes life cycle while, at the same time, ensure reliable and satisfactory nutrient removal. Besides this robust control, the IEDSS allows to detect sensor faults, controller faults and process faults.

The exhaustive list of functionalities that IEDSS for membrane bioreactors should cover includes:

- data acquisition systems throughout communication with SCADAs,
- detection of sensor-related faults,
- advanced (robust) control, establishing the set points of the automatic control loops implemented in the SCADA,
- detection of control loop faults,
- detection of biological, filtration and backwashing process faults,
- automatic learning system,
- to perform daily reports about the plant performance and,
- integration and remote (via Internet) user communication.

4. ARCHITECTURE OF THE IEDSS

The proposed architecture of the IEDSS for membrane bioreactors is based on three levels:
- The first level, called data acquisition, includes the communication and detection of sensor fault modules. The data communication module carries out the tasks involved in the remote data reading and writing throughout communication with SCADA via OPC server and client, and its recording. Different time scales and heterogeneous (both quantitative and qualitative) data should be acquired.
- The second level, called knowledge management, implements the reasoning tasks which, from the available information, allow inferring the process state and afterwards proposing, implementing and simulating a control strategy. The tasks from this level can be carried out thanks to the integrated use of a plant-wide mathematical model (biological process + filtration and backwashing), advanced control algorithms, multivariate statistical techniques and tools from artificial intelligence (rule-based expert systems, model-based reasoning systems and case-based reasoning systems) that allow to optimise the filtration/backwashing and complete nutrient removal processes, whilst supervising controller and process faults (Figure 3).

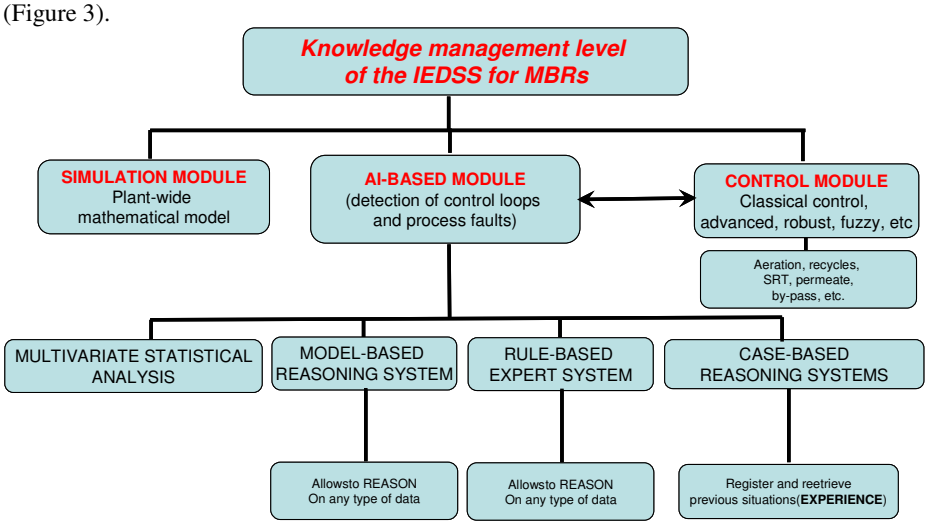


Figure 3. Knowledge management level of the IEDSS for membrane bioreactors.

The rule-based and model-based systems should include empirical and tacit knowledge to relate parameters of biological control (sludge retention time, mass load, dissolved oxygen concentration, temperature, process performance, proliferation of filamentous organisms, etc.) with those relevant for the membrane unit operation (TMP, permeability, extracellular polymeric substances -EPS-, washing frequency and type, particle size distribution, etc.). At the same time, the IEDSS is meant to be equipped with an automatic learning module of the membranes life cycle that records and analyses accumulated experiences and allows acquired knowledge to be reused in the future whenever similar situations arise.
- The third level, the support system level, performs the supervision and integration tasks, processing the conclusions from the previous level and proposing the control strategy (set

points modification etc.). This level also allows evaluating the control alternatives by means of simulation. The interaction with the end-user is essential in this level, through a remote application (via Internet, www.colmatar.es) with user-friendly and interactive interfaces (Figure 4).

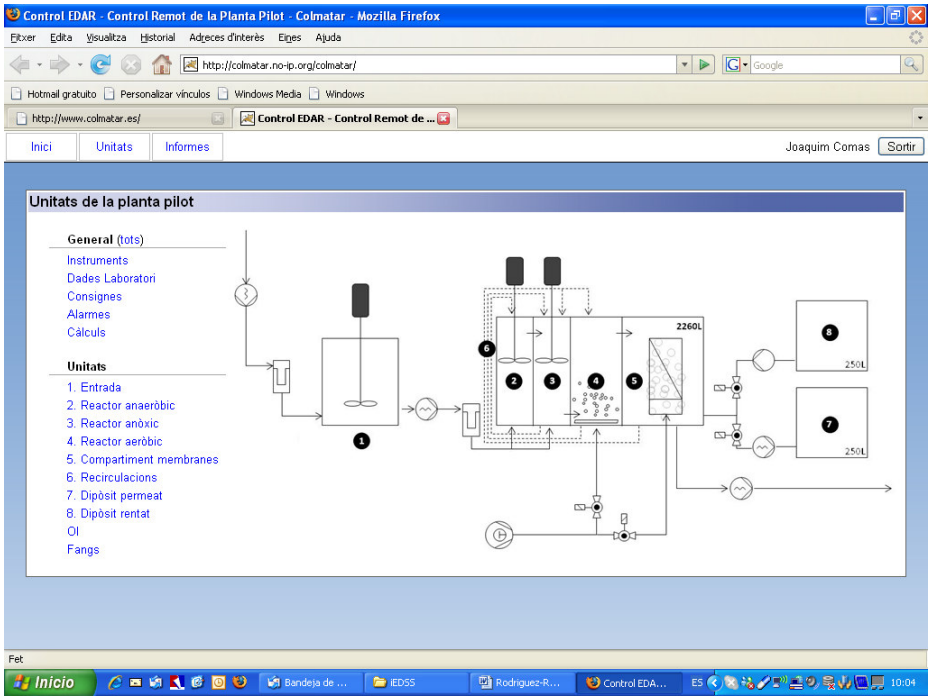


Figure 4. Homepage of the web-based IEDSS for membrane bioreactors.

5. CONTROL MODULE OF THE IEDSS

The control module of the IEDSS governs the automatic or semi-automatic control loops of the most important parameters for the biological and physical processes of membrane bioreactors: dissolved oxygen control, being the aeration flow rate in the aerobic reactor as control action, closed-loop control of permeate and backwashing fluxes, control of feed flow rate (and thus hydraulic retention time), closed-loop control of sludge retention time (or mixed liquor suspended solids or biomass concentration), based on wasting sludge flow rates, and closed-loop control of sludge and nitrate recycles.

The control module of the IEDSS includes the necessary knowledge to fix the set points for all the automatic or semi-automatic conventional controllers of membrane bioreactors (e.g. PI or PID, usually implemented in the SCADA systems) and the new algorithms developed for more advanced controllers (e.g. FLC). The adjustment of control set points (DO, permeate flux, sludge retention time, etc.) can be related to the detection of faults by means of the AI-based module. This way the control module ensures an optimal and robust integrated operation of the membrane bioreactors.

6. AI-BASED MODULE OF THE IEDSS

The use of AI tools and models provides direct access to knowledge and expertise and makes the IEDSS capable of supporting learning and decision making processes. Their integration with numerical and/or statistical models in a single system provides higher reliability [Poch *et al.*, 2004]. This module becomes thus the reasoning core of the IEDSS and it comprises multivariate data analysis, rule-based reasoning, model-based reasoning

and case-based reasoning. The aim of this reasoning module is to establish the set of set points for all the control loops or other manual or semi-automatic control actions (e.g. chemical cleaning of the membranes) that continuously ensure lower operational and energy costs while at the same time maintains or improves effluent quality.

6.1 Multivariate data analysis

Multivariate analysis, based on both statistical and AI-related methods, explore historical databases comprised of several types of variables, including physical, chemical and biological data. The aim is to find clear interrelationships among variables in a low-dimensional space and ultimately to study the possibility to ascertain a model of sub-process faults (e.g. problems in the DO control loop due to sensor fault) or process faults (e.g. biofouling indicated by a significant decrease of membrane permeability over time).

6.2 Rule-based expert system

A Rule-Based Expert System (RBES) is a computer program that emulates human expert reasoning processes when making decisions to confront problems, in this case the integrated operation of nutrient removal and membrane filtration. The key part of a RBES is its knowledge base. In this case the knowledge base, mainly built upon specialised literature and multivariate data analysis, will include symptoms and reasoning process to diagnose and propose control strategies for:

- detection of control loop faults,
- detection of typical process faults of the biological nutrient removal (loss of nitrification capacity, bulking, organic shock, etc.).

The detection of these faults may obviously imply set point modifications of the automatic control systems.

- estimation of costs and risks.

6.3 Model-based reasoning system

Due to the ease of its maintenance and reusability, model-based reasoning technique has been identified as more adequate for the development and implementation of the knowledge base for the optimal operation of the physical membrane filtration [Peischl and Wotawa, 2003]. It implies developing an appropriate model for both the filtration and backwashing processes which are abstract enough to be computationally tractable but detailed enough to provide significant diagnosis results for the membrane biofouling rate. Ultimately this part of the knowledge base, based on the factors influencing biofouling (mainly biomass characteristics, wastewater characteristics e.g. EPS, permeate flow and aeration intensity) must maximize the net permeate flux whereas the biofouling and energy consumption are minimized, and thus membrane life-time maximized.

6.4 Case-based reasoning system

The Case-Based learning and Reasoning System (CBRS) of the membranes life-cycle will allow analyzing and registering those significant events relative to the filtration and backwashing processes with the aim to reuse knowledge gained in previous experiences, given that similar situations have to be confronted. This system allows the accumulation of useful experiences for optimisation of the pilot performance. The identification of the most relevant variables in filtration/backwashing processes (such as flux, TMP, permeability, etc.) and the type of experiences that may occur (mainly management of the filtration/backwashing cycle according to the process variables to minimize biofouling and reduce costs) is indispensable for the development of a functional CBRS.

7. SIMULATION MODULE OF THE IEDSS

The simulation module of the IEDSS comprises an integrated model for membrane bioreactors. This integrated model involves the description of the biological nutrient

removal processes, based on ASM models, together with the physical processes of filtration and backwashing, including the membrane fouling or clogging phenomena, which are still currently under study [Guglielmi *et al.*, 2007 or Li and Wang, 2007]. The availability of such a model will enable to perform several simulations aimed at acquiring relevant knowledge for the optimisation of the operation and control of membrane bioreactors. Thus the simulation module of the IEDSS for membrane bioreactors is not intended to be used for online control, but for offline scenario analysis.

8. CONCLUSIONS

This paper presents the architecture and development of an IEDSS for the integrated operation of membrane bioreactors. The proposed web-based IEDSS is based on a three-level architecture to integrate on-line, off-line and qualitative information with different knowledge management techniques, control algorithms and mathematical models. The intelligent core of the IEDSS will be based on the acquisition of the relevant knowledge to optimise the membrane life cycle as well as keeping a good level of nutrient removal performance and its representation by means of the rule-based, model-based and case-based AI paradigms.

ACKNOWLEDGEMENTS

This research project has been funded by the Spanish Ministerio de Educación y Ciencia research project (DPI2006-15707-C02-01) and the Spanish Ministerio de Medio Ambiente project (018/SGTB/2007/3.1). Also, the authors want to acknowledge the support of ASEPMA, ATM, Consorci de la Costa Brava, Depuradores d'Osona, INIMA, S.A. and Sisltech, S.L and of the computer scientists (Pere Clara and Jordi Majó).

REFERENCES

Ayesa E., Grau P., Sagarna J.M., Salterain A., de la Sota A. and Suescun J. (2006). Supervisory control strategies for the new WWTP of Galindo-Bilbao: The long run from the conceptual design to the full-scale experimental validation. *Water Science and Technology*, 53(4-5), 193-201.

Cho J., Song K.G., Yun H., Ahn K.H., Kim J.Y. and Ghung T.H. (2005). Quantitative analysis of biological effect on membrane fouling in submerged membrane bioreactors. *Water Science and Technology*, 51(6-7), 9-18.

Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M., Cortés, U., Freixó, A., Arráez, J. y Poch, M. (2003). A knowledge-based approach to the deflocculation problem: integrating on-line, off-line, and heuristic information. Water Research, 37 (10), 2377–2387.

Comas J., R.-Roda I., Cortés C., Sànchez-Marrè M., Cortés U., Poch M. Implementation of an Intelligent System to supervise the Granollers WWTP: two years of experience. 74th Annual Conference & Exposition on Water Quality and Wastewater Treatment, WEFTEC 2001. A Conference on Instrumentation, Control and Automation. Seminar #105 Implementing Instrumentation, Control, and Information Technology in Activated Sludge Systems. Atlanta, Georgia (USA), October 2001.

Genovesi A., Harmand J., Steyer J-Ph. (2000). Integrated fault detection and isolation: application to a winery's wastewater treatment plant. *Applied Intelligence Journal*, 13, 59-76.

Guglielmi G., Saroj D.P., Chiarani D. and Andreottola G. (2007). Sub-critical fouling in a membrane bioreactor for municipal wastewater treatment: Experimental investigation and mathematical modelling. *Water Research*, 41(17), 3903-3914.

Jiang T., Kennedy M.D., Guinzbourg B.F., Vanrolleghem P.A. and Schippers J.C. (2005). *Water Science and Technology*, 51(6-7), 19-25.

Judd S. (2005). Fouling control in submerged membrane bioreactors. *Water Science and Technology*, 51(6-7), 27-34.

Lesjean B., Rosenberger S., Laabs C., Jekel M., Gnirss R. and Amy G. (2005). Correlation between membrane fouling and soluble/colloidal organic substances in membrane bioreactors for municipal wastewater treatment. *Water Science and Technology*, 51(6-7), 1-8.

Li X.-y. and Wang, X.-m. (2007). Modelling of membrane fouling in a submerged membrane bioreactor. *Journal of Membrane Science*, 278(1-2), 151-161.

Martínez, M., Sànchez-Marrè, M., Comas, J. and Rodríguez-Roda, I. (2006a). Case-based reasoning, a promising tool to face solids separation problems in the activated sludge process. *Water Science and Technology*, 53(1), 209-216.

Martínez, M., Rodríguez-Roda, I., Poch, M., Cortés, U. and Comas, J. (2006b). Dynamic reasoning to solve complex problems in activated sludge processes: a step further in Decision Support Systems. *Water Science and Technology*, 53(1), 191-193.

Peischl B. and Wotawa F. (2003). Model-Based Diagnosis or Reasoning from First Principles. *IEEE Intelligent Systems*, 18(3), 32-37.

Poch M., Comas J., Rodríguez-Roda I., Sànchez-Marrè M. and Cortés U. (2004). Designing and building real environmental decision support systems. *Environmental Modelling & Software*, 19(9), 857-873.

R.-Roda I., Sànchez-Marrè M., Cortés U., Lafuente J and Poch M. (1999). Consider a Case-Based System for Control of Complex Processes, Chemical Engineering Progress, 95(6), 39-45.

R.-Roda I., Comas J., Poch M., Sànchez-Marrè M. and Cortés U. (2001). Automatic Knowledge Acquisition from Complex Processes for the development of Knowledge Based Systems. *Industrial & Engineering Chemical Research*, 40(15), 3353-3360.

Rodríguez-Roda I., Sànchez-Marrè M., Comas J., Baeza J., Colprim J., Lafuente J., Cortés U. and Poch M. (2002). A Hybrid Supervisory System to Support Wastewater Treatment Plant Operation. *Water Science and Technology*, 45(4/5), 289-297.

Sànchez-Marrè M., Cortés U., R.-Roda I. and Poch M. (1999). Sustainable Case Learning for Continuous Domains. *Environmental Modelling and Software*, 14(5), 349-357.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A new definition of Minimum Sustainable Flow based on water quality modelling and fuzzy processing

[a] S. Marsili-Libelli[a], A. Nocita[b], E. Giusti[a] and M. Saccà[a]

[a] Department of Systems and Computers, University of Florence, Via S. Marta 3, 50139 Firenze, Italy (marsili/giusti@dsi.unifi.it)

[b] Museo di Storia Naturale, University of Florence, Via Romana 17, 50125 Firenze, Italy (nocita@unifi.it)

Abstract: Water resources in temperate and Mediterranean regions are increasingly scarce and water authorities need efficient criteria for the management of this precious commodity, guaranteeing an amount of water which ensures the preservation of aquatic life. This minimal discharge is referred to as the Minimum Sustainable Flow (MSF) and has been subject to many differing definitions in differing contexts. Gradually the purely hydraulic definition, very popular in the past, has given way to more comprehensive indicators taking into account the river ecosystem, and fish in particular, leading to very popular methods such as the IFIM (Instream Flow Incremental Methodology). The drawback of this approach is that it requires extensive field measurements and is very localized in space. This paper presents an extension of IFIM, extending its validity from the microhabitat to the macrohabitat and introducing several water quality parameters. From the methodological viewpoint, the new MSF definition is based on a Sugeno fuzzy inferential system, enhancing the system flexibility in producing environmental scenarios. As an example the method is applied to the middle-lower course of the Arno river, flowing in central Italy, showing that this method is better able to detect critical situations than the conventional IFIM approach.

Keywords: Minimum Sustainable Flow, water quality modelling, fuzzy reasoning, fuzzy inference, artificial intelligence.

1. INTRODUCTION

The growing concern about water shortage has fostered an increasing interest in the parameters defining the ecological status of natural water bodies, and rivers in particular, as tools in implementing water conservation strategies. The concept of Minimum Sustainable Flow (MSF), see e.g. [Stalnaker et al., 1995], once defined in purely hydraulic terms, in receiving renewed interest and several new definitions are being proposed to extend its ecological significance in addition to the original hydraulic meaning. The foremost effort to produce an all-encompassing habitat-oriented definition is the IFIM (In-stream Flow Incremental Methodology). It is a complex analytical and conceptual structure, conceived for the management of river flow regimes variations caused by human exploitation of the resource (Stalnaker *et al*., 1995). Therefore, IFIM can provide answers to problems of river management in relation to the aquatic ecosystem [Bovee *et al*., 1998]. However, its limitation, as a biological method, is its focus on the high-end of the food chain, i.e. fish, disregarding all the underlying water quality parameters and the aquatic food chain, which contribute to the welfare of the whole ecosystem. Further, the practical implementation of the IFIM procedure requires extensive field work and its results are very limited in space, being microhabitat-oriented. Conversely, the growing need for regulatory tools calls for

simpler assessment tools, which can be used for scenario generation using synthetic water quality data.

In this paper, a new MSF definition is proposed, which takes into account some ecological parameters than were previously disregarded in the conventional MSF definition based on the IFIM approach. In this proposal, not only flow, velocity and depth are considered, but also water quality parameters such as temperature, pH, dissolved oxygen and unionized ammonia. All these parameters contribute to the computation of the Composite Suitability Index (CSI) from which the MSF is obtained. This new methodology results from the integration between an extension of the IFIM approach and a water quality model (QUAL2Kw) [Pellettier *et al.*, 2006] fed with hydraulic data computed with HEC-RAS 6 [HEC, 1991], a widely-used one-dimensional open channel flow computational scheme. Both packages are of public domain and can be freely downloaded from the internet. A fuzzy computational method based on Sugeno fuzzy inference [Babuska, 1998] is used to combine the individual suitability contributions and yield the extended MSF.

The paper is organized as follows: after describing the three steps of the computational method, this is applied to the computation of the MSF for the lower part of the Arno catchment, for which critical conditions often develop during the summer months. It is shown that the proposed method is more flexible and comprehensive in determining the critical MSF during the low-flow season.

2. A COMPUTATIONAL SCHEME FOR THE MINIMUM SUSTAINABLE FLOW

The proposed method is an extension of the well-known IFIM (In-stream Flow Incremental Methodology) approach [Stalnaker *et al.*, 1995; Bovee *et al*, 1998], which is supposed to be implemented in five sequential phases: problem identification, study planning, study implementation, alternatives analysis, and problem resolution. The main difference between this approach and IFIM is the spatial scale, which in our case is catchment-oriented as contrasted to the micro-habitat approach adopted by IFIM.

The MSF computation develops in three successive steps, as illustrated in Figure 1.

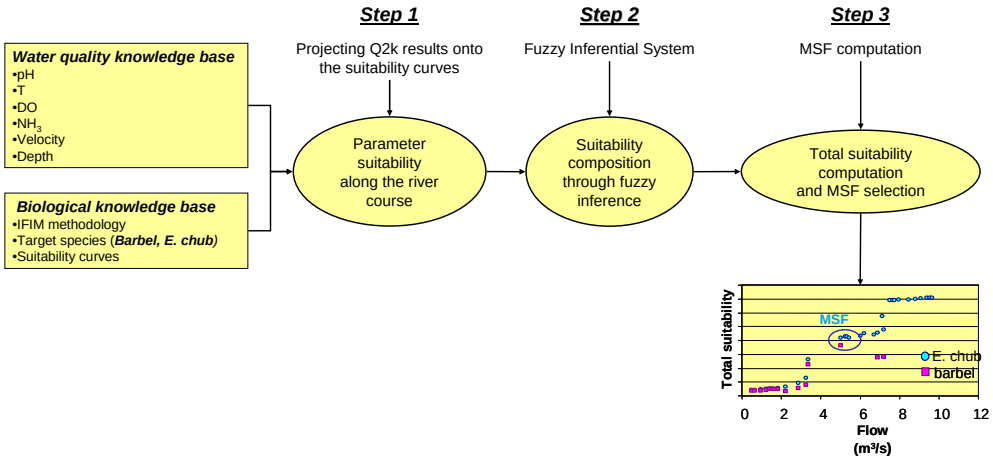


Figure 1. The three step in the procedure for the MSF computation.

The example figures shown in this section to illustrate the method refer to the Arno river case study and the two most typical fish species abundant in that environment: Barbel (*Barbus tyberinus*) and European Chub (*Leuciscus cephalus*), the most widespread autochthonous and reophyl species. The three steps of Figure 1 are now examined.

2.1. Step 1: Projecting the quality data onto the suitability curves.

The suitability curves for each parameter shown in Figure 2 were obtained from experimental observation regarding the fish habitat. Combining this information with the catchment quality obtained through simulation of the hydrodynamic (Hec-Ras) and quality (QUAL2Kw) models yields the river-wide suitability curve for each parameter, shown in

Figure 3. These curves are obtained separately from field studies and literature data and refer to the considered species (Barbel and European Chub). Each hydraulic or quality parameter contributes to the habitat suitability in a different manner, and for each of them the fish response has been normalized between 0 (totally unsuitable habitat) and 1 (perfectly comfortable habitat). The combination of individual suitability indexes will be described in the next step.

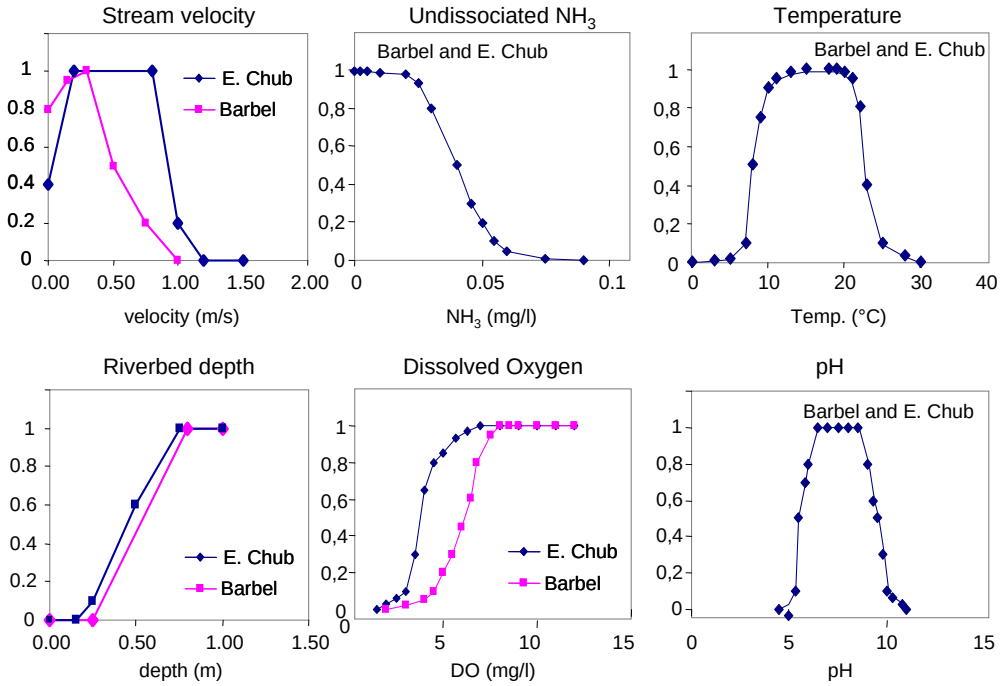


Figure 2. Normalized suitability curves for the river quality parameters considered in the study for European Chub and Barbel. Wherever a single line is shown both species exhibit the same suitability.

Feeding each diagram of Figure 2 with the quality data produced by QUAL2Kw the overall suitability for that parameter can be obtained for the whole river length, as shown by the examples of Figure 3.

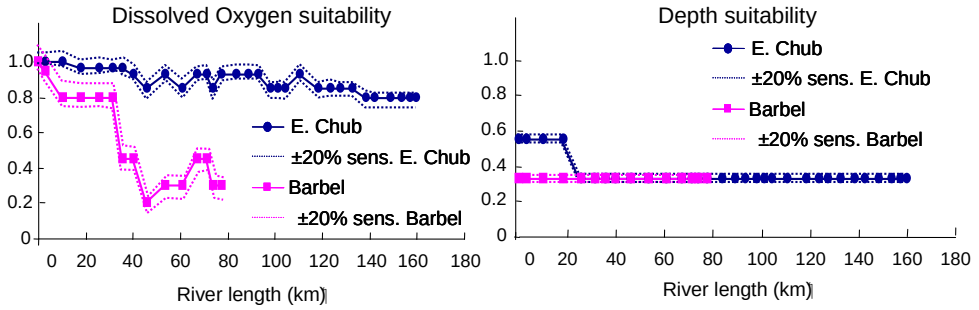


Figure 3. Example of single parameter suitability curves computed along the whole river length. The Barbel suitability data cover only the first 80 km because its habitat is limited to this upstream part. The dotted curves represent the ± 20% sensitivity with respect to the fuzzy membership functions: it can be seen that dissolved oxygen has a higher sensitivity than depth.

2.2. Step 2: The fuzzy inferential system.

Apart from the hydrologic data, which are processed separately, the Sugeno fuzzy inferential engine has four water quality suitability indexes as antecedents (dissolved oxygen, unionized ammonia, velocity and depth, all defined in the (0, 1) support) and four singleton consequents. Therefore the complete inferential set would be of 3×4×4 = 48 rules.

The knowledge base behind Table 1 was extracted by direct experimentation and prior biological experience, both direct and from the literature. By direct experimentation, however, this set was reduced to 31 practically useful rules, listed in Table 1, where H, M, L are Gaussian-shaped membership functions defined on the support of each antecedent variable qualifying suitability to DO, NH3, velocity U and depth H as High, Medium and Low respectively. The four consequent singletons are set at 1 (CSI_H), 0.8 (CSI_MH), 0.5 (CSI_ML) and 0.2 (CSI_L).

Table 1. Rules of the Sugeno fuzzy inferential system. The suffix 's' after each antecedent parameter stands for 'suitability'.

1. If (DOs is H) and (NH3s is M) and (Us is M) and (Hs is H) then (CSI is CSI_H)
2. If (DOs is H) and (NH3s is M) and (Us is H) and (Hs is M) then (CSI is CSI_MH)
3. If (DOs is M) and (NH3s is H) and (Us is M) and (Hs is H) then (CSI is CSI_MH)
4. If (DOs is H) and (NH3s is L) and (Us is H) and (Hs is L) then (CSI is CSI_ML)
5. If (DOs is L) and (NH3s is H) and (Us is H) and (Hs is L) then (CSI is CSI_ML)
6. If (DOs is L) and (NH3s is M) and (Us is L) and (Hs is M) then (CSI is CSI_L)
7. If (DOs is M) and (NH3s is L) and (Us is M) and (Hs is L) then (CSI is CSI_ML)
8. If (DOs is L) and (NH3s is M) and (Us is M) and (Hs is L) then (CSI is CSI_L)
9. If (DOs is H) and (NH3s is H) and (Us is M) and (Hs is M) then (CSI is CSI_MH)
10. If (DOs is H) and (NH3s is H) and (Us is L) and (Hs is L) then (CSI is CSI_ML)
11. If (DOs is M) and (NH3s is M) and (Us is H) and (Hs is H) then (CSI is CSI_H)
12. If (DOs is M) and (NH3s is M) and (Us is L) and (Hs is L) then (CSI is CSI_L)
13. If (DOs is L) and (NH3s is L) and (Us is M) and (Hs is M) then (CSI is CSI_ML)
14. If (DOs is H) and (NH3s is H) and (Us is H) and (Hs is L) then (CSI is CSI_MH)
15. If (DOs is M) and (NH3s is M) and (Us is M) and (Hs is H) then (CSI is CSI_MH)
16. If (DOs is M) and (NH3s is M) and (Us is M) and (Hs is L) then (CSI is CSI_ML)
17. If (DOs is H) and (NH3s is M) and (Us is M) and (Hs is M) then (CSI is CSI_MH)
18. If (DOs is H) and (NH3s is L) and (Us is L) and (Hs is L) then (CSI is CSI_L)
19. If (DOs is M) and (NH3s is H) and (Us is M) and (Hs is M) then (CSI is CSI_MH)
20. If (DOs is L) and (NH3s is H) and (Us is L) and (Hs is L) then (CSI is CSI_L)
21. If (DOs is M) and (NH3s is L) and (Us is L) and (Hs is L) then (CSI is CSI_ML)
22. If (DOs is M) and (NH3s is M) and (Us is H) and (Hs is M) then (CSI is CSI_H)
23. If (DOs is L) and (NH3s is L) and (Us is H) and (Hs is L) then (CSI is CSI_ML)
24. If (DOs is H) and (NH3s is M) and (Us is H) and (Hs is H) then (CSI is CSI_H)
25. If (DOs is H) and (NH3s is H) and (Us is M) and (Hs is H) then (CSI is CSI_MH)
26. If (DOs is L) and (NH3s is M) and (Us is L) and (Hs is L) then (CSI is CSI_L)
27. If (DOs is L) and (NH3s is L) and (Us is M) and (Hs is L) then (CSI is CSI_ML)
28. If (DOs is L) and (NH3s is M) and (Us is M) and (Hs is M) then (CSI is CSI_ML)
29. If (DOs is M) and (NH3s is M) and (Us is L) and (Hs is M) then (CSI is CSI_ML)
30. If (DOs is M) and (NH3s is M) and (Us is M) and (Hs is M) then (CSI is CSI_MH)
31. If (DOs is L) and (NH3s is L) and (Us is L) and (Hs is L) then (CSI is CSI_L)

The fuzzy inference engine and its preliminary part, generating the river quality parameters, is shown in Figure 4. The result of this procedure is the production of the **C**omposite **S**uitability **I**ndex (CSI)

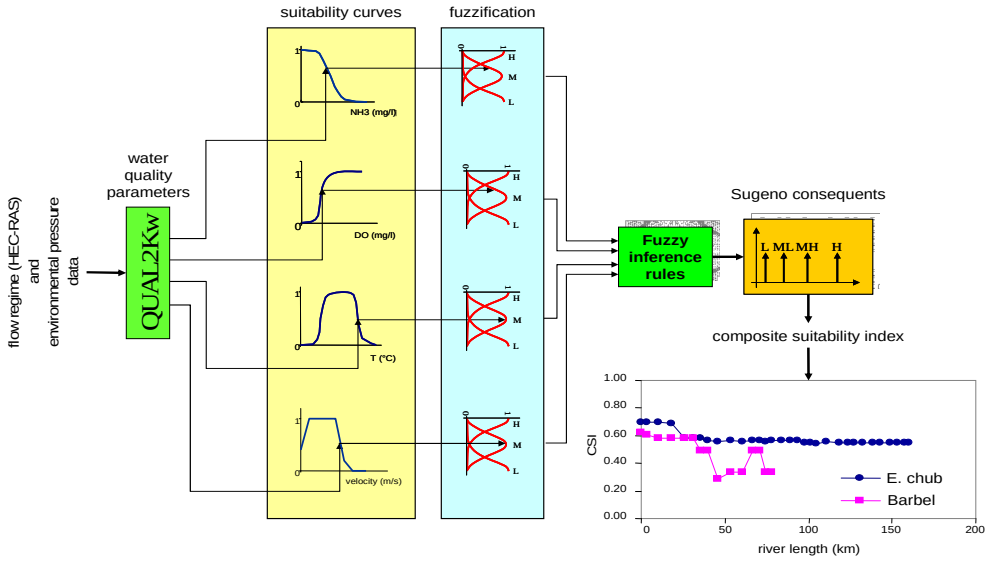


Figure 4. Computational scheme producing the Composite Suitability Index.

In the fuzzy rules of Table 1 the AND and THEN connectives are implemented with the 'minimum' operator. Then the *Composite Suitability Index* (CSI) is obtained by weighted average defuzzification

$$CSI = \frac{\sum_{i=1}^{31} \mu_i b_i}{\sum_{i=1}^{31} \mu_i}, \tag{1}$$

where μ_i is the degree of fulfilment of each rule in Table 1 and b_i is the corresponding singleton previously defined.

2.3. Step 3: Computation of the actual Minimum Sustainable Flow

The final goal of the study is to obtain the *weighted usable area* (WUA) as a function of flow, from which the minimum sustainable flow can be determined as the minimum discharge still providing a fish suitable habitat. However, the CSI defined by eq. (1) as the output of the computational scheme of Figure 4 is not yet in terms of WUA vs. flow, therefore we are not yet in a position to select the MSF on such a graph. It is in this final step that the present method differs considerably (apart from the fuzzy approach) from the conventional IFIM procedure. Contrary to the classical IFIM method, where the WAU is computed as the weighted sum of the products of the CSI by the pertinent cells, here the WUA is obtained as the product of the reach CSI multiplied by the average reach cross section, namely

$$WUA = CSI \times S_r, \tag{2}$$

where S_r is the average river section computed along the whole reach of interest. Thus there will be one WUA for each reach. In this way, the emphasis is shifted from the microhabitat, typical of IFIM, to a catchment-wide dimension.

In the WUA/flow diagram, the Minimum Sustainable Flow (MSF) is usually determined as the first break-point in the WUA-discharge diagram [Milhous et al., 1989]. To obtain such a diagram requires an iterated computation of the WUA for the whole range of expected flows. The break-point normally represents the point where a major change in the WUA/flow slope occurs. It represents the boundary between the monotonic part and the region where WUA increases can be obtained only at the expense of a major flow increase. Thus from a cost-benefit viewpoint the breakpoint flow represents the best compromise between environmental sustainability and the economic value of the water resource and as such it can be regarded as the Minimum Sustainable Flow. Figure 5 shows a typical WUA vs. flow diagram with the location of the breakpoint, which can be interpreted as the MSF. The figure refers to a typical section of the Arno river during the summer low-flow condition.

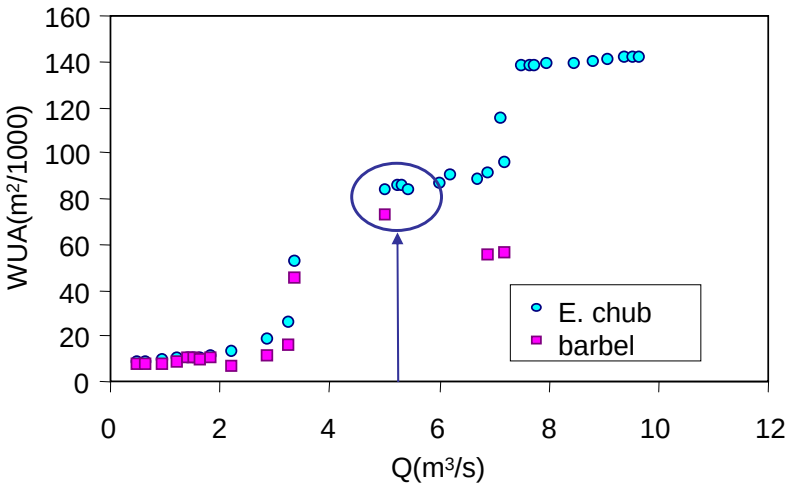


Figure 5. MSF computed for both target species at a typical section of the Arno river during low-flow conditions.

3. APPLICATION TO THE ARNO RIVER

The procedure outlined in Section 2 is now demonstrated with an application to the middle-lower course of the Arno river, flowing through Tuscany, central Italy, for an overall length of nearly 180 km. Figure 6 depicts the extent of the study together with reach boundaries and water exchanges.

3.1. Characteristics of the Arno catchment

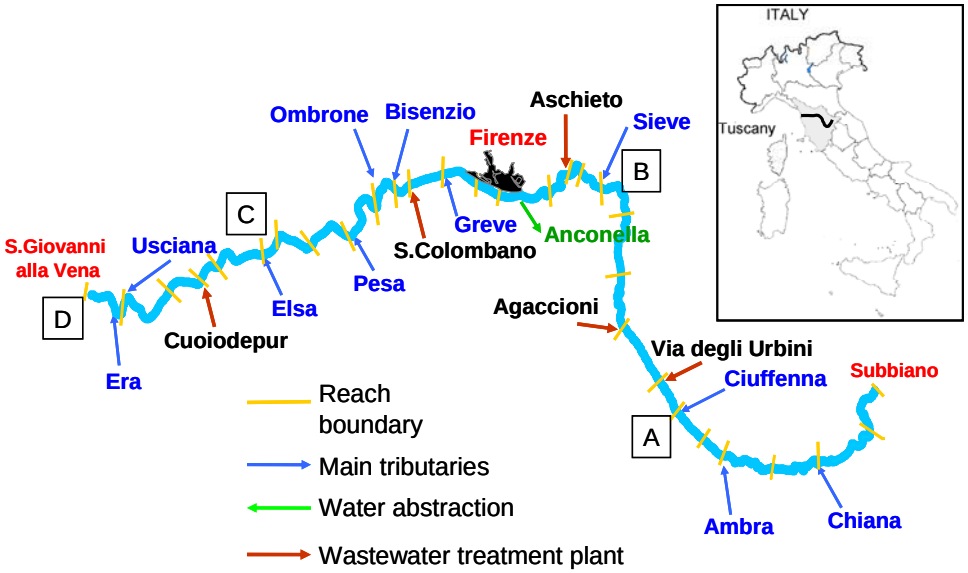


Figure 6. Reach boundaries and main features of the part of the Arno river considered in this study. The letters A, B, C, D refer to the MSF computations shown in Figure 7.The bars indicate the reach boundaries and the arrows show the water exchange points.

Running the procedure of Section 2 for a number of typical flow regimes, mostly in the low-to-middle range, the WUA/flow curves were obtained for each section. The most meaningful of them, corresponding to the reaches labeled as A, B, C, and D in Figure 6, are shown in Figure 7 where the typical low and medium flow are also indicated for comparison.

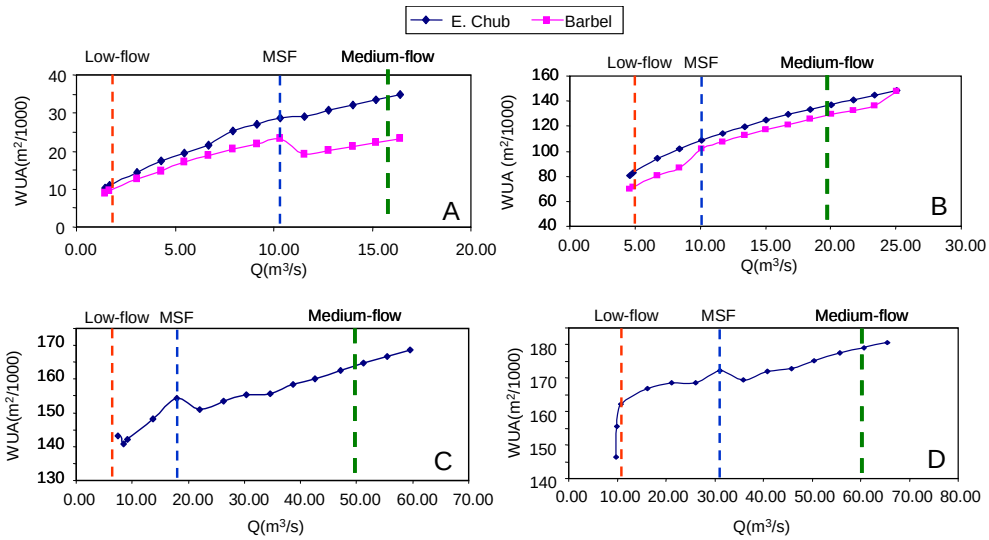


Figure 7. MSF values for four meaningful reaches labelled with the lettering in the map of Figure 6 resulting from the described algorithm. It can be seen that the low-flow condition is much lower than the recommended MSF. In reaches C and D only the European Chub is present, because in the downstream reaches Barbel is absent.

The results of Figure 7 show that during the low-flow condition the Arno river is in a highly critical condition. In fact, according to the method outlined in Section 2, the typical low-flow condition is considerably worse than the MSF, whereas in the medium-flow regimes the situation improves considerably.

4. CONCLUSION

The purpose of the study was twofold, one methodological and one applicative. From the methodological viewpoint, described in Section 2, a new definition of the Minimum Sustainable Flow (MSF) was proposed, based on water quality parameters and not only on an hydraulic basis. Further, the coupling with a water quality model (QUAL2Kw) allows the production of a large number of scenarios for management decision support. The aim of producing a new definition of MSF was to extend the IFIM concept from the microhabitat to the macrohabitat and introduce more water quality parameters in its definition. The Sugeno fuzzy inferential engine represents a neat way for combining the suitability index for each parameter to derive the composite index, from which the weighted usable area is eventually computed, enhancing the system flexibility and paving the way toward more comprehensive MSF definitions.

The applicative aspect, described in Section 3, was the MSF computation for the middle and lower Arno river course flowing through central Italy, showing its critical condition during the summer low flow, since in almost every reach the flow is well below the MSF.

5. ACKNOWLEDGEMENTS

The authors wish to acknowledge the collaboration of the Arno River Water Authority in providing data and information useful for this research.

6. REFERENCES

Babuska R., *Fuzzy modeling for control.* Kluver Publ. Co., Amsterdam, 260 pp, 1998.

Bovee K.D., B.L. Lamb, J.M. Bartholow, C.B. Stalnaker, J. Taylor, and J. Henriksen, *Stream Habitat Analysis Using the Instream Flow Incremental Methodology.* Information and Technology Report USGS/BRD/ITR-1998-0004. Fort Collins, CO: U.S. Geological Survey-BRD. 130 pp, 1998.

Chapra, S. and G. Pelletier, *QUAL2K: A Modeling Framework for Simulation River and Stream Water Quality: Documentation and User Manual*. Civil and Environmental Engineering Dept., Tufts University, Medford, MA., 2003.

Milhous R.T., M.A. Updike and D.M. Schneider, *Physical habitat simulation system reference manual – version II.* U.S. Fish and Wildlife Service Biological Report 89 (16). Washington, D.C., 1989.

Pelletier G.J., S.C. Chapra, and H. Tao, QUAL2Kw e A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. Environmental Modelling & Software 21 (2006) 419 – 425, 2006.

Stalnaker, C., B.L. Lamb, J. Henriksen, K. Bovee, and J. Bartholow, *The instream flow incremental methodology: a primer for IFIM.* National Ecology Research Centre, Internal. Publication. U.S. Department of the Interior, National Biological Service, Fort Collins, Colorado: 43 pp, 1995.

Shepard R.B., *Quantifying Environmental Impact Assessments Using Fuzzy Logic.* Springer Science+ Businness Media, Inc., 2005.

Svobodova Z., R. Lloyd, J. Machova, and B. Vykusova, *Water Quality and fish health.* EIFEC Technical Paper, FAO, Rome, n.54, 1993.

U.S. Army Corps of Engineers, Hydrologic Engineering Center (HEC), *HEC-6, Scour and Deposition in Rivers and Reservoirs, User's Manual*, Davis, CA., 1991.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Improving process performance in wastewater treatment plants using an intelligent environmental decision support system

Lluis Corominas[1,2], Joel Denamur[1], Xavier Flores-Alsina[1] and Joaquim Comas[1]
[1] *Laboratory of Environmental and Chemical Engineering, Faculty of Sciences, University of Girona, Campus Montilivi s/n 17071. Girona (Spain) {joel, xavi, quim}@lequia.udg.cat*

[2] *modelEAU, Département du Génie civil, Université Laval, Québec, G1V 0A6, QC (Canada) lluis.corominas@ gci.ulaval.ca*

Abstract: The objective of this paper is to present the simulation results of implementing and intelligent environmental decision support system (IEDSS) to overcome the limitations of the manual control at the wastewater treatment plant (WWTP) of Vall del Ges (North-East, Spain). The proposed IEDSS integrates online information with knowledge management techniques and aims to reduce the pollution load leaving the plant, at the same time that keeps the operational costs at minimum. To this end, a fuzzy logic engine (FLE) controls the effluent nitrate and ammonium concentration manipulating both aeration and internal recycle flow respectively. The results show significant savings in terms of operational costs and a slightly enhancement of the nitrification/denitrification (N/D) efficiency improving the overall nitrogen removal.

Keywords: wastewater treatment; modelling; fuzzy logic; control; intelligent environmental decision support system (IEDSS).

1. INTRODUCTION

Upgrading or optimisation of biological nutrient removal (BNR) is a common issue in existing wastewater treatment plants (WWTP). The variability in quality and quantity of the influent, the dynamics of the biological processes and the limited knowledge of the physical-chemical and biological interactions that take place in the biological reactor make the nitrification/denitrification (N/D) process complex and therefore difficult to control. The even more widespread use of on-line sensors over the last decades, the automation of a number of operations and progress in automatic control technology have enabled real time process information and thus significantly improved automatic control of the N/D process in the majority of the medium- and large-sized treatment plants [Olsson *et al.,* 2005]. The majority of feedback automatic controllers are based on the data provided with indirect variables such as dissolved oxygen (DO) or the oxidation-reduction potential (ORP) (e.g. Jung Kim *et al*., [2005]). However, an adequate knowledge of N/D process performance requires data about ammonium, nitrates, sludge retention time or even respiration rates to be monitored. Much of the latter data, however, is obtained with some delay after taking samples and analysing specific parameters in the laboratory. This fact limits automatic controllers when certain disturbances in the process occur due to the influent (e.g. toxic events) or to operational problems (filamentous bulking or incomplete denitrification). In addition, the majority of small plants (less than 2000 inhabitant equivalents, IE) typically do not contain even a small part of the instrumentation found in large plants, so that the knowledge, experience and reasoning processes of the human expert continues being of crucial importance.

N/D process control involves high reliability of the process performance in any operational situation and in real time, minimizing the impact on the water body and generating minimal operating costs. This can be only achieved if more and better information from automatic analysers is integrated with typically human characteristics such as analytical capacity, interpretation and the reasoning skills of a treatment process expert. Literature already provides some examples where automatic analysers of nutrients or suspended solids enable new control structures to be designed and implemented to optimize nitrogen removal (e.g. Ayesa *et al*., [2006]). On the other hand, some authors propose control strategies based on changes of aeration volume to improve the biological nitrogen removal in full-scale plants [Thomsen *et al*., 1998] and by simulation [Samuelsson *et al.,* 2002].

Nevertheless, there are still very few examples of the use of tools based on expert knowledge and human reasoning models to improve the full-scale WWTPs control. These tools are based on principles of Artificial Intelligence (AI), the branch of computer science concerned with creating or mimicking intelligent behaviour or "thought" in computers. Among those that have recently generated great expectations, Intelligent Environmental Decision Support Systems (IEDSS) for environmental problems stand out [Poch *et al*., 2004]. These systems are computer programs that attempt to emulate the human capacity to solve problems as if they were experts. The core of the IEDSS incorporates reasoning models used to infer the state of the process. The models used involve the integration of statistical and AI models (e.g. Rule-Based or Fuzzy Logic Systems) with mathematical models (e.g. Activated Sludge Models, ASM) and with advanced control algorithms.

This paper illustrates (via simulation) a successful application of an IEDSS to ease the integration of online information with knowledge management techniques to overcome the limitations of the current manual control in the WWTP of Vall del Ges. The existing mode of operation is based on an open loop regime making the plant non-capable to handle with process disturbances. The final aim of the IEDSS is to reduce the operational costs and slightly enhance the N/D efficiency, therefore keeping both environmental and financial cost at minimum. To this end, a fuzzy logic engine (FLE) controls the effluent nitrate and ammonium concentration manipulating both aeration and internal recycle flow respectively. The paper is organized as follows: First of all, the facility under study is described detailing the design and operating characteristics. Then, the development of the deterministic model and of the new IEDSS, especially the reasoning core (fuzzy-logic engine and rule-based system), are described. Next, the simulation procedure and evaluation criteria used are briefly introduced. Finally, the simulation results of the IEDSS implementation in the WWTP of Vall del Ges are presented and discussed.

2. METHODS

2. 1. Environmental System under study

The WWTP of Vall del Ges (North-East, Spain) is designed to remove biologically organic carbon and nitrogen. The large number of industries nearby makes chemical precipitation the most suitable technology to achieve a relatively low concentration of phosphorus in the effluent. It was constructed in 1993 for a design capacity of 42000 IE but it receives only 30.000 IE with an average flow rate of 4800 $m^3 \cdot d^{-1}$, and an organic and nitrogen load of 1946 kg $COD \cdot d^{-1}$ and 167 kg $N \cdot d^{-1}$ respectively. Figure 1 shows the dynamic profile of influent COD, BOD5 and TKN together with the influent flow and the temperature during the studied period. The studied WWTP is depicted in Figure 2. The water line comprises a pretreatment with sand and fat removal units followed by and activated sludge system. The latter has an Orbal configuration (see for example Metcalf & Eddy, [2003]) with three concentric channels (ASU1, 2 & 3) of 3370, 2607, and 1843 m^3 and the aeration is transferred through surface rotors. The external channel has 6 rotors that remain idle because it is used as the anoxic zone of the treatment. The inner and the middle channels are aerobic with an average DO concentration of 0.6 and 1.5 $g \cdot m^{-3}$ respectively. The aerobic zone presents 8 rotors (four in each channel) but working in pairwise fashion because there are only 4 engines. Finally, two secondary settlers are followed, with a circular area of 628 m^2 and a total volume of 2370 m^3.

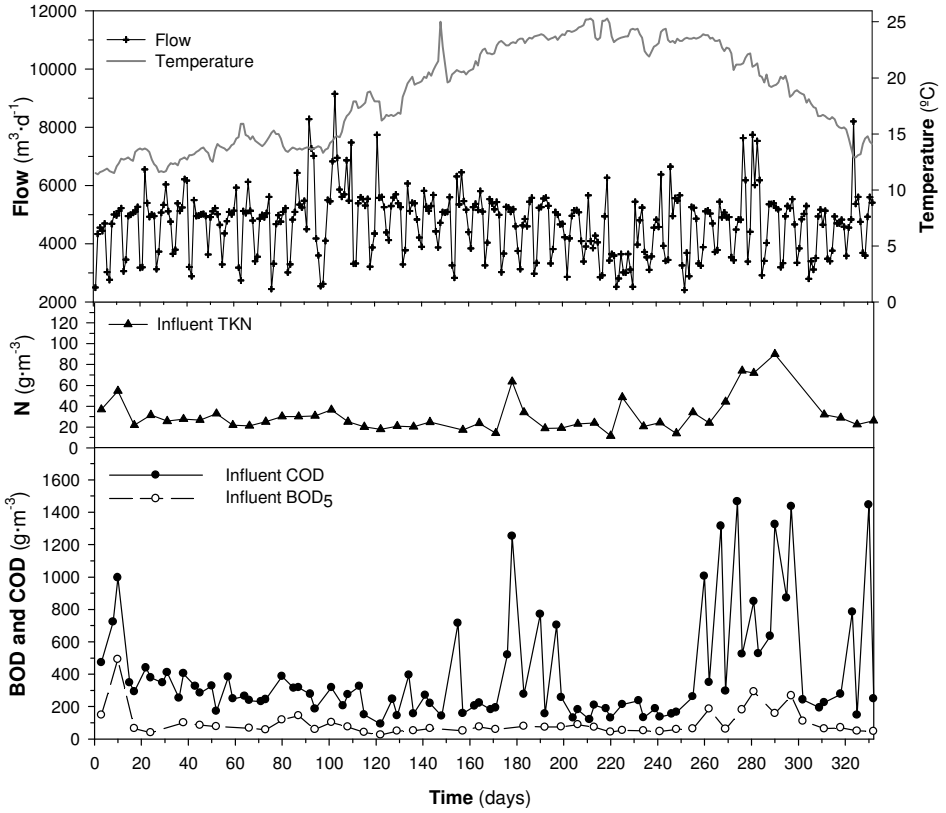


Figure 1. Influent wastewater flow and components dynamic profile

The current control of aeration is based on setting time intervals and thus switching the aeration ON and OFF. An internal recirculation (Q_{intr}) links the inner (ASU3) and the external (ASU1) channels with a constant flow of 23760 $m^3 \cdot d^{-1}$. Also, there is an external recirculation (Q_r) from the bottom of the settlers to the external channel (ASU1) with a constant value of 4320 $m^3 \cdot d^{-1}$, and the wastage flow rate is 12$m^3 \cdot h^{-1}$. A DiaMon analyser is providing on-line data for the effluent ammonium and nitrate but is not linked to any type of automatic control.

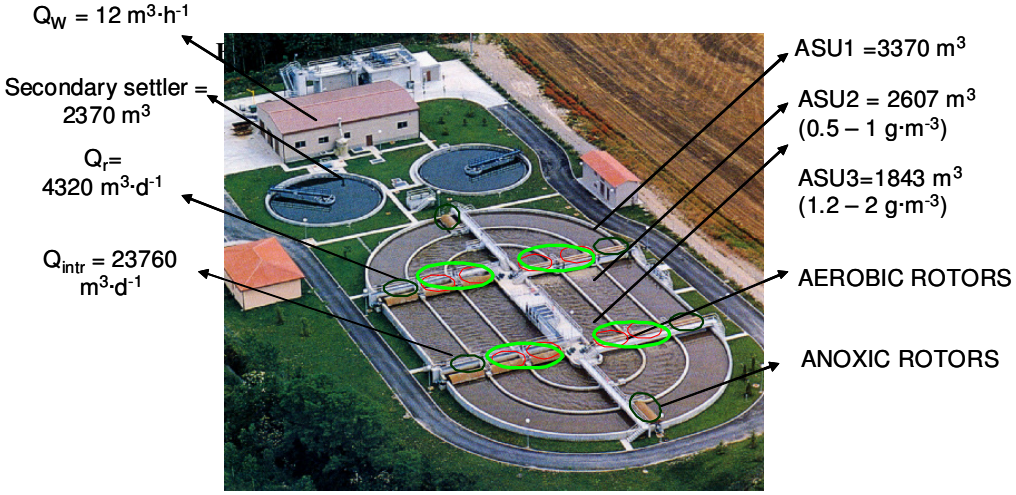


Figure 2. Representation of the studied WWTP

2. 2. Determistic model of the WWTP of Vall the Ges

The model is calibrated following the Unified Calibration protocol proposed by the IWA Task Group of Good Modeling Practices (www.modeleau.org/GMP_TG/). This calibration protocol involves several steps: i) project definition, ii) data collection and reconciliation, iii) model set-up, iv) calibration and validation and, v) simulation and result interpretation.

Project definition: The main goal of this study is to obtain a deterministic model that allows comparing the pefomance of the Vall del Ges WWTP when the proposed IEDSS is either included or not.

Data collection and reconciliation: One year historical data was collected for the influent and effluent characteristics, as well as the operating parameters. Two measuring campaigns (of one day length each) were also carried out to check the quality of the data and proceed with the data reconciliation. In this sense, phosphorus mass balances were conducted with the aim of determining the proper sludge retention time. All these data were also used for influent wastewater characterization, which was based on the STOWA protocol (Hulsbeek et *al.* 2002). Finally, a hydrodynamic study (tracer test) of the WWTP was conducted.

Model set-up: The complete model of a WWTP is based on the hydraulic, oxygen transfer, biokinetic, settling and sensor models. The results of the tracer test came up with the structure of 3 anoxic reactors ($ASU3_1$=$ASU3_2$=$ASU3_3$=1123m^3) for the external channel, and 2 aerobic reactors (ASU2=2607 m^3 and ASU1=1843 m^3) for the middle and inner channels. Aeration capacity was modelled as the mass transfer coefficient (K_La), and its value was obtained adjusting the experimental concentration profiles of DO in the liquid phase. The biological model ASM1 [Henze *et al*., 2000] was used. The effect of temperature on the kinetics was also considered in the model implementation using the Arrhenius equation. Regarding the settling model, the double exponential settling velocity of Takács *et al.* [1991] was selected as a fair representation of the settling process with a ten layer discretization. Sensor models were also implemented for effluent ammonia and nitrate to emulate the performance of the DiaMon analyser. These sensor models include response time, sample & hold, noise, measuring range as described in Rieger *et al*., [2003].

Calibration and validation: The simulations are carried out using the MatLab-simulink environment. Dynamic simulations were preceded of steady state simulations. This ensures and eliminates the bias due to the selection of the initial conditions for the dynamic results. Steady state was reached simulating 100 days with constant influent flow and concentrations and using default values at 15^oC for the kinetic and stoichiometric parameters [Copp 2002]. Next, a dynamic influent of 332 days was simulated considering that i) for the days without influent concentrations data the mean value between the previous and next day was taken or in some cases the mean values for the whole period, and ii) the daily dynamics in the influent flow were applied based on percentages of the mean flow (according to Metcalf & Eddy, [2003] values). Figure 3 shows a good fit between the experimental and simulated values and the dynamics are properly described, so the model is ready for evaluation of the IEDSS's performance. The observed mismatches are not relevant for accomplishing the objectives of the study. They are related to the nature of the data used to carry out the simulation i.e incomplete, ambiguous and uncertain.

2. 3. Fuzzy Logic Engine and Rule-Based System

The IEDSS is based on a FLE and integrates knowledge management techniques and online information. The IEDSS controls the effluent nitrate and ammonium concentration (input variables) manipulating both aeration and internal recycle flow (output variables). The effluent nitrate and ammonium signals coming from the DiaMon analyzer are used as input variables of the IEDSS. Historical data of effluent ammonium and oxidized nitrogen were analyzed in order to define the membership functions of the FLE. Both ammonium and oxidized nitrogen range from 0 to 10 $gN \cdot m^{-3}$ and they are distributed into 3 labels of concentrations (Low, Normal and High). The membership functions are trapezoidal for low and high labels and triangular for the Normal label.

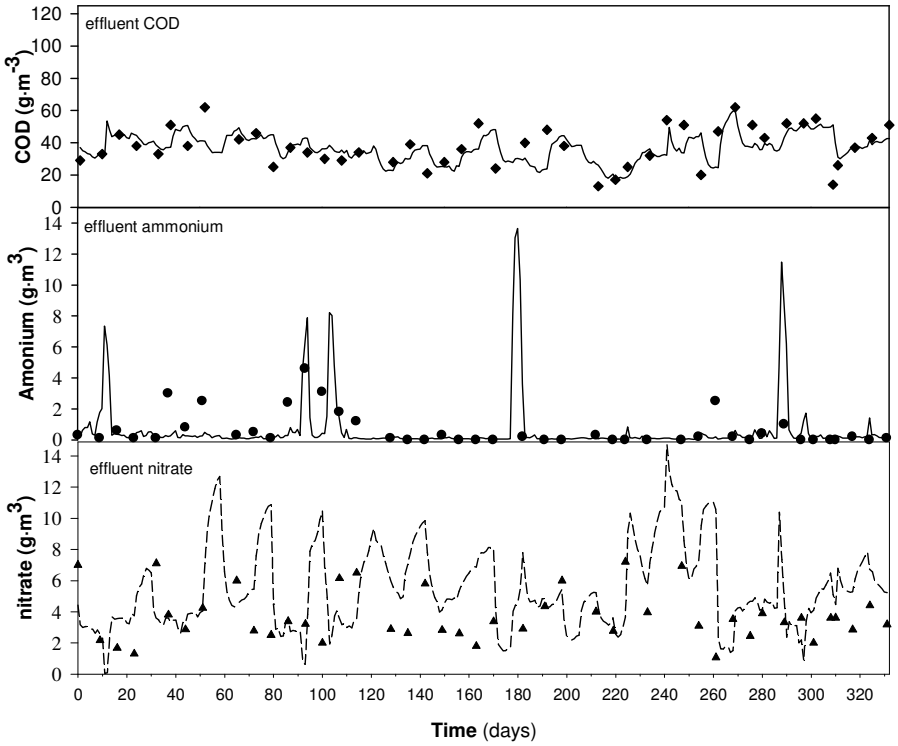


Figure 3. Experimental (dotted) and simulated (line) data for COD, ammonium and nitrate

Aeration (K_La) and internal recycle flow are the output variables. The K_La ranges from 0 to 80 d^{-1}. Thus, 0 and 80 represents when the four rotors are switched OFF and ON respectively. The intermediate values are associated to different working and speed combinations of rotors. On the other hand, the internal recycle flow ranges from 0 to 31680$m^3 \cdot d^{-1}$. Both aeration and internal recycle flow ranges are subdivided into three labels (low, medium and high) in the definition of the output triangular membership functions.

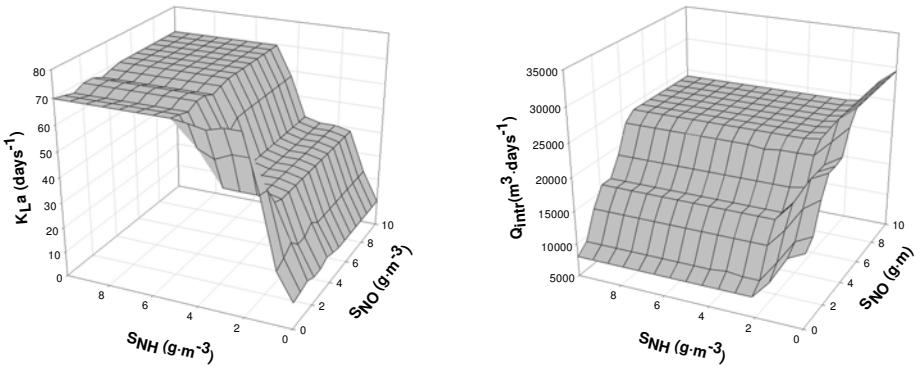


Figure 4. Graphic representation of the decision matrix for (left) aeration flow and (right) internal recycle

Information and good practices from the experts working in the WWTP are the sources to construct the knowledge base including IF-THEN rules that link the values of the input and the output variables. The mapping between the inputs (fuzzification) and the ouputs (defuzzification) is carried out following the mandami approach. The fuzzy toolbox of Maltab-simulink is used to develop the FLE. The graphic representation of the decision matrix can be found in Figure 4 for the aeration (left) and recycle flow (right) respectively.

Thus, on the one hand, when the effluent ammonium nitrogen is high the desired control action would be increasing the aeration flow in order to improve nitrification. High dissolved oxygen concentration in ASU2 & 3 favours the growth of the autotrophic bacteria in charge of nitrification process. On the other hand, when the effluent nitrate nitrogen is high, the control action would be increasing the internal recycle flow in order to improve denitrification. High internal recycle flow increases the quantity of nitrate going from ASU3 to ASU1 and potentially usable for anoxic growth of heterotrophic bacteria. On top of the IEDSS a RBS is developed to check the quality of the data before being introduced into the FLE. This RBS checks the measuring range for the ammonia and oxidized nitrogen values. When implementing the IEDSS in full-scale this will be complemented with other methods e.g. control charts described in Thomann *et al.,* [2002].

2.4. Simulation Procedure and Evaluation criteria

The model described in section 2.2 is used for evaluating the IEDSS performance. The simulations are conducted for the 332 days using dynamic conditions. Hence, it is possible to evaluate long term performance of the system. A set of evaluation criteria, defined by the IWA Task Group on benchmarking strategies and described in Copp [2002], is used to test the proposed IEDSS in the Vall del Ges WWTP. In this case the selected criteria are the effluent quality index (EQ), the aeration energy cost (AE), and the pumping energy cost (PE).

3. RESULTS

This section summarizes the simulation results and shows the plant performance comparing the current manual mode of operation with the proposed IEDSS. It is organized in the following way: Firstly the effluent quality is quantified averaging the concentration of the different pollutants leaving the plant. Secondly, both approaches are compared in terms of operation costs and effluent quality indices. Finally, the behaviour of the system using the IEDSS is shown under dynamic conditions demonstrating its adaptation to different plant perturbations i.e changes in the influent load, effect of the temperature.

Table 1 presents the averaged effluent concentrations when the proposed IEDSS is either included or not. The plant is in both cases operating close to the effluent optimum conditions and complies with the limits set by the European Directive (91/271/EEC). However, one can notice that the alternative with the IEDSS is marginally better than the other in terms of enhancing N/D, which is one of the goals of the study. Organic matter and solids are not a big issue here since the concentrations are much lower than the legislation limits.

Table 1. Simulated effluent concentrations (average of the 332 days simulation)

EFFLUENT CONCENTRATIONS	Without IEDSS	With IEDSS
average total N ($g \cdot m^{-3}$)	7.38	6.28
average total COD ($g \cdot m^{-3}$)	37.35	37.81
average TSS ($g \cdot m^{-3}$)	6.72	6.73
average BOD5 ($g \cdot m^{-3}$)	0.61	0.73

Figure 5 (top and middle) shows the dynamics of the manipulated variables according to the defined response surfaces and the decision matrixes (see Figure 4). Regarding aeration, K_La fluctuates between 13 and 67 d^{-1}, with an average value of 26 d^{-1}, which is lower than the open loop value of 40 d^{-1}. The internal recycle flow ranges from 3376 to 27997 $m^3 \cdot d^{-1}$, with an average value of 19106 $m^3 \cdot d^{-1}$. Again, this value is lower than the open loop constant value of 23760 $m^3 \cdot d^{-1}$ used for manual mode of operation, indicating that savings could be achieved.

Also, Figure 5 (bottom) shows the dynamic profile of the total nitrogen in the effluent with and without IEDSS. Nitrification efficiency is enhanced when using IEDSS because the

rotor provides the necessary oxygen and thus overcoming the apparent limitations of the plant running in an open loop regime in front of the plant perturbations e.g. decrease of the autotrophic activity during the winter periods because of the temperature. Further, the control of internal recycle improves the denitrification rates in the anoxic section, maximizing the conversion of nitrogen nitrate to nitrogen gas. Thus, the effluent total nitrogen is lower during almost the whole studied period when using the IEDSS. However, it is important to point that for days 180 and 290 the FLE is not capable to keep effluent nitrogen concentrations at the desired values. These periods are characterized for having a step increase in the influent nitrogen concentration but not a significant increase in the BOD (see Figure 1). Hence, although the control action is to increase the internal recycle, there is not enough organic carbon in the influent (the BOD/TKN ratio is less than 3) to be used as electron donor to denitrify the entire arriving nitrogen nitrate.

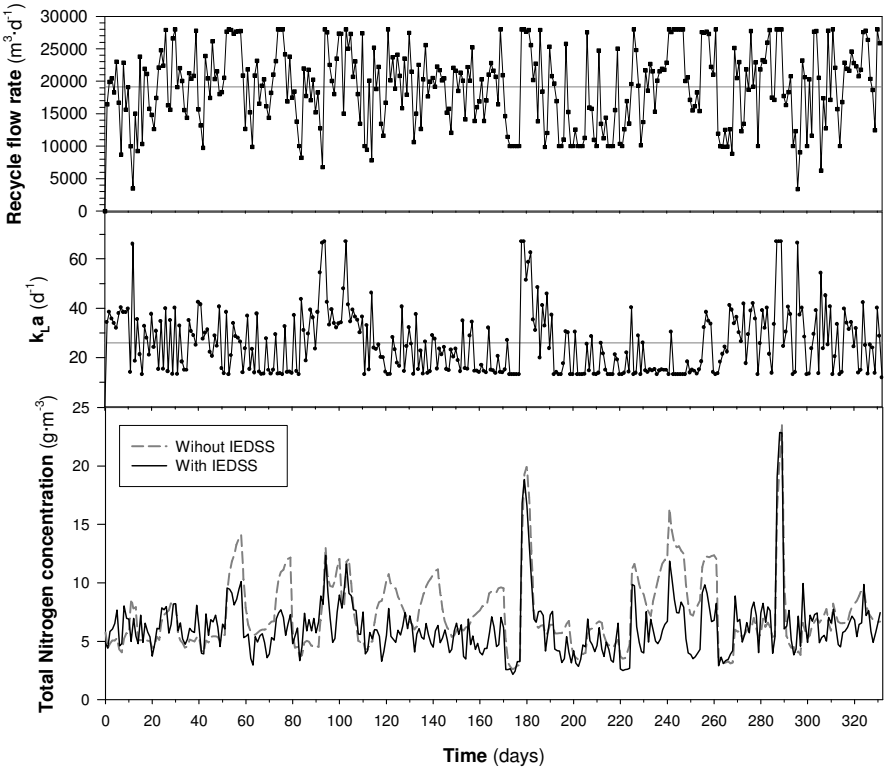


Figure 5. Effect of the IEDSS on K_La, internal recycle flow and effluent total nitrogen

The performance of the IEDSS is also assessed by the calculated performance indices presented in Table 2. An improvement of 10% is observed for the EQ index when implementing the IEDSS. Nitrogen ammonium and nitrogen nitrate are the pollutants with the highest impact on water body and for this reason they are weighted the most in the EQ calculation. In terms of economic objectives, it is important to point out the reduction of the operational costs that the implementation of the controller supposes. The aeration and the pumping energy costs are reduced by 37% and 14% respectively. Assuming a value of 0.12€/kW·h, the savings would be around 19000€·year^{-1}.

Table 2. Comparison of the calculated performance indices, with and without IEDSS

PERFORMANCE INDICES	Without IEDSS	With IEDSS	%
Effluent Quality (EQ) (Kg pollution units·d^{-1})	918	828	-10
Aeration energy cost (AE) (kW·h·d^{-1})	1137	718	-37
Pumping energy cost (PE) (kW·h·d^{-1})	135	116	-14

As a final remark, forthcoming work will be conducted for full-scale implementation of the IEDSS. Then, the limitations of the IEDSS performance due to model assumptions and deviations between measurements and model variables will be investigated.

4. CONCLUSIONS

This paper has presented the advantages of implementing an IEDSS to overcome the limitations of the current manual control of the WWTP of Vall de Ges. The proposed IEDSS integrates on-line information with knowledge management techniques and reduce the pollution load leaving the plant at the same time that keeps operational costs at minimum. Additionally, it must be pointed out a reduction of the operation costs in terms of aeration and pumping energy and the slightly improvement of the overall nitrogen removal. This is mainly due to a better adaptation to the different process disturbances (e.g. changes in the influent load, temperature...) using the knowledge extracted from both the experts and the deterministic simulations.

ACKNOWLEDGEMENTS

The authors would like to thank Depuradores d'Osona, S.L. for their valuable contribution.

REFERENCES

Ayesa, E., De la Sota, A., Grau, P., Sagarna, J.M., Slaterian, A. and Suescun, J., Supervisory control strategies for the new WWTP of Galindo – Bilbao: The long run from the conceptual design to the full-scale experimental validation, *Wat. Sci. Tech., 53*, 193-201, 2006.

Copp, J.B., *The COST Simulation Benchmark: Description and Simulator Manual*, Office for Official Publications of the European Community, Luxembourg, 2002.

Henze, M., Gujer, W., Mino, T. and van Loosdrecht, M.C.M., *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Scientific and Technical Report No. 9*, IWA Publishing, London, UK, 2000.

Hulsbeek, J.J.W., Kruit, J., Roeleveld, P.J. and Van Loosdrecht, M.C.M., A practical protocol for dynamic modeling of activated sludge systems, *Wat. Sci. Tech., 45* (6), 127-136, 2002.

Jung Kim, Y., Geol Lee, J., Hoon Lee, J., Young Lee, C., Gyun Choi, Y., and Hak Chung, T., Development of a real-time aeration control system based on ORP and DO for enhanced Nitrogen Removal, The 2nd IWA Conference on Instrumentation, Control and Automation, May 29 – June 2, Busan (Korea), 677-685, 2005.

Mecalf & Eddy, *Wastewater Engineering: Treatment, Disposal and Reuse*, 3rd edition . McGraw-Hill. New York, 2003.

Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A. and Steyer, J-P., *Instrumentation, Control and Automation in Wastewater Systems. Scientific and Technical Report nº15*, IWA Publishing, London, UK, 2005.

Poch, M., Comas, J., Rodríguez-Roda, I., Cortés, U. and Sànchez-Marrè, M., Designing and building real environmental decision support systems, *Env. Mod. Soft., 19*, 857-873, 2004.

Rieger, L., Alex, J., Winkler, S., Böhler, M., Thomann, M. and Siegrist, H., Progress in sen-sor technology – progress in process control? Part I: Sensor property investigation and classification, *Wat. Sci. Tech. 47*(2), 103-112, 2003.

Samuelsson, P. and Carlsson, B., Control of the aeration volume in an activated sludge process for nitrogen removal, *Wat. Sci. Tech., 45*, 45-52, 2002.

Takács, I., Patry, G.G. and Nolasco, D., A dynamic model of the clarification-thickening process. *Wat. Res. 25*(10):1263, 1991.

Thomann, M., Rieger, L., Frommhold, S., Siegrist, H. and Gujer, W., An efficient monitoring concept with control charts for on-line sensors, *Wat. Sci. Tech. 46*(4-5), 107-116, 2002.

Thomsen, H.A., Nielsen, M.K., Nielsen, E.H. and Hansen, N.P., Load dependent control of BNR-WWTP by dynamic changes of aeration volumes, *Wat. Sci. Tech., 37*, 157-164, 1998.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Genetic algorithm based technique for defining threshold regression models

J.C.M. Pires, F.G. Martins, M.C.M. Alvim-Ferraz and M.C. Pereira

LEPAE, Departamento de Engenharia Química, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
(dce05005/fgm/aferraz/mcsp@fe.up.pt)

Abstract: This study proposes a new technique based on genetic algorithms to define threshold regression models (TR-GA). The threshold regression assumes that the behaviour of the dependent variable changes when it enters in a different regime. The change from one regime to another depends of a specific value (threshold value) of an explanatory variable (threshold variable). In this study, the threshold regression models were composed by two linear equations. The application of genetic algorithms allows evaluating, at the same time: (i) the threshold variable; (ii) the threshold value; and (iii) the statistically significant regression parameters in each regime. The aim of this study was to evaluate the performance of TR-GA models in the prediction of next day hourly average ozone ($O_{3|d+1}$) concentrations. The considered predictors were hourly average concentrations of carbon monoxide (CO), nitrogen oxide (NO), nitrogen dioxide (NO_2) and ozone (O_3) and meteorological data (temperature - T, solar radiation - SR, relative humidity - RH and wind speed - WS). The studies were performed in the period from May 2004 to July 2004. Different TR-GA models were obtained, corresponding to different threshold variables and threshold values. Considering both training and test periods and comparing with multiple linear regression (MLR) approach, better performance indexes were achieved in four of these models. The model that presented the best results showed that the $O_{3|d+1}$ change its behaviour at the temperature of 23 ºC. For temperatures below 23 ºC, $O_{3|d+1}$ depended on CO, NO, NO_2, SR, WS and O_3, for higher temperatures, it depended on CO, NO, T, WS and O_3, while for MLR, the most important variables were NO, NO_2, SR, RH, WS and O_3.

Keywords: Threshold regression models; Genetic algorithms; Multiple linear regression; Ozone concentrations prediction.

1. INTRODUCTION

Ozone (O_3) is a photochemical oxidant found in different layers in the atmosphere. In the stratosphere, O_3 plays an important role for the protection of the human health, absorbing much harmful ultraviolet radiation. However, in the troposphere this reactive gas has negative impacts on human health, climate, vegetation and materials [Alvim-Ferraz et al., 2006; Chan et al., 1998]. Concerning human health, O_3 is responsible for the irritations in the respiratory system [Kley et al., 1999; Leeuw, 2000]. It is considered a greenhouse gas, being also associated to the increase of the temperature [Bytnerowicz et al., 2006]. Additionally, it is considered toxic to plants, reducing the crop yields [Alvim-Ferraz et al., 2006] and as a strong oxidant, it contributes for the increase of the corrosion of the materials [Leeuw, 2000].

O_3 is the result of the combination of three basic processes: (i) photochemical production by the interaction of hydrocarbons and nitrogen oxides (emitted by gasoline vapours, fossil fuel power plants, refineries, and other industries) under the action of suitable ambient

meteorological conditions [Guerra et al., 2004; Strand and Hov, 1996; Zolghadri et al., 2004]; (ii) tropospheric/stratospheric exchange that causes the transport of stratospheric air, rich in ozone, into the troposphere [Dueñas et al., 2002]; and (iii) horizontal transport due to the wind that transports ozone produced in other regions.

Many studies were done aiming the prediction of O_3 concentrations [Abdul-Wahab et al., 2005; Sousa et al., 2006 and 2007; Schlink et al., 2003]. Linear (multiple linear and principal component regressions) and nonlinear (neural networks, time series and generalised additive models) models were applied and no significant difference was found between their performances. In this study, O_3 concentrations were predicted using a threshold regression model composed by two linear equations. These models are not only applied to economy [Fouquau et al., 2007; Hansen, 1999; Lee and Chang, 2007], but also in biology [Gutzwiller and Barrow, 2003], medicine [Whitmore and Su, 2007] and environment [Cakmak et al., 1999]. The threshold regression assumes that the behaviour of the dependent variable changes when it enters in a different regime [Terui and Dijk, 2002]. The change from one regime to another depends of a specific value (threshold value) of an explanatory variable (threshold variable) [Fouquau et al., 2007]. The dependent variable is given by:

$$y_i = \begin{cases} \hat{\alpha}_0 + \sum_{i=1}^{k} \hat{\alpha}_i x_i + \varepsilon_1, \ if \ x_d \le r \\ \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \varepsilon_2, \ if \ x_d > r \end{cases} \quad (1)$$

where x_i (i=1,…,k) are the explanatory variables, $\hat{\alpha}_i$ and $\hat{\beta}_i$ (i=0,…,k) are the regression parameters, ε_1 and ε_2 are the errors associated with the regressions, r is the threshold value and x_d is the threshold variable (one of the explanatory variables) that determines the division of the original data in two parts [Terui and Dijk, 2002]. Multiple linear regression (MLR) was then applied to each part of the data and the regression parameters were determined through the minimization of the sum of square errors (SSE) [Pires et al., 2008]. In both regression equations, only the statistically significant regression parameters should be considered. The statistical significance of the regression parameters was evaluated through the calculation of their confidence interval. Thus, the parameter $\hat{v}_i$ is statistically significant if [Hayter et al., 2006]:

$$|\hat{v}_i| > \frac{t_{n-k-1}^{\alpha/2}\hat{\sigma}}{\sqrt{Sxx_i}} \quad (2)$$

where t is the Student t distribution, n is the number of points, k is the number of parameters, α is the significance level, $\hat{\sigma}$ is the standard deviation given by $\sqrt{SSE/(n-k-1)}$ and Sxx_i is the sum of squares related to x_i given by $\sum_{j=1}^{n}\left(x_{i,j} - \bar{x}_i\right)^2$.

Genetic algorithms (GA) were applied to the threshold regression model (TR-GA model), aiming to optimize the values of r (threshold value) and d (index of threshold variable) with the constraint of all regression parameters must be statistically significant. The main objective of this study was to evaluate the performance of TR-GA models in the prediction of the next day hourly average ozone ($O_{3|d+1}$) concentrations.

2. GENETIC ALGORITHMS AND TR-GA PROCEDURE

GA is a search algorithm based on the mechanics of natural selection and population genetics [Goldberg, 1989; Holland, 1975]. This method starts with a set of individuals (the population) chosen randomly. These individuals (also called chromosomes) have genes that represent a solution of a given problem. GA generates a sequence of populations (the generations) by applying the genetic operators (selection, crossover and mutation) to the individuals. GA presents the following advantages: (i) optimization with continuous or

discrete variables; (ii) derivative function not necessary; (iii) dealing with a large number of variables; (iv) optimization of variables with extremely complex cost surfaces; and (v) providing a list of optimal solutions (not just a single solution). Even that, GA is not the best method to solve all the problems. For example, traditional methods quickly find the solution of a well behave convex analytical function with few variables, while GA is still evaluating the initial population [Haupt and Haupt, 2004]. The optimizer should select the best method to solve the problem that has in hands. In this study, GA was selected due to the different type of parameters to optimize and the complexity of the constraints (ensure that all regression parameters are statistically significant).

The population size is the number of individual chromosomes that is presented in a population. A large number of chromosomes increases the population diversity, but it also increases the computation time due to the fitness evaluation step. Goldberg [1989] reported that the population size selected by many GA researchers usually ranges from 30 to 200. In this study, the population size was fixed to 100 chromosomes. Preliminary simulations showed that for this population size the number of generations should be high to achieve convergence. Thus, the number of generations was 500. Figure 1 shows the codification of chromosomes. Each chromosome was divided in four sub-strings that correspond to: (i) the value of *d*; (ii) the value of *r*; (iii) the explanatory variables used in the first regression (1 – consider the correspondent explanatory variable); and (iv) the explanatory variables used in the second regression (for $x_d > r$).

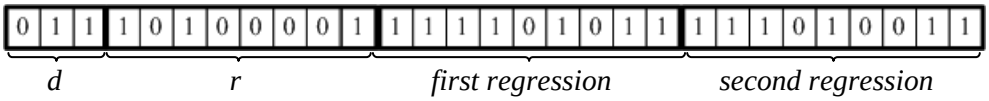


Figure 1. Codification of chromosomes.

The selection operator determines which chromosomes are used to generate the next population based on their fitness in the current generation (survival of the fittest). The fitness function measures the performance of the individual with respect to the particular search problem. The fitness function was defined as:

$$\arg\min f = \sqrt{\frac{SSE_1 \times 10^{ip_1} + SSE_2 \times 10^{ip_2}}{nl}} \quad (3)$$

where *ip* is the number of statistically insignificant regression parameters and *nl* is the number of the training points. The indexes *1* and *2* correspond to the first and second regressions, respectively. In many selection methods, the best solutions can be not selected to reproduce. Therefore, these solutions can be lost after the application of crossover and mutation. To avoid this situation, the best elements were copied to the next generation (elitism). However, this procedure decreases the population diversity in the next generations. To reduce this effect, all the chromosomes in the current generation had equal probability to be chosen by crossover and mutation procedures.

The crossover operator consists in exchanging genetic material (binary substrings) of two parents (two chromosomes of the current generation), creating two new individuals. High crossover rates increase the population diversity, promoting the mixing of chromosomes [Siriwardene and Perera, 2006]. The used crossover rate was 0.7.

The mutation operator consists in modifying the chromosomes at random. In bit string representation, the mutation is done by changing 0 to 1 and vice versa in one or more bits. High mutation rates increase the probability of destruction of the best chromosomes [Siriwardene and Perera, 2006]. The used mutation rate was 0.1.

Figure 2 shows how GA is applied for defining threshold regression models. First, the initial population is randomly created. Then, for each individual in the population, the values of *r* and *d* must be calculated to divide the initial data in two parts. Applying MLR to each part (taking into account only the explanatory variables selected by the chromosome), the

regression parameters are determined and also their statistical significance. After, the individual fitness is evaluated. Finally, the genetic operators are applied to create new individuals in the next generation. This procedure stops when the maximum number of generations is achieved.

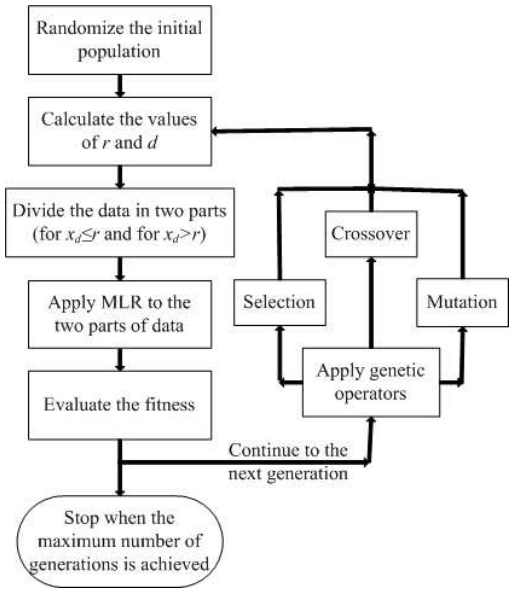


Figure 2. Procedure to apply GA for defining threshold regression models.

3. DATA

Tropospheric O_3 is a secondary pollutant, predominantly formed by photochemical reactions involving other air pollutants, under suitable meteorological conditions. Thus, the MLR and TR-GA models should present relationships between O_3 concentrations and environmental and meteorological variables.

The environmental data, hourly average concentrations of carbon monoxide (CO), nitrogen oxide (NO), nitrogen dioxide (NO_2) and O_3, were collected in an urban site (*Antas*) with traffic influences situated in Oporto, Northern Portugal (Figure 3). This site belongs to the air quality monitoring network of Oporto Metropolitan Area that is managed by the Regional Commission of Coordination and Development of Northern Portugal (*Comissão de Coordenação e Desenvolvimento Regional do Norte*), under the responsibility of the Ministry of the Environment. The meteorological data, temperature (T), solar radiation (SR), relative humidity (RH) and wind speed (WS), were recorded on the left edge of Douro River, at an altitude of 90 m approximately.

The analysed period was from May to July 2004. It was divided in the training (1 May 2004 to 15 July 2004) and test (16 to 31 July 2004) periods.

4. RESULTS AND DISCUSSION

Different TR-GA models were obtained corresponding to different threshold variables. As MLR was the model selected for each regression in TR-GA models, it was considered the basis for comparison for the achieved models. For all models, a t-test (with α=0.05) was performed to evaluate the statistical significance of the regression parameters. Table 1 presents the statistically significant regression parameters for TR-GA (1 to 6) and MLR models and corresponding root mean squared error (RMSE) values in the training period. The regression parameters $\hat{v}_i$ (i=1 to 8) correspond to CO, NO, NO_2, T, SR, RH, WS and



Figure 3. Location of the air quality monitoring site (*Antas*).

O_3, respectively. As all regression parameters were considered statistically significant, the fitness value (calculated in GA procedure) corresponded to the RMSE value in the training data. Therefore, all TR-GA models presented slightly better performances than MLR approach in the given period.

In almost TR-GA models, the MLR parameters were very similar (when validated at same time) to the parameters of the first regression of TR-GA models. Thus, the improvement of the achieved models was expected in the prediction of O_3 concentrations corresponding to their second regression.

For test period, the regression equations obtained in the training step were applied to predict the $O_{3|d+1}$ concentrations. The performance of the models was evaluated through the calculation of the commonly used statistical indexes: mean bias error (MBE), mean absolute error (MAE), RMSE, Pearson correlation coefficient (R) and index of agreement (d_2) [Chaloulakou et al., 2003; Gardner and Dorling, 2000; Sousa et al., 2006]. Table 2 shows the performance indexes presented by TR-GA and MLR models. MBE was always positive, showing that, in average, the predicted ozone concentrations were overestimated. The MAE, RMSE (absolute error measures), R and d_2 give a global idea of the difference between the observed and modelled values. Thus, slightly better model predictions were

Table 1. Statistically significant regression parameters for TR-GA (1 to 6) and MLR models and correspondent RMSE value in the training data

Model	$\hat{\nu}_0$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\nu}_3$	$\hat{\nu}_4$	$\hat{\nu}_5$	$\hat{\nu}_6$	$\hat{\nu}_7$	$\hat{\nu}_8$		RMSE
TR-GA1	42.1	1.7	-4.1	8.5		3.6		3.2	17.8	$if\ T \leq 23.0$	19.9
	22.6	-8.3	8.5		22.4			5.3	9.0	$if\ T > 23.0$	
TR-GA2	43.5	2.1			4.5	2.4		4.5	15.1	$if\ RH \leq 82.0$	20.0
	26.8		-5.4	14.5		3.7	13.7		17.2	$if\ RH > 82.0$	
TR-GA3	41.2		-3.9	7.3		6.1		5.4	12.5	$if\ O_3 \leq 75.0$	20.0
	77.2		16.4		10.6					$if\ O_3 > 75.0$	
TR-GA4	43.7	2.6	-5.1	8.2		3.5	-5.3	3.0	14.2	$if\ NO_2 \leq 72.3$	20.1
	40.2	-6.9		6.9		5.1		8.4	6.0	$if\ NO_2 > 72.3$	
TR-GA5	12.5				8.4	5.3	-15.6	4.6	10.0	$if\ RH \leq 55.7$	20.1
	43.3	1.7	-3.8	7.0		3.3	-3.9	3.4	14.9	$if\ RH > 55.7$	
TR-GA6	57.5	23.9		5.4	4.9	2.2	-4.0	4.3	15.7	$if\ CO \leq 304.4$	20.1
	44.7		-2.8	4.3		3.1	-2.2	3.8	14.4	$if\ CO > 304.4$	
MLR	42.6		-2.2	5.3		3.9	-3.7	3.5	13.9		20.7

obtained in four TR-GA models when compared to MLR. The TR-GA1 model presented the best results in both training and test periods. Figure 1 shows the codification of the correspondent chromosome. This model assumed that the $O_{3|d+1}$ behaviour changed at the temperature of 23 ºC. For temperatures below 23 ºC, $O_{3|d+1}$ depended on CO, NO, NO_2, SR, WS and O_3, while for higher values, it depended on CO, NO, T, WS and O_3.

Table 2. Performance indexes of the TR-GA and MLR models in the test period

Model	MBE	MAE	RMSE	R	d_2
TR-GA1	0.31	15.42	19.66	0.74	0.83
TR-GA2	1.61	15.63	19.77	0.74	0.83
TR-GA3	0.81	15.92	20.68	0.71	0.81
TR-GA4	0.14	16.01	20.94	0.70	0.80
TR-GA5	0.43	15.51	19.98	0.73	0.81
TR-GA6	1.66	15.45	19.80	0.74	0.83
MLR	0.25	15.59	20.29	0.73	0.81

The main differences of the regressions in TR-GA1 model were the incorporation of NO_2 concentrations and SR as important variables in O_3 formation for temperatures below 23 ºC, being T only important for temperatures above this threshold value. These observations could be explained by the relative influence of volatile organic compounds (VOC) and nitrogen oxides (NO_x) in O_3 formation. Seinfeld and Pandis [1998] presented the complex chemical reactions involved in this system and showed that there is a competition between VOC and NO_x for the hydroxyl radical (OH), very important in the O_3 formation. At high VOC to NO_x ratio, OH reacts with VOCs; otherwise, the NO_x reaction predominates. In general, increasing VOC concentrations means the appearance of more ozone. The effect of

NO_x concentration increase depends on the VOC to NO_x ratio (positive correlation for high ratios and negative correlation for low ratios) [Seinfeld and Pandis, 1998]. The temperature has a great influence on O_3 formation. High temperatures favour VOC to NO_x ratio because VOC concentrations increase more with temperature than NO_x concentrations. Accordingly, the results showed that temperature increase lead to higher O_3 concentrations (positive correlation between O_3 concentrations and temperature was observed in the second regression of TR-GA1 model). Simultaneously, as expected, a positive correlation between NO and O_3 concentrations was also observed. At lower temperatures, VOC to NO_x ratio decrease and O_3 concentrations are greatly dependent on NO_x concentrations. As the correspondent chemical reactions are catalysed by solar radiation [Seinfeld and Pandis, 1998], this variable was considered statistically significant in the first regression of TR-GA1 model, presenting a positive correlation with O_3 concentrations. Furthermore, as expected for lower VOC to NO_x ratio, a negative correlation between NO and O_3 concentrations was observed.

5. CONCLUSIONS

GA was applied to define threshold regression models for prediction of the next day hourly average ozone ($O_{3|d+1}$) concentrations. These models assume that the dependent variable changes its behaviour when an explanatory variable takes a specific value. Applying the procedure presented in this study, different TR-GA models were obtained corresponding to different threshold variables and threshold values. In both training and test periods, four of these models presented slightly better results than MLR approach. Additionally, the best model showed that $O_{3|d+1}$ changed its behaviour at the temperature of 23 ºC. For temperatures below that value, $O_{3|d+1}$ depended of CO, NO, NO_2, SR, WS and O_3, while for higher temperatures, it depended of CO, NO, T, WS and O_3.

ACKNOWLEDGEMENTS

Authors are grateful to *Comissão de Coordenação da Direcção Regional-Norte* and to *Instituto Geofísico da Universidade do Porto*, for kindly providing the air quality and meteorological data. This work was supported by *Fundação para a Ciência e Tecnologia* (FCT). *J.C.M. Pires* also thanks the FCT for the fellowship SFRD/BD/23302/2005.

REFERENCES

Abdul-Wahab, S.A., C.S. Bakheit, and S.M. Al-Alawi, Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations, *Environmental Modelling & Software 20(10)*, 1263-1271, 2005.

Alvim-Ferraz, M.C.M., S.I.V. Sousa, M.C. Pereira, and F.G. Martins, Contribution of anthropogenic pollutants to the increase of tropospheric ozone levels in Oporto Metropolitan Area, Portugal since the 19th century, *Environmental Pollution, 140(3)*, 516-524, 2006.

Bytnerowicz, A., K. Omasa, and E. Paoletti, Integrated effects of air pollution and climate change on forests: A northern hemisphere perspective, *Environmental Pollution 147*, 438-445, 2006.

Cakmak, S., R.T. Burnett, and D. Krewski, Methods for detecting and estimating population threshold concentrations for air pollution - related mortality with exposure measurement error, *Risk Analysis, 19(3)*, 487–496, 1999.

Chaloulakou, A., M. Saisana, and N. Spyrellis, Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. *The Science of the Total Environment, 313*, 1-13, 2003.

Chan, L.Y., H.Y. Liu, S. Lam, T. Wang, S.J. Oltmans, and J.M. Harris, Analysis of the seasonal behavior of tropospheric ozone at Hong Kong, *Atmospheric Environment 32(2)*, 159-168, 1998.

Dueñas, C., M.C. Fernández, S. Cañete, J. Carretero, and E. Liger, Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast, *The Science of the Total Environment, 299(1-3)*, 97-113, 2002.

Fouquau, J., C. Hurlin, and I. Rabaud, The Feldstein–Horioka puzzle: A panel smooth transition regression approach, *Economic Modelling, 25(2)*, 284-299, 2007.

Gardner M.W., and S.R. Dorling, Statistical surface ozone models: an improved methodology to account for non-linear behaviour, *Atmospheric Environment, 34*, 21-34. 2000.

Goldberg, D.E, *Genetic Algorithms in Search, Optimization and Machine Learning*, Kluwer Academic Publishers, Boston, MA, 1989.

Guerra, J.C., S. Rodríguez, M.T. Arencibia, and M.D. García, Study on the formation and transport of ozone in relation to the air quality management and vegetation protection in Tenerife (Canary Islands), *Chemosphere, 56*, 1157-1167, 2004.

Gutzwiller, K.J., and W.C. Barrow, Bird communities, roads and development: prospects and constraints of applying empirical models. *Biological Conservation, 113(2)*, 239-243, 2003.

Hansen, B.E., Threshold effects in non-dynamic panels: Estimation, testing, and inference, *Journal of Econometrics, 93(2)*, 345-368, 1999.

Haupt, R.L., and S.E. Haupt, *Practical Genetic Algorithms, second edition*, John Wiley & Sons, 27-49, New Jersey, 2004.

Hayter, A.J., H.P. Wynn, and W. Liu, Slope modified confidence bands for a simple linear regression model, *Statistical Methodology, 3*, 186-192, 2006.

Holland, J.H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.

Kley, D., M. Kleinmann, H. Sanderman, and S. Krupa, Photochemical oxidants- state of the science, *Environmental Pollution, 100(1-3)*, 19-42, 1999.

Lee, C.C., and C.P. Chang, The impact of energy consumption on economic growth: Evidence from linear and nonlinear models in Taiwan, *Energy, 32(12)*, 2282-2294, 2007.

Leeuw, F.A.A.M., Trends in ground level ozone concentrations in the European Union, *Environmental Science & Policy, 3(4)*, 189-199, 2000.

Pires, J.C.M., F.G. Martins, S.I.V. Sousa, M.C.M. Alvim-Ferraz, M.C. Pereira, Selection and validation of parameters in multiple linear and principal component regressions, *Environmental Modelling & Software, 23(1)*, 50-55, 2008.

Schlink, U., S. Dorling, E. Pelikan, G. Nunnari, G. Cawley, H. Junninen, A. Greig, R. Foxal, K. Eben, T. Chatterton, J. Vondracek, M. Richter, M. Dostal, L. Bertucco, M. Kolehmainen, and M. Doyle, A rigorous inter-comparison of ground-level ozone predictions, *Atmospheric Environment, 37(23)*, 3237-3253, 2003.

Seinfeld, J.H., and Pandis S.N., *Atmospheric Chemistry and Physics – from Air Pollution to Climate Changes*, A Wiley-Interscience Publication, USA, 1998.

Siriwardene, N.R., and B.J.C. Perera, Selection of genetic algorithm operators for urban drainage model parameter optimisation, *Mathematical and Computer Modelling, 44(5-6)*, 415-429, 2006.

Sousa, S.I.V., F.G. Martins, M.C. Pereira, and M.C.M. Alvim-Ferraz, Prediction of ozone concentrations in Oporto city with statistical approaches, *Chemosphere, 64(7)*, 1141-1149, 2006.

Sousa, S.I.V., F.G. Martins, M.C.M. Alvim-Ferraz, M.C. Pereira, Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environmental Modelling & Software, 22*, 97-103, 2007.

Strand, A., and O. Hov, The impact of man-made and natural NO_x emissions on upper tropospheric ozone: a two-dimensional modal study, *Atmospheric Environment, 30*, 1291-1303, 1996.

Terui, N., and H.K. Dijk, Combined forecasts from linear and nonlinear time series models, *International Journal of Forecasting, 18(3)*, 421-438, 2002.

Whitmore, G., Y. Su, Modeling low birth weights using threshold regression: results for U. S. birth data, *Lifetime Data Analysis, 13(2)*, 161-190, 2007.

Zolghadri, A., M. Monsion, D. Henry, C. Marchionini, and O. Petrique, Development of an operational model-based warning system for tropospheric ozone concentrations in Bordeaux, France, *Environmental Modelling & Software, 19(4)*, 369-382, 2004.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Optimum Water Allocation at Basin Scale Using Meta-Modeling

M. Shourian[a] **and S. Jamshid Mousavi**[b]

[a] *PhD. Candidate, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, (shourian@aut.ac.ir)*

[b] *Associate Professor, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic),Tehran, Iran, (jmosavi@aut.ac.ir)*

Abstract: Evolutionary optimization algorithms, in many cases, suffer from a high computational cost due to the high-fidelity simulation models performed for objective function evaluations. Meta-modeling is one of the useful approaches to overcome this problem in which a surrogate model, running of which is much faster than the exact model, is used in lieu of the simulation model. To build a meta-model it is required to use a function-approximation procedure by which the expensive simulation model is approximated. Artificial Neural Networks (ANNs), as meta-models, have shown different applicability in various engineering design problems. However, training ANNs needs enough input-output data (design of experiments) each of which is obtained by running the expensive simulation model. A methodology is presented in this study in which the problems of design of experiments, function approximation and function optimization are sequentially solved in a feedback loop so that a much fewer number of experiments is required for the task of function approximation. The proposed approach adaptively utilizes the information obtained from function optimization, finds the regions where more data are needed, updates the training data set to fill the space and sequentially improves the accuracy of the meta-model. The performance of the proposed approach is analysed using a optimization problem on a benchmark multi-modal mathematical function and a real-world water allocation optimization problem at basin scale.

Keywords: River Basin; Decision Support Systems; Optimization, Meta-Modelling; Neural Networks, Design of Experiments.

1. INTRODUCTION

Water resources systems optimization problems at basin scale, containing environmental systems optimization, due to high number of the systems components and variables and discontinuous, nonsmooth, nonlinear or nonconvex functions are complex and hard to solve. There are various challenging faces to deal with in these optimization problems arising. Various highly-related components in water resources systems make the simulation-based or evolutionary optimization approach a promising and useful method in solving these problems. But, high computational costs associated with the use of high-fidelity river basin simulation models pose a serious impediment to the successful use of evolutionary algorithms (EAs) to optimization of water resources systems design and operation. In such complex problems, EAs typically require thousands of function evaluations to locate a near optimal solution. Hence, when computationally expensive simulation models, common in water resources systems analysis, are used for simulating system's performance, the use of EAs may be computationally prohibitive. One approach to reduce the computational expense is to approximate the simulation model by a less

expensive meta-model among which artificial neural networks (ANNs) have been gaining significant attention.

In field of using meta-models, Jin et al. [2002] presented a framework for coupling EAs and neural network-based surrogate models. This approach uses both the expensive and approximate models throughout the search, with an empirical criterion to decide the frequency at which each model should be used. Yan and Minsker [2006] proposed a dynamic modeling approach, called adaptive neural network GA in which ANNs networks are adaptively and automatically trained directly within a GA to replace a time-consuming groundwater remediation simulation model. A more detailed survey of the state-of-the-art in meta-modeling can be found in Simpson et al. [2002].

In this paper, an adaptive sequentially space filling (ASSF) method is presented for training an ANN, with minimum possible design of experiments, which is used as the surrogate model in a function optimization procedure. Performance of the proposed methodology is demonstrated at first in a benchmark mathematical problem with many local minima and then the developed model is applied in solving the problem of optimum water allocation at upstream Sirvan river basin system in Iran as a real case study. In the considered case study, the trained ANN is used for replacing the MODSIM [Labadie, 1995] model, a generalized river basin DSS (decision support system), in order to achieve the best possible accuracy with the least experiments. The size of planned dams and water transfer systems in a river basin, as design variables, and the values of some operational variables regarding relative water allocation priorities are searched for by the particle swarm optimization (PSO) [Kennedy and Eberhart, 1995] algorithm as the optimization engine in which MODSIM or its approximation is used to simulate the river basin system operation. Results of three modeling tools including simulation-optimization without meta-modeling [Shourian et al., 2008], adaptive sequentially space filling (ASSF) and pure random data generating meta-modeling are analysed and compared in the mentioned problems.

2. ADAPTIVE SEQUENTIALLY SPACE FILLING META-MODELING

The formulation of an optimization model whose objective function evaluation requires performance of a simulation model may be represented as below, referred to as problem (1):

$$
\begin{aligned}
&\textit{Min } f(y) \\
&\textit{Subject to:} \\
&h_i(x) = 0, \qquad i = 1, \ldots, m \\
&g_j(x) \geq 0, \qquad j = 1, \ldots, n \\
&y = sim(x)
\end{aligned}
\tag{1}
$$

In problem (1), $f(y)$ is the objective function to be minimized and $h_i(x)$ and $g_j(x)$ are the equality and inequality constraints of the problem. In order to evaluate $f(y)$, in many cases, there is a need to use a high-fidelity, computationally expensive simulation model. Moreover, the mathematical functions and relationships in the simulation model, $sim(x)$, may not be algebraic functions suitable to optimization by gradient-based techniques. There are also many cases in which the simulation model is a black box expensive function in the optimization procedure. In that case the modeler may not have access to computer codes and routines (functions) of the model. It should be noted that function $f(y)$ may be multimodal with respect to decision vector x and also some of functions $h_i(x)$ or $g_j(x)$ may be nonsmooth or the feasible space of the problem may be nonconvex. Meta-heuristic and evolutionary optimization techniques are promising in solving these types of optimization problems as they can be easily linked with any simulation model without the need to have access to computer codes or details of the function $sim(x)$. However, meta-heuristic algorithms may require thousands of objective function ($f(y)$) evaluations to converge to a good solution. Each evaluation needs running the high-fidelity simulation model $sim(x)$. This may make solving problem (1) using a meta-heuristic optimization

algorithm computationally prohibitive. To address this difficulty, the incorporation of meta-models has been suggested. The purpose of meta-models is to approximate the relationship between *f(y)* and the vector of input decision variables *x* by using efficient mathematical models. Therefore, a function approximation technique may be used to solve problem (1) where the approximation function $\tilde{f}(y)$ is optimized instead of exact function *f(y)*. The optimization problem in this case referred to as problem (2) may be expressed as following:

$$Min \quad \tilde{f}(y) \tag{2}$$

Subject to:

$$h_i(x) = 0, \qquad i = 1, ..., m$$

$$g_j(x) \geq 0, \qquad j = 1, ..., n$$

In problem (2), $\tilde{f}(y)$ is used in lieu of the exact function *f(y)*. Hence, to solve main optimization problem (1), we need to determine the surrogate function $\tilde{f}(y)$. To have an accurate approximation the following condition should be satisfied:

$$\left| \tilde{f} - f \right| \leq e, \qquad \forall x \in X \tag{3}$$

where ε is the accuracy parameter of the approximation and *X* is the search space of problem (1). In order to obtain the required precision for function approximation used in optimization problem (2), the only possible approach is to design experiments such as to fill the search space uniformly. These experiments are used to construct the meta-model utilized as the function approximator (e.g. in case of using ANN, it is the data set used for training the network). Therefore, one could say that there is a need to perform a number of experiments, *D*, which means that the function *f(y)* should be evaluated *D* times. In other words, in order to get the approximate function $\tilde{f}(y)$ in problem (2), a set of experiments known as $S_D = \{(x_k, f(sim(x_k))\}, k=1, ..., D$, should be prepared, which requires evaluating function *f* for *D* times. This problem is referred to as design of experiments, or problem (3).

In a classical method of design of experiments, the set S_D is made by a random data generation process. This means that in order to cover the search space of the problem, *X*, with a required homogeneity, a uniform random data generation process is used to locate points of x_k and then for each point of x_k the function *f(sim(x))* is evaluated. Two main characteristics are important to determine the size of S_D: nonlinearity of function *f* and the dimension of the input vector *x*. The higher nonlinearity of function *f* and also the larger number of decision variables, the more experiments are needed to be designed for construction of the meta-model.

Finding a way to solve problem (2) with the minimum number of experiments is one of the main objectives of the problem solving approaches, which this paper intends to address. A methodology is used to design experiments for data training in which problems (1), (2) and (3) are not viewed as separate problems; but each of them sequentially takes advantages of the results of the previous ones.

The procedure starts at first with generating a small number of input-output data in the search space, just enough to construct an approximate surrogate model. The experiments are designed randomly and the function *f(sim(x))* is evaluated for each experiment by running the original simulation model (design of experiments and function evaluation, problem (3)). A meta-model, which is a multi-layer perceptron ANN in this paper, is trained using the data which can be used for approximating function *f* (function approximation that is referred to as problem (2a)). The minimum value of approximate function $\tilde{f}$ is obtained using the PSO algorithm (approximate function optimization, problem (2)). This value will be near to the minimum value of the exact function *f*, if the approximation procedure is effective. To verify this, one should evaluate the error of the

approximation (ε) between $\tilde{f}$ and f for the solution of problem (2). If ε is small enough, there is a high chance that the located minimum of $\tilde{f}$ in problem (2) is a near-optimal solution for the main problem (1); Otherwise, the trained meta-model has not been able to learn and accurately approximate the exact function f. If so, the solutions obtained for problem (2) are not accepted as the solutions of problem (1) and thus more experiments should be designed. These new experiments are added to the former set of experiments, S_D, and the procedure is iterated until the whole search space of the problem and especially the regions with good solutions of the main problem (1) are covered with the required accuracy. The relationships between the different parts of the problem solving procedure are shown in Figure 1.

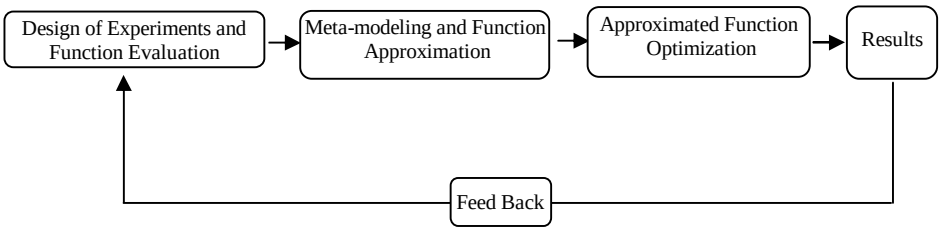


Figure 1. Relation between different parts of the proposed methodology for meta-modeling

As seen in Figure (1), the three sub-problems of (3), (2a) and (2) are solved in a feed-back loop such that solving each sub-problem is based on the information obtained from the other parts. The final procedure converges to find the solution of problem (1) using a meta-model with the minimum number of experiments.
Solving problem (2) finds an optimal solution of variables, x_*^j, for the surrogate function $\tilde{f}$. In order to accept the solution of the problem (2) as the answer of the problem (1) two important constraints must be satisfied. The first is that it is checked if there is a minimum pre-defined number of experiments in the training data set of S_D in neighborhood of x_*^j, as the constraint showing that the meta-model has found the optimal solution based on what it has learned from the search space of the problem. The second rule is that the exact objective value for x_*^j is calculated using the original simulation model. The error of approximation must be less than the required accuracy criterion. If these constraints are satisfied, x_*^j is accepted as a good optimizer for both $\tilde{f}$ and f and the meta-model has been able to approximate well the promising regions of the search space. Otherwise, x_*^j is within a gap of the search space which is not well estimated and it is required to design a number of experiments in a covering range to be included in the training set for the next construction of the meta-model. In this case, x_*^j is just stored and $\tilde{f}(x_*^j)$ in problem (2) is temporarily penalized to let the search engine (PSO) escape from x_*^j and search for other solutions. The next section describes the PSO algorithm used as the search engine to solve problem (2) within the proposed methodology just explained above.

3. PARTICLE SWARM OPTIMIZATION (PSO) ALGORITHM

The PSO, originally introduced in terms of social and cognitive behavior by Kennedy and Eberhart [1995], has proven to be a powerful competitor to other evolutionary algorithms. The PSO algorithm simulates social behavior among individuals (particles) "flying" through a multi-dimensional search space, each particle representing a single intersection of all search dimensions. The particles evaluate their positions relative to a goal (fitness) at every iteration. Particles in a local neighborhood share memories of their "best" positions and then use those memories to adjust their own velocities and positions as shown in

equations (4) and (5) below. The PSO algorithm defines each particle as a potential solution to a problem in a *D*-dimensional space, with the *i*th particle represented as $x_i = (x_{i1}, x_{i2}, ..., x_{iD})$. Each particle also remembers its previous best position, designated as *pbest*, $p_i = (p_{i1}, p_{i2}, ..., p_{iD})$, and its velocity, $v_i = (v_{i1}, v_{i2}, ..., v_{iD})$. In each generation, the velocity of each particle is updated, being pulled in the direction of its own previous best position (p_i) and the best of all positions (p_g) reached by all particles until the preceding generation.

The original PSO formulae were modified by Shi and Eberhart [1998] with the introduction of an inertia parameter, ω, that was shown empirically to improve the overall performance of PSO as follows:.

$$v_{id}^{n+1} = c\left(wv_{id}^{n} + c_1 r_1^{n}(p_{id}^{n} - x_{id}^{n}) + c_2 r_2^{n}(p_{gd}^{n} - x_{id}^{n})\right) \quad (4)$$

$$x_{id}^{n+1} = x_{id}^{n} + v_{id}^{n+1} \quad (5)$$

where $d = 1, 2, ..., D$; $i = 1, 2, ..., N$, and N is the size of the swarm; χ is a constriction factor which is used in constrained optimization problems in order to control the magnitude of the velocity (in unconstrained optimization problems it is usually set equal to 1.0); ω is called *inertia weight*; c_1, c_2 are two positive constants, called *cognitive* and *social* parameters respectively; r_1, r_2 are random numbers uniformly distributed in [0,1]; and $n = 1, 2, ...$, determines the iteration number.

4. FUNCTION OPTIMIZATION USING ADAPTIVE SEQUENTIALLY SPACE FILLING META-MODELING METHOD

4.1 Levy # 5 Function Optimization Problem

At first, to demonstrate the robustness of the proposed methodology, a well-known mathematical function is considered to locate its global minimum. The function is a notorious two dimensional test function, called the *Levy No. 5* [Levy et al., 1981] as follows:

$$f(x) = \sum_{i=1}^{5}[i\cos((i-1)x_1 + i)]\sum_{j=1}^{5}[j\cos((j+1)x_2 + j)] + (x_1 + 1.42513)^2 + (x_2 + 0.80032)^2 \quad (6)$$

where $-10 \le x_i \le 10$, and $i = 1, 2$. There are about 760 local minima and one global minimum with function value $f(x^*) = -176.1375$, at the point $x^* = (-1.3068, -1.4248)$. The large number of local minimizers makes it difficult for any method to locate the global minimizer. In Figure 2, the original plot of the *Levy No. 5* function, within the cube $[-10, 10]^2$, is shown.

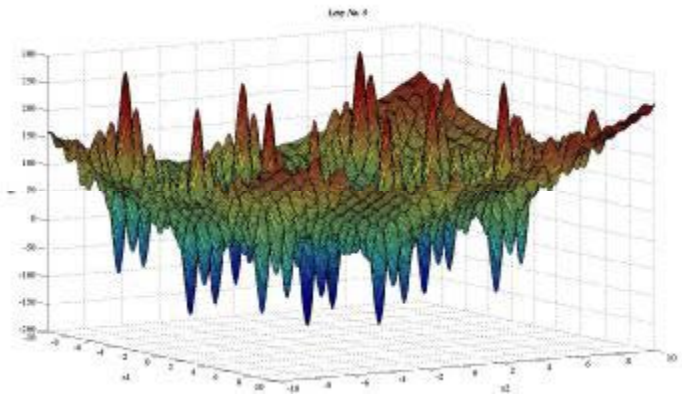


Figure 2. Plot of the *Levy No. 5* function

Results of using the ASSF meta-modeling for this bench-mark problem when compared to using meta-modeling with random design of experiments (DoE) process is represented in Table 1. Also, the solution obtained by the PSO algorithm using the exact function is presented. In this table, PSO-LEVY~ANN stands for optimization of the approximated function of Levy with the ANN.

Table 1. results of the PSO-LEVY and PSO-LEVY~ANN models

Item	PSO-LEVY	PSO-LEVY~ANN	
		ASSF DoE	Random DoE
Minimum exact objective function f^*	–176.1375	-175.9248	-2.5892
No. of training data	-	2’170	20’000
x^*_1	-1.3068	-1.3228	-1.8829
x^*_2	-1.4248	-1.4233	-1.316

As seen in Table 1, the ASSF meta-modeling method is able to locate a near optimal solution just using a number of 2’170 experiments for training the ANN while the network trained by 20’000 random data is not able to obtain a good solution. Also, it is observed that the PSO algorithm is able to locate the global solution of the problem.

4.2 Basin Scale Water Allocation Optimization Problem

Optimum water allocation planning in the upstream Sirvan river basin system in Iran has been considered as the case study problem in this paper. Excess surface water resources in the basin and growing water uses in neighboring basins have made important the subject of water transfers and water allocation. In the proposed development plan for the basin, construction of dam storage and water transfer projects has been considered. Deciding on the size of the project and how water resources should be allocated over time and space considering coordinated operation of the system components are some of the challenging issues in management of the basin. They could be dealt with by formulating a large scale optimization model for the Sirvan river basin system.

There are 9 reservoirs in the basin. Among the reservoirs, Zhaveh, Palangan and Gerdalan (S9-1) reservoirs are not constructed yet and they have to be sized. Also, relative priorities of target reservoir storages are operational variables through which the network flow program in MODSIM decides whether water stored in reservoirs in a time period should be released to meet water demands of that time period or it should be kept in the reservoirs for future uses. These priorities indicate the significance of storing in or releasing water from reservoirs in every time step in comparison with the priorities of the water demands. Therefore, there are two main sets of decision variables: (i) design variables, which include the sizes of planed dams and water transfer systems and (ii) operational variables, which include the relative priorities of target reservoir storage volumes. This optimization model in which MODSIM is linked with the PSO algorithm, PSO-MODSIM, has the form of problem (1) in which optimization the main function of f (the net benefit of the system) requires the performance of the expensive simulation model of MODSIM. Details of integrating MODSIM and PSO and structure of the PSO-MODSIM model and its application to optimal water allocation at basin scale could be found in Shourian et al. [2008].

High computational cost of the PSO-MODSIM model is the main prohibitive factor for its application to real problems. The procedure needs to call and perform a MODSIM simulation for each searching particle, which results in the total of 3000 to 4000 times of MODSIM runs with a PSO model using 20 particles. For instance, the computational time required by an up-to-date PC for solving the optimization model of the upstream Sirvan river basin over a 46 years planning horizon exceeds 20 hours. Therefore, using a meta-model to replace the high-fidelity MODSIM DSS is followed to reduce the computational cost of PSO-MODSIM model. An ANN is used as the surrogate model for substituting MODSIM in the PSO-MODSIM model. Error of training the ANN is one of the most affecting parameters in performance of the PSO-MODSIM~ANN model (~ indicates

approximation of MODSIM by ANN). Because of using an ANN as the meta-model in the PSO algorithm, it is desired to minimize training error of the ANN. Obviously, it is more beneficial to reach the required accuracy for the surrogate model with the minimum possible number of experiments used for training.

Each pair of $(\underline{x}_k , f(sim(\underline{x}_k)))$ ($\underline{x}_k$ is the vector of variables) in the training set S_D requires MODSIM to be executed. The sample training data set may be randomly generated. In this way, it is not guaranteed that the search space is covered appropriately unless the space is filled with a large number of generated data sets. Therefore, adaptive sequentially space filling meta-modeling presented in the second section may be employed to obtain the required surrogate model with the minimum number of experiments.

In the considered problem, capacities of Zhaveh, Gerdalan and Palangan reservoirs as design variables and priorities of the Gavshan, Zhaveh, Palangan, Azad and Gerdalan reservoirs are considered as the operational decision variables, resulting in an 8-dimensional search space. Firstly, the PSO-MODSIM model was performed and its solution was stored for comparison purposes. Then, ANN model was constructed to replace MODSIM using 100 random exemplars of the decision variables and their associated objective values for which MODSIM has been run. Subsequently, the PSO-MODSIM~ANN model was employed in which the trained ANN is called instead of MODSIM. Although the ANN model had been trained well in terms of both training and validating errors, the PSO-MODSIM~ANN model found solutions around which there were no experiments in the training set. As a result, the solutions were taken as gap representatives at which function $\tilde{f}$ is not adequately correct and therefore more experiments are needed to fill the gaps. By using the proposed ASSF technique and increasing the number of exemplars generated around the gaps found by the PSO to 1890, the PSO-MODSIM~ANN located a near optimal solution in comparison with the answer found by the PSO-MODSIM model. The new solution, whose exact objective value (using the original MODSIM simulation model) was very close to that reported by PSO-MODSIM~ANN, was also very close to the best objective value of the PSO-MODSIM model. As described before, once the PSO algorithm finds a solution it is checked if it is a gap representative. If so, its objective value is changed temporarily such that the PSO does not come back to this solution and then the search process is continued to locate another solution. Gradually, the number of solutions in gaps decreases and finally the PSO converges to a solution which satisfies both of the described constraints to be accepted as the optimum solution for the problem. This shows that the search space has been reasonably covered and there is no need to generate extra experiments.

In order to survey the robustness of the ASSF methodology, the PSO-MODSIM~ANN model was performed again, but using a pure random data generation process to prepare the training set. This means that using an initial set of 100 data points the ANN was trained and the PSO-MODSIN~ANN model run was performed. The solution was erroneous and thus a new set of experiments was designed randomly, instead of using the ASSF method, and added to the former set of training data. This procedure was continued until the difference between values of $\tilde{f}$ and f for the best solution found by the PSO-MODSIM~ANN model met the required accuracy criterion. In this case, it was seen that by designing 20000 random experiments the algorithm was able to find a good solution. Table 2 reports the solutions found by the proposed models and the number of data used in training the meta-models.

Table 2: results of the PSO-MODSIM and PSO-MODSIM~ANN models

Item	PSO-MODSIM	PSO-MODSIM~ANN	
		ASSF DoE	Random DoE
Minimum exact objective function f^*	808	811	822
No. of training data	-	1890	20000
Zhaveh capacity (MCM)	247.85	297.97	304.53
Gerdalan capacity (MCM)	253.8	253.8	247.8
Palangan capacity (MCM)	80	80	80

Gavshan priority	24	24	16
Zhaveh priority	25	24	25
Palangan priority	25	25	24
Azad priority	25	25	25
Gerdalan priority	30	30	30

It is seen random design of experiments for meta-modeling fails to find solutions as good as the ASSF method. The process of random data generation was stopped after generating 20000 experiments as the reported solution is close to the best solution found by the PSO-MODSIM model. However, the ASSF method has been able to locate a good solution (less than 1% difference in terms of objective function value) after only 1890 experiments which is about 9.45% of the number of experiments generated randomly. The value of the objective function reported in Table 2 has been calculated by MODSIM simulating the solution obtained by the approximate model. The approximate objective values evaluated by ANN are equal to 813.03 and 794.1 for sequential and random DoE models, respectively.

5. CONCLUSIONS

In this paper, a methodology for minimizing the required number of experiments for constructing a meta-model, especially in high-dimensional optimization problems, is presented. Three sub-problems of function evaluation, function approximation and function optimization are linked with each other and take advantage from each part's results in order to locate the regions in the search space of the optimization problem which needs to design more experiments. The proposed methodology is applied to a benchmark mathematical optimization problem and then to a basin scale optimum water allocation problem. The results show that the ASSF (adaptive sequentially space filling) model performs significantly better than random data generation-based meta-modeling. The ASSF meta-modeling is able to find solutions which are very close or even the same as the ones of the exact simulation-optimization model with less than 10% of experiments required in random generating based meta-modeling. The generalization aspect of the proposed methodology to be applied in any problem of evolutionary optimization with approximate fitness functions makes it an effective tool for solving large scale water systems optimization problems which contain environmental problems as a within subject.

REFERENCES

Jin, Y., Olhofer, M. and Sendhoff, B., A Framework for Evolutionary Optimization with Approximate Fitness Functions, *IEEE Transactions on Evolutionary Computation*, 2002.

Kennedy, J. and Eberhart, R. C., Particle Swarm Optimization, *Proceedings of IEEE International Conference on Neural Networks, IV*, 1942-1948, IEEE Service Center, Piscataway, NJ, 1995.

Labadie, J., MODSIM: River Basin Network Flow Model for Conjunctive Stream-aquifer Management, *Program User Manual and Documentation, Colorado State University*, 1995.

Levy, A., Montalvo, A., Gomez, S. and Galderon, A., Topics in Global Optimization, *Springer-Verlag*, New York, 1981.

Shi, Y. and Eberhart, R.C., A modified Particle Swarm Optimizer, *Proceedings of the 1998 IEEE Conference on Evolutionary Computation*, AK, Anchorage, 1998.

Shourian, M., Mousavi, S. J. and Tahershamsi, A., Basin-wide Water Allocation Planning by Integrating PSO Algorithm and MODSIM, *Water Resources Management, In press, available* at http://www.springerlink.com., 2008.

Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N. and Yang, R.-J., Approximation Methods in Multidisciplinary Analysis and Optimization: *A Panel Discussion, Proceedings of the Third ISSMO/AIAA Internet Conference on Approximations in Optimization*, October 14-25, 2002.

Yan, S., and Minsker B., Optimal groundwater remediation design using an Adaptive Neural Network Genetic Algorithm, *J. Water Resources Research*, 42, W05407, doi: 10.1029/2005WR004303.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Machine Learning Algorithms for GeoSpatial Data. Applications and Software Tools

M. Kanevski, A. Pozdnoukhov, V. Timonin

Institute of Geomatics and Analysis of Risk (IGAR), Faculty of Geosciences and Environment, University of Lausanne. Amphipole building, 1015 Lausanne, Switzerland
(Mikhail.Kanevski@unil.ch)

Abstract: Nowadays machine learning (ML), including Artificial Neural Networks (ANN) of different architectures and Support Vector Machines (SVM), provides extremely important tools for intelligent geo- and environmental data analysis, processing and visualisation. Machine learning is an important complement to the traditional techniques like geostatistics. This paper presents a review of several contemporary applications of ML for geospatial data: regional classification of environmental data, mapping of continuous environmental and pollution data, including the use of automatic algorithms, optimization (design/redesign) of monitoring networks.

Keywords: Machine learning algorithms; Spatial predictions and mapping; Software tools.

1. INTRODUCTION

One of the very important problem which is faced nowadays is how to handle, to understand and to model the data if there are too many or too few of them. Significant problems are arisen while dealing with large data bases or long period of observation, e.g. pattern recognition, geophysical monitoring, monitoring of rare events (natural hazards), etc. The major problems in such case are how to explore, analyse and visualise the oceans of available information. Nowadays machine learning (ML), including Artificial Neural Networks (ANN) of different architectures, and Support Vector Machines (SVM), are extremely important tools for intelligent geo- and environmental data analysis, processing and visualisation.

Several important applications of Machine learning algorithms for geospatial data are presented in the paper: regional classification of environmental data, mapping of continuous environmental data including automatic algorithms, optimization (design/redesign) of monitoring networks. ML is an important complement to the traditional techniques like geostatistics [Cressie, 1993; Chiles and Delfiner, 1999; Kanevski et al 2004]. They are nonlinear, adaptive robust and universal tools for patterns extractions and data modelling. Therefore they can be easily implemented in environmental decision support systems as data-driven modelling tools. In general, geospatial data are not only data considered in a geographical two- or three-dimensional space but data in a high-dimensional spaces composed of geo-features (see below)

In the present paper only some principal tasks and applications are considered along with the presentation of corresponding software tools. Theoretical topics and corresponding methodological details can be found in the corresponding references.

2. MACHINE LEARNING FOR GEOSPATIAL DATA

First, let us mention some typical characteristics of geospatial phenomena and environmental data: nonlinearity (linear models have limited applicability); spatial and temporal non-stationarity, i.e. in many cases hypotheses of spatio-temporal stationarity (second-order stationarity, intrinsic hypotheses) can not be accepted; multi-scale variability (high variability at several geographical scales), presence of noise and extremes/outliers; multivariate nature, etc. These "particularities" violate applications of traditional methods (including many geostatistical models) and highly complicates analysis, modelling and visualisation of geo- and environmental data. As it was mentioned above, in many real situations problems have to be considered in a high-dimensional feature (geo-features) spaces (very often the dimension of this space can be more than 10). It includes original geographical space and many features derived from science-based models or additional sources of information, for example, remote sensing images; slope, curvature, etc. derived from digital elevation models. In the latter case traditional (geo)statistical models either are too complicated to be applied or it is not possible to apply them. For example, the variography can be applied efficiently in the space of the dimension less than 3. Therefore, the important questions of spatial and in general spatio-temporal data analysis and modelling (including predictions and forecasting) deal with the development and implementation of data-adaptive, nonlinear, robust and multivariate models working in high dimensional spaces and having good generalisation properties.

One of the possible solutions can be based on machine learning algorithms, in particular, Artificial Neural Networks of different architectures and Statistical Learning Theory (e.g. kernel-based methods: Support Vector Machines, Support Vector Regression, etc.). Let us mention, that such approaches, being a data driven (black/grey boxes) highly depend on the quality and quantity of data. Therefore, it is useful and necessary to apply different statistical/geostatistical tools to control the quality of data analysis and modelling using ML. For example, variography helps to understand and to model spatial anisotropic correlations, spatial trends, local variability and the level of noise.

There are many resources available on machine learning algorithms including theoretical tutorials, scientific publications, and software tools. The theoretical topics applied in the present research are covered at a good level in recently published books [Bishop, 2007; Haykin, 1998; Hastie et al, 2001; Vapnik, 1999; Shawe-Taylor and Cristianini, 2004]. Internet site www.kernel-machines.org contains information and references on kernel-based data modelling; mloss.org is a site on machine learning open source software modules. Good tutorials on statistical data mining can be found on www.autonlab.org/tutorials/list.html; on-line proceeding and conference tutorials of NIPS conferences (Neural Information Processing Systems) are available, at nips.cc. A site www.cs.waikato.ac.nz/~ml/weka/ is dedicated to Weka software which is a collection of machine learning algorithms for general data mining tasks. Taking into account the importance of environmental applications recently two special issues of Neural Networks journal were devoted to earth sciences and environmental applications [Cherkassky et al., 2006; Cherkassky et al. 2007].

2. 1 Geospatial Data Analysis Tasks

In general, there are three fundamental tasks of statistical learning from data: classification, regression, and probability density modelling. Another two basic problems are of great importance: monitoring networks design/redesign and assimilation/integration of data and science-based models, e.g. physical pollution diffusion models, meteorological models, etc.

Usually learning machines are universal tools, i.e. in principle they can model any mapping (either categorical or continuous data) with any desired precision. The problem is how to select good structure of the machine (for example, multilayer perceptron) and how to tune its parameters. In machine learning there are several approaches for hyperparameters tuning, e.g. splitting of data, cross-validations, etc. For geospatial data geostatistical tools, for example, variography is a valuable tool to control the quality of machine learning procedures and parameters tuning [Kanevski and Maignan 2004].

Now, let us enumerate some typical geospatial data analysis problems and corresponding approaches/methods which can be used to solve them:

- Spatial predictions/interpolations: deterministic interpolators, geostatistics, machine learning. Spatial predictions in a high dimensional geo-feature space – machine learning.
- Modelling and spatial predictions with uncertainties (e.g. taking into account measurement errors): geostatistics, machine learning.
- Multivariate joint predictions of several variables: geostatistics (co-kriging), machine learning (multi-task learning).
- Risk mapping – modelling of local probability density function: geostatistics (indicator kriging, simulations), machine learning (Mixture Density Networks).
- Modelling of spatial variability and uncertainty, conditional simulations (spatial Monte Carlo simulations): geostatistical conditional stochastic simulations (sequential Gaussian simulations, indicator simulations, etc.).
- Optimisation of monitoring networks (spatial sampling design/redesign): space filling models, geostatistics (kriging, simulations), machine learning (Support Vector Machines). The basic idea of using Support Vector Machines for spatial sampling design is that only support vectors are important measurement points contributing to the solution of mapping problem [Vapnik, 1999].
- Data mining in a high dimensional geo-feature space: machine learning (supervised and unsupervised learning algorithms).

There are several important steps in performing spatial predictions of categorical and continuous data which are briefly described in the next section.

2.2 Methodology

The generic methodology of spatial data analysis and modelling is presented in Figure 1. As usually, exploratory spatial data analysis (ESDA) is a first step of the study. Quantitative analysis of monitoring networks using topological, statistical and fractal measures helps to describe data representativity, to remove biases in modelling distributions and to select declustering procedures [Kanevski and Maignan 2004].

The variography (well known geostatistical tool to analyse and to model anisotropic spatial correlations) is proposed to be used both at the phase of exploratory spatial data analysis and at the evaluation of the results. Despite of a variogram is a linear two-point statistics, like auto-covariance function for time series, it characterises the presence of spatial structures, anisotropy and scales [Chiles and Delfiner 1999; Cressie 1993]. Variogram analysis of the residuals, in addition to the traditional statistical analysis, is an important step in understanding the quality of modelling results: variograms of the residuals should demonstrate pure nugget effect (i.e. no spatial structures) on training and validation data sets. Variography can be used as an independent tool during tuning of machine learning hyper-parameters. In this case the cost function can be modified taking into account the difference between desired theoretical variogram based on data and a variogram based on ML results.

In fact, hybrid models (machine learning + geostatistics), e.g. Neural Network Residual Kriging/Co-kriging (NNRK/NNRCK), have proven their efficiency in many real-world mapping problems [Kanevski and Maignan, 2004]. They can overcome some difficulties of both approaches: in geostatistics – problem of spatial stationarity; in machine learning - interpretability. Moreover, such models are well adapted to multi-scale mapping of highly variable spatial data. Currently, new methodology which takes into account several aspects of geospatial data mentioned above and geographical/spatial constraints (geomorphology, networks, DEM, GIS thematic layers. etc.) is under development (see www.geokernels.org).

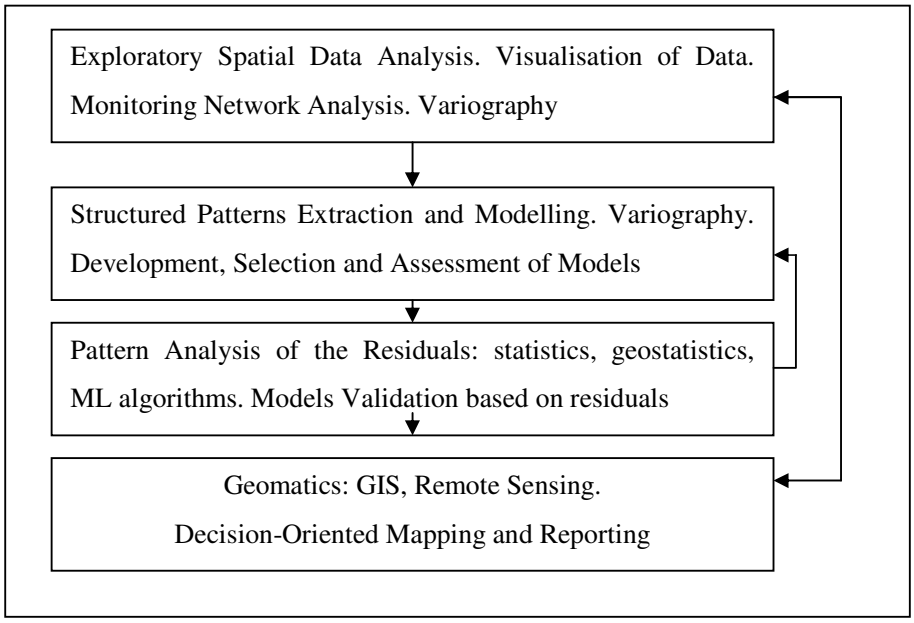


Figure 1. Spatial data analysis and predictions: generic methodology.

Let us consider some examples of machine learning application for spatial data. For the visualisation purposes mainly two-dimensional data are exploited.

2.3 Neural Networks for Environmental GeoSpatial Data

As it was mentioned above the first step deals with the exploratory data analysis and visualisation of raw and transformed (if necessary) data. At this step we propose also to use a non-parametric k-nearest neighbour method as a benchmark for data modelling and to check the availability of structured patterns [Kanevski et al., 2008]. Usually cross-validation (leave-one-out) is used to find the optimal k number. K-NN model can be used in a high dimensional space as well.

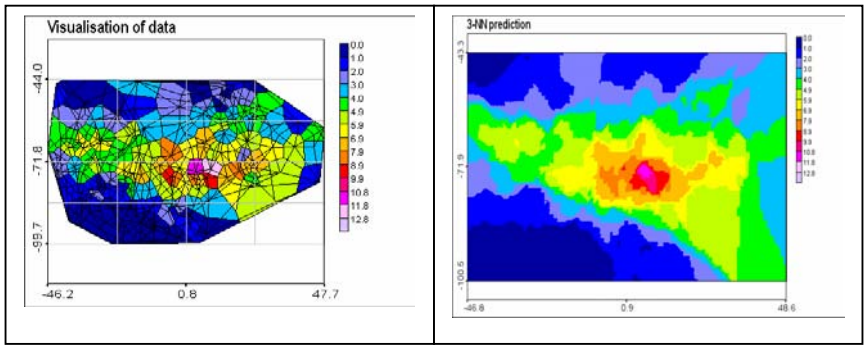


Figure 2. Visualisation of data using Voronoi polygons (left) and k-NN optimal modelling (k=3, see below).

In Figure 2 (left) raw data are visualised using Voronoi polygons, and Delaunay triangulation in order to visualise the topology of the monitoring network. The optimal number of k-NN modelling equals to 3. The result of k-NN mapping is given in Figure 3 (right). In a more general content, k-NN can be proposed to be used instead of the

variography to control the quality of mapping by ML algorithms: theoretically k-NN cross-validation curve has no minimum when data/residuals are not correlated.

Next important step deals with the quantitative analysis of monitoring networks and its clustering properties. This phase of the analysis is important in order to: 1) characterise a spatial resolution and topological properties of monitoring networks; 2) to characterise dimensional resolution (fractal dimension of monitoring network), 3) to characterise spatial representativity of data and corresponding biases, and 4) to define declustering procedures for clustered monitoring networks [Kanevski and Maignan, 2004]. In fact, the problem clustering and not i.i.d.-ness of data is an open question for the future research.

Recently great attention was paid to the possibility of on-line automatic mapping using different techniques. One the most efficient approach to solve such tasks is based on General Regression Neural Networks – GRNN [Kanevski and Maignan 2004; Kanevski et al., 2008]. The procedure of automatic tuning of anisotropic GRNN (based on Mahalanobis distances) is presented in Figure 3 (left), and corresponding optimal mapping in Figure 3 (right).

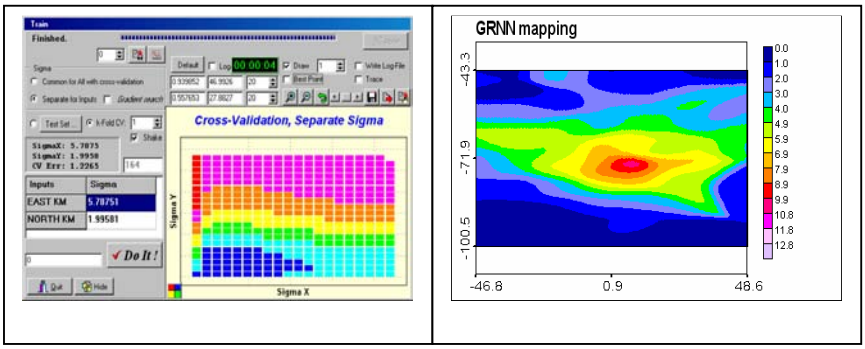


Figure 3. Automatic mapping using General Regression Neural Networks: training (left), optimal mapping (right).

In order to check the quality of modelling let us apply to proposed tests based on the residuals: variography and k-NN modelling. The directional variograms (variograms computed in several directions) are shown in Figure 4 (left). A straight solid line corresponds to a priori variance of the training residuals. All variograms demonstrate pure nugget effect with fluctuations around a priori variance, i.e. no overfitting. The k-NN test – cross-validation curve of the training residuals is presented in Figure 4 (right). There is no minimum on the curve, which demonstrates the absent of spatially structured pattern in these residuals. Therefore, all structured information was extracted form the data by GRNN model.

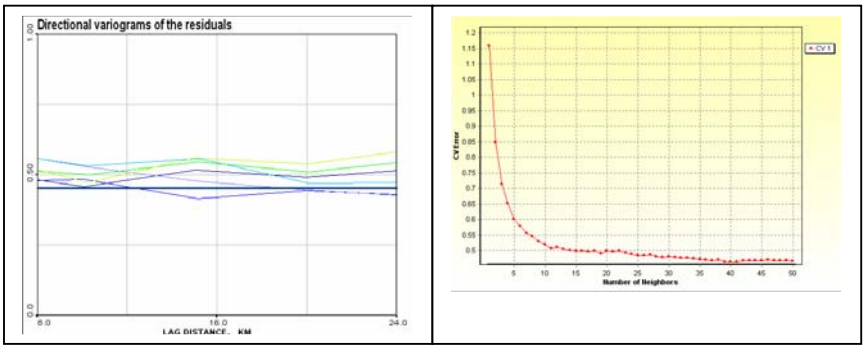


Figure 4. Variography of the residuals (left). K-NN cross-validation curve of the residuals (right).

k-NN test is a simple and efficient test but it does not provide details about patterns and their structures if they are present. It discriminates only between presence/absence of the structured patterns in the residuals. Variography is more powerful approach which takes into account anisotropies and provides detailed information about spatial correlations if they are present in the residuals. They can be considered as complementary diagnostic tools. Drawback (comparing with k-NN) is that dimension of the data should be less 3. Finally, let us mention that GRNN model itself can be used as well: if there are no spatial structures GRNN cross-validation curve has no minimum as well (when there are no spatial structures – in fact no space, the only prediction possible is a mean value of data).

2.4 Statistical Learning Theory for Environmental Spatial Data

Statistical Learning Theory (Vapnik-Chervonenkis theory) has a solid mathematical background for dependencies estimation and predictive learning from finite data sets. The workhorse of statistical learning theory – Support Vector Machines (SVM) – is based on the structural risk minimisation principle, aiming to minimise both the empirical risk and the complexity of the model, thereby providing high generalisation abilities. SVM provides non-linear classification and regression by mapping the input space into a higher-dimensional feature space using kernel functions, where the optimal solutions are constructed. The theoretical details on Statistical Learning Theory and corresponding models can be found in [Vapnik, 1998].

During recent years it was shown that SVM modelling of geospatial data has a great potential, especially when data are nonlinear, high-dimensional, and noisy.

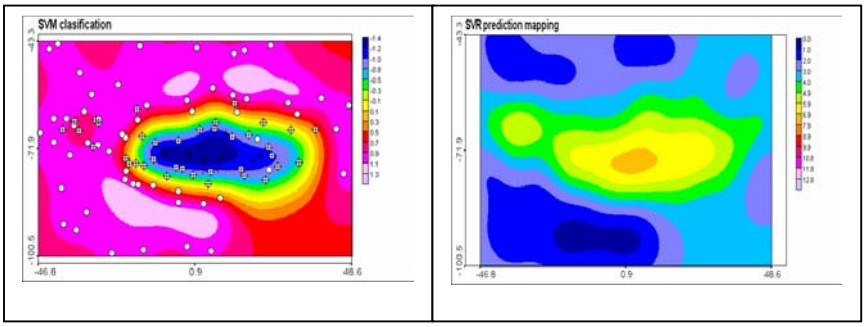


Figure 5. Support Vector Machines for geospatial data: decision function for two class spatial classification problem (white circles - support vectors) (left). Spatial regression/mapping of soil pollution data (right).

It was shown that SVM has good generalisation (predictability of validation data) properties on geospatial data analysis and modelling. Examples of the results of spatial predictions (classification and mapping) using SVM are given in Figure 5. Only 56% of data are support vectors which contribute to the solution. Other data points do not contribute to the decision boundary definition. Final two-class classification is carried out by taking sign[decision_function].

An important new application of SVM for geospatial data was proposed in [Pozdnoukhov and Kanevski, 2006]. It deals with monitoring networks design/redesign based on the properties of sparseness of SVM. Only Support Vectors are important data points contributing to the solution. The task is to find potential positions of Support Vectors and to select them as a new measurement points. From the learning point of view this problem can be considered as an active learning task.

An important contemporary developments concern semi-supervised or manifold learning. During following years machine learning will considerably contribute to modelling of high dimensional (geo-feature space of dimensions more than 10) nonlinear phenomena in the earth and environmental sciences. Such high-dimensional geospatial data are typical for

topo-climatic modelling and mapping, natural hazard analysis and risk susceptibility mapping (landslides, avalanches, forest fires, etc.), assessment of renewable resources (e.g. wind-power and solar engineering).

In a high dimensional space when the number of data is limited there is an important problem concerning the curse of dimensionality [Hastie et al, 2001; Haykin, 1998]. Therefore and important questions dealing with dimensionality reduction (in general nonlinear) should be considered and corresponding methods and tools selected [Lee and Verleysen, 2007]. Moreover, methods like Support Vector Machines, which are not sensitive on the dimension of the space, are preferable [Vapnik 1998]. Therefore generic methodology should be correspondingly modified taking into account the dimension of space and geo (natural) manifolds [GeoKernels, 2008, Kanevski et al., 2008].

3 SOFTWARE TOOLS

An extremely important part of intelligent data analysis using ML algorithms concerns software tools. Implementation of algorithms and development of software tools is an important step in machine learning studies. At present there are many ML software modules both commercial and freeware. Geospatial data has some specificity that can be taken into account developing corresponding modules: interactivity of training and validation, visualisation of data and analysis of the residuals, control of modelling procedures using geostatistical tools, generation of new thematic GIS layers for real decision making process etc. In [Kanevski et al., 2008] the Machine Learning Office (MLO) is presented as a complete set of tools to solve the majority of these tasks.

Let us mention just some modules ands their rather unique properties: GeoMISC - utilities to perform exploratory data analysis, to manage and to visualise data and the results/residuals, generation of new GIS layers using raw data and modelling results; GeoKNN - k-Nearest Neighbour algorithm for regression and classification, training based on cross-validation, different types of Minkowski distances; GeoMLP - Multilayer Perceptron Neural Network, training using 1^{st} and 2^{nd} order gradient training algorithms, application of simulated annealing in order to generate starting point for the weights, different types of regularizations including noisy injection; GeoSVM - Support Vector Machines/Regression, equipped with two conventional QP solvers and tools for tuning the hyper-parameters, visualisation of support vectors; GeoGRNN - General Regression Neural Network with 6 types of kernels, cross-validation tuning of parameters, automatic detection of anisotropy; GeoGMM - Gaussian Mixture Model density estimator with full covariance matrix; GeoMDN - Mixture Density Network, based on Radial Basis Function Neural Network. GeoMDN is a valuable tool to model local probability density functions which is important for real risk mapping.

An important extension of the GeoGRNN module deals with the estimation of higher moments and not only mean values and estimation of the prediction uncertainties, as well as an estimation of the validity domain which in general corresponds to the density of measurements in the input space.

Software tools developed within the framework of Machine Learning Office are currently used for teaching and research in geospatial data modelling, such as topo-climatic modelling, natural hazard assessments (landslides, avalanches), pollution mapping (indoor radon, heavy metals, air and soil pollution), natural resources assessments, remote sensing images classification, socio-economic data analysis and visualisation, etc.

4. CONCLUSIONS

Machine learning algorithms are extremely powerful adaptive, nonlinear, universal tools. They were successfully used in many geo- and environmental applications. In principle, they can be efficiently used at all stages of environmental data mining: exploratory spatial data analysis, recognition and modelling of spatio-temporal patterns, decision-oriented mapping. Current trends in ML applications for geo- and environmental sciences deal with:

nonlinear dimensionality reduction and data visualisation; analysis and modelling of data in high-dimensional geo-feature spaces; fast modelling of physical and other processes in hybrid models; spatio-temporal patterns/structures extraction, modelling and predictions (data mining and forecasting).

Finally it should be noted, that being a data driven models they need deep expert knowledge in order to be applied correctly and efficiently starting from data pre-processing to the interpretation and justification of the results.

ACKNOWLEDGEMENTS

The authors wish to thank Swiss National Science Foundation (SNF) for the partial financial support of the projects "GeoKernels"(200021-113944) and "ClusterVille" (100012-113506). We thank reviewers for some useful comments which helped to improve the paper.

REFERENCES

Bishop, C., Patter Recognition and Machine Learning. Springer, Singapore, 2007.

Cherkassky, V., V. Krasnopolsky, D. Solomatine and J. Valdes (Eds.), Special Issue: Earth Sciences and Environmental Applications of Computational Intelligence. Introduction, *Neural Networks*, 19(2), 111-250, 2006.

Cherkassky, V., W. Hsieh, V. Krasnopolsky, D. Solomatine and J. Valdes, (Eds.), Special Issue: Computational intelligence in earth and environmental sciences. *Neural Networks*, 20(4), 433-558, 2007.

Chiles, J.P. and P. Delfiner, Geostatistics. Modelling Spatial Uncertainty, A Wiley-Interscience Publication, New York, 1999.

Cressie N., Statistics for spatial data, John Wiley & Sons, New-York, 1993.

Haykin S., Neural Networks. A Comprehensive Foundation. Second Edition. Macmillan College Publishing Company, New York, 1998.

Geokernels: http://www.geokernels.org, 2008.

Hastie T., R. Tibshirani and J. Friedman, The elements of Statistical Learning. Springer, 2001.

Kanevski M. and M. Maignan, Analysis and Modelling of Spatial Environmental Data. EPFL Press, Lausanne, 2004.

Kanevski M., A. Pozdnoukhov and V. Timonin, Machine Learning Algorithms for Spatial Environmental Data. Applications and Software Tools. (in press), EPFL Press, Lausanne, 2008.

Lee J. and M. Verleysen, Nonlinear dimensionality reduction, Springer, New York, 2007.

Pozdnoukhov A. and M. Kanevski, Monitoring Network Optimisation for Spatial Data Classification Using Support Vector Machines. *International Journal of Environment and Pollution*, 28(3/4), 465-484, 2006.

Shawe-Taylor J. and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, Cambridge, 2004.

Vapnik V., Estimation of Dependencies Based on Empirical Data. Springer, New York, 1998.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Semi-Supervised Support Vector Machine for Natural Hazards Forecasting. Case Study: Snow Avalanches

A. Pozdnoukhov[a], **R. S. Purves**[b], **M. Kanevski**[a]

[a] *Institute of Geomatics and Analysis of Risk, University of Lausanne, Switzerland (Alexei.Pozdnoukhov@unil.ch)*

[b] *Department of Geography, University of Zurich, Switzerland*

Abstract: This paper explores the use of Support Vector Machine (SVM) as a predictive engine for natural hazards forecasting. It particularly discusses the issues of incorporating this classification method into a decision-support system for operational use in avalanche forecasting. The recent developments concerned with semi-supervised and transductive SVM-based learning targeted at applications in natural hazards forecasting on geomanifolds are presented. The real case study on spatio-temporal avalanche forecasting deals with the development of a predictive engine for the decision support system used at the avalanche-prone site of Ben Nevis, Lochaber region in Scotland.

Keywords: environmental data mining, support vector machine, avalanche forecasting, semi-supervised and transductive learning.

1. INTRODUCTION

Amongst different natural hazards the events like snow avalanches are of particular interest. These events can be characterized by relatively low frequency, complex non-linear relationships with meteorological conditions, geomorphology and a large variety of other factors including human activity on the site. In terms of data-driven modeling, the avalanche forecasting can be considered as a classification problem, where one needs to find a decision boundary in the feature space of factors which discriminate the "safe" and "dangerous" conditions.

In this paper we explore the use of Support Vector Machine (SVM), a method from the field of Machine Learning, as a predictive engine for natural hazard forecasting. We discuss the issues of incorporating the developed model into a decision-support system for operational use in avalanche forecasting, and present the recent achievements. The real case study on the application of SVM is devoted to temporal and spatio-temporal avalanche forecasting at the avalanche-prone site of Ben Nevis, Lochaber region in Scotland, where avalanche forecasts are produced daily in winter months.

The paper is organized as follows. First, in the next section, we introduce the data-driven classification as an approach to decision support. We present there the basic features of a particular machine learning classification method, SVM, including the probabilistic interpretation of its outputs. Next, in Section 3, we motivate the use of semi-supervised and transductive learning in environmental data-driven modelling and describe the related contemporary approaches. We finally review the recent results on the application of SVMs for decision support in avalanche forecasting and provide the preliminary results on the use of semi-supervised and transductive SVM learning (Section 4). The paper is summarized with directions to the further developments and the conclusions in Section 5.

2. BINARY CLASSIFICATION AND DECISION SUPPORT

A wide range of numerical models and tools have been developed over the last decades to support the decision making process in environmental applications ranging from physical models, through expert systems, to a variety of statistically-based methods. In operational forecasting a mixture of all three approaches are often used, with process chains involving physical models and statistical or expert systems being relatively common.

As model complexity has increased, so to has our ability to collect real time, spatially distributed data describing a wide range of parameters through technological advances in sensor networks and automated environmental monitoring, and one can thus expect data-driven models to become increasingly important. Binary classification problems (the task to find a decision rule to discriminate the data into two classes such as "dangerous" and "safe" based on available empirical data), are widely met in environmental decision support. Interestingly, this target-oriented approach to decision support (direct inference from data to binary decisions without considering the intermediate modelling steps which complicate the model and bring uncertainty) is justified by the Occam Razor principle.

Below we present one of the most powerful data-driven classifiers, the Support Vector Machine, and discuss its use in decision support including an important issue of the interpretation of the data-driven forecasts produced by SVM.

2. 1 Support Vector Machine

SVM is a machine learning approach derived from Statistical Learning Theory aimed to deal with data of high dimensionality by approaching the nonlinear problems in a robust and non-parametric way. An interested reader is kindly asked to refer to some of the profound introductions to the theory of SVMs and related algorithms [Vapnik, 1998], [Scholkopf and Smola, 2002]. Here we only mention the main principles of SVMs which will find important applications in their applications in decision-oriented forecasting.

Suppose we deal with the linearly separable data (x_1, y_1), … (x_N, y_N), where x are the input features and $y \in \{+1, -1\}$ are the binary labels. By "linearly separable" we mean data that can be discriminated into two classes by a hyperplane. The idea of SVM is to separate this dataset by finding the hyperplane that is, roughly speaking, the farthest apart from the closest training points. The minimal distance between the hyperplane and the training points is called the margin, which is maximized by the SVM algorithm (Figure 1).

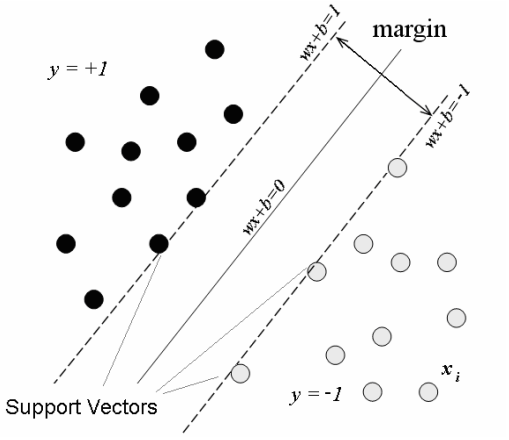


Figure 1. Margin maximization principle: the basic idea of Support Vector Machine.

It is proven in scope of Statistical Learning Theory that the maximum margin principle prevents over-fitting in high-dimensional input spaces, thus leads to good generalization abilities.

The decision function used to classify the data is a linear one, as follows:

$$f(x, w) = w \cdot x + b\,, \qquad (1)$$

where coefficient vector w and threshold constant b are optimized in order to maximize the margin. This is a quadratic optimization problem with linear constraints which has unique solution. Moreover, w is a linear combination of the training samples, many of them having zero weights α_i:

$$w = \sum_{i=1}^{N} y_i \alpha_i x_i \, . \qquad (2)$$

The samples with non-zero weights are the only ones which contribute to this maximum margin solution. They are the closest samples to the decision boundary and called Support Vectors.

To make this classifier non-linear, the so-called kernel trick is used. Kernel is a symmetric semi-positive definite function K(x,x'). According to the Mercer theorem, this implies that it corresponds to a dot product in some space (Reproducing Kernel Hilbert Space, RKHS). Generally, given a (linear) algorithm, which includes data samples in the form of dot products only, one can obtain a (non-linear) kernel version of it by substituting the dot products with kernel functions. This is the case for linear SVM, where the decision function (1) relies on the dot products between samples, as clearly seen by substituting (2) into (1). The final classification model is a kernel expansion:

$$f(x, \alpha) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \qquad (3)$$

The choice of the kernel function is an open research issue. Using some typical kernels like Gaussian RBF, one takes into account some knowledge like distance-based similarity of the samples. The parameters of the kernel are the hyper-parameters of SVM and have to tuned using cross-validation or a testing dataset.

2.2 Probabilistic Post-processing

Though the SVM is specifically constructed to solve the classification task, that is, to *discriminate the binary events*, the outputs of SVMs can be probabilistically interpreted by post-processing. To introduce a characterization of uncertainty, the values of the decision function (1) or (3) can be transformed into the smooth confidence measure, $0<p(y=1|x)<1$. This is done, for example, through taking a sigmoid transformation of $f(x,\alpha)$ [Platt, 1999]:

$$p(y=1|x) = \frac{1}{(1+\exp(a \cdot f(x) + b))}, \qquad (4)$$

where a and b are constants. These constants are tuned using a maximum likelihood (usually, the negative log-likelihood to simplify the optimization) on the testing dataset. The value of a is negative, and if b is found to be close to zero, then the default SVM decision threshold $f(x)=0$ coincides with a confidence threshold level of 0.5.

The major advantage of the latter interpretation is a possibility to introduce a decision threshold for the smooth confidence outputs $p(y=1|x)$. This threshold may later be tuned to satisfy the desired forecast quality measures.

2.3 Interpretation for Decision Support

The output of the binary classification system can be characterized by several basic measures, which are shown in the Table 1, known as contingency table. Concerning natural hazards, different possible forms and interpretations of the forecast are usually considered. Firstly in *categorical forecasts* a decision boundary is directly used to classify the region\time as being either dangerous or not. Secondly, in *probabilistic forecasts* the output of the system has to be interpreted as the probability of an event in the temporal or spatio-temporal domain of the forecast. Such forecasts can be used, for example, for risk assessment. Thirdly, a so-called *descriptive forecast* is often desirable, since experts wish to interpret and incorporate, for instance, a detailed list of similar events into their decision-making process. Concerning the last category, the Nearest Neighbour methods and their

variations commonly named as the "methods of analogues" are extensively used in a number of applications, with their probable roots in early atmospheric predictions [Lorenz, 1969].

Table 1. Basic measures for binary forecasts (contingency table or confusion matrix).

	Forecast	
Observed	Yes (+1)	No (-1)
Yes (+1)	Hits	Misses
No (-1)	False Alarms	Correct Negative

Support Vector Machines are well-suited to produce the abovementioned forms of the forecasts [Pozdnoukhov et.al., 2008]. The categorical forecast is just the predicted class, then the probabilistic interpretation can be used for predicting the event probability. The descriptive forecasts can be produced by providing the corresponding Support Vectors (which are the most valuable discriminative events in the past), though it still needs to be properly verified in a dialogue with a forecaster.

The Table 2 below provides some conventional forecast quality measures which can be used to tune the optimal hyper-parameters of SVM and the decision threshold.

Table 2. Forecast verification measures [Wilks, 1995].

Forecast accuracy measures	
POD - Probability of detection	The probability that the event was forecast when it occurred. POD = Hits/(Hits+Misses)
SR - Success rate	The probability that the event occurred when it was forecast. SR = Hits/(Hits+False Alarms).
HR - Hit rate	The proportion of correct forecasts. HR = (Hits+Correct Negative)/(Total Number of Days)

3. SEMI-SUPERVISED AND TRANSDUCTIVE LEARING

The problem of using unlabeled data is of increasing attention in Machine Learning. By unlabeled data, we mean those data samples which consist of the input values only, while the desired output value is unknown. The methods making use jointly of labelled and unlabeled data are called semi-supervised. When predictions have to be made to given unlabeled locations only, this particular situation is called transductive learning. Most real-life learning problems are actually semi-supervised, which gives rise to the developments of large-scale semi-supervised methods nowadays. For example, the semi-supervised setting of the problem for forecasting natural hazards such as snow avalanches is illustrated in Figure 2. The available data often consists of historical observations of avalanche activity (the list of events registered at given location under given meteorological conditions), the locations where avalanches were not observed under these conditions and the locations where the inputs (location and meteorological situation) are known but no information on avalanche activity is available. The ratio between the amount of available labelled and unlabelled data will always be in favour of the last.

The information one obtains from the unlabeled part of the dataset can be of different nature. A common approach is to consider the *manifold assumption*. This implies that data actually belong to some lower dimensional manifold in high dimensional input space. A large body of literature is devoted to the exploration of such an approach; see [Belkin, 2003], [Chapelle et.al., 2006] and references therein. Another use of unlabelled data is the so-called *cluster assumption*, which implies that the data are structured (clustered) in the input space. This structure can not be observed given the limited amount of labelled data, though the large amount of unlabelled samples would help exploring it. Concerning SVMs, the use of unlabelled data is illustrated in Figure 3. The methods

developed to exploit the cluster assumption include Trasductive SVMs and Low Density Separation methods. The overview on semi-supervised learning methods and particularly the approaches to Semi-Supervised SVMs (S^3VMs) can be found in [Chapelle et.al., 2006].

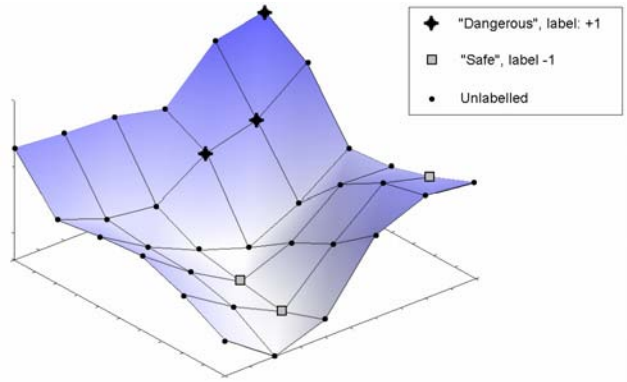


Figure 2. An example of typical natural hazards data. Snow avalanches on the digital elevation model of the terrain: the dangerous (avalanches were observed in the past under similar conditions), certainly safe (expert knowledge or no avalanches observed in the past under similar conditions) and unlabelled samples (those where input features are known but the outcome is unknown or/and has to be predicted).

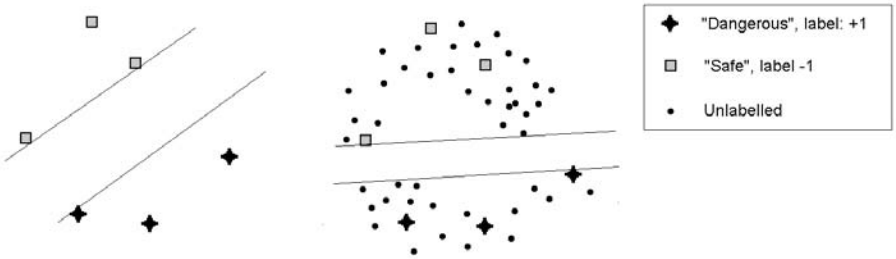


Figure 3. Unlabelled data can precise the data classification.

Semi-supervised learning methods become particularly useful in approaching environmental modelling in data-driven manner. With the finite set of available measurements or observations, the amount of related information (Digital Elevation Models and remote sensing images) one may consider is almost unlimited. It should be noted that the manifold assumption often can be considered to be satisfied while the likelihood of cluster assumption is less evident. The first one can be successively used both in classification and regression problems, while the second one is specific to classification. For both approaches, the amount of unlabelled data to model a manifold or to use with the methods relying on cluster assumption is almost exhaustive. We briefly introduce the model and the implementation of the S^3VMs which we use in the case study below.

3.1 Transductive Support Vector Machine

Several attempts were reported to implement the idea presented in Figure 3. The general approach is to formulate the margin maximization problem including the penalty given if the unlabelled sample appears to be inside the margin. This constraint is similar to the constraint applied to the labelled samples in standard formulation of the SVM. However, the optimization problem complicates significantly. In the experiments below we have used the approach of [Collobert et.al., 2006]. It formulates the optimization problem using the ramp loss as a Constrained Concave-Convex Programming (CCCP), and has two major advantages, first of them being reasonably fast computational speed. The second one is that the noisy and mislabelled data samples do not become the support vectors of the model as it is the case in standard SVM formulation. This model increases the number of hyper-

parameters to tune, including the width of the ramp loss and the trade-off cost constant which specifies the sensitivity of the model to misfits on unlabelled data samples. Despite of these practical difficulties, this implementation is used in the experiments below.

4. CASE STUDY: SNOW AVALANCHES

Concerning snow avalanche forecasting, different approaches have being proposed and being used in operational practice. To name some, these are the interpretations of the physical models of the development of the snowpack [Durand et al., 1999], expert systems which attempt to integrate expert knowledge [Schweizer and Föhn, 1996], and Nearest Neighbor methods, [Purves et.al., 2003]. Nearest Neighbour methods accord well with conventional inductive avalanche forecasting processes and are thus relatively popular with forecasters. In machine learning this is a relatively simple pattern classification technique. Moreover, both through theoretical considerations and in forecasting practice [McCollister et.al., 2003] it has been noted that such methods may be prone to over-fitting when dealing with highly-dimensional data.

4.1 Avalanche Forecasting in Lochaber Region

Avalanche forecasts are produced daily in the Lochaber region of Scotland, and the Nearest Neighbour based system is currently used there for decision support [Purves et.al., 2003]. The original data on avalanche activity in the region consist of daily measurements of 10 meteorological and snowpack variables starting from the winter seasons of year 1991. There is a database for this period with 712 registered avalanche events. These were happening at the particular 49 avalanche paths located at mainly north-east oriented slopes and gullies. The Digital Elevation Model of the region is available (Figure 4.1).

By using these data, the binary classification problem was formulated. The input feature vector contains several spatial features (coordinates of the events, elevation, slope, aspect, and convexity of the path) and about 30 temporal features, including meteorological observations for the previous 3 days for each event and derived “expert features”, constructed in a dialogue with a local avalanche forecaster who was asked to list important indicators of avalanche activity. The next step in identifying suitable features used recursive feature elimination in conjunction with a SVM to filter redundant features. This feature selection method iteratively omits the variables with the smallest influence on the decision surface of the SVM classifier. The final feature vector included a total of 26 variables and was used to produce the temporal forecasts with SVM [Pozdnoukhov et. al., 2008].

4.2 Spatially Variable Forecasts

An important step in producing the spatially variable forecasts consists in the spatialization of the meteorological data over the forecasting region. It can be done using physical models, heuristics, or data-driven approaches and is a matter of profound independent research not considered in this paper.

While it was relatively straightforward to put the registered avalanche events into a dataset as a class representing avalanche events, it is much harder to describe the “safe” conditions. Here lies an important issue - the samples of the “safe” class have to be both discriminative and have a proper label. In other words, to include a sample of the “safe” class one has to be sure that the snowpack at the given slope is stable under given conditions, while still representing a “non-trivial” data sample (as the slope with no inclination or without any snow at all). Here the unlabelled samples come into play - if no one is completely sure about the stability of the slope (no avalanches were actually observed due to bad weather conditions but the avalanche activity was suspected), we suggest including the sample as the one with known inputs but no output - the unlabelled one. As for the days with good visibility when no avalanche events were observed and no new snowfalls registered, the “safe” samples were constructed by combining the spatial features of all the potential avalanche paths and the current meteorological features. The intuition behind this was to provide the boundary and most discriminative samples to the system: those which are “safe” but still closest to turning into dangerous ones given changing weather conditions. It resulted in 712 positively labelled samples, about 30000 negatively labelled and about 12000 unlabelled samples. The problem is thus very unbalanced. In supervised SVM, it is

usually approached by modifying the misclassification costs of the different classes [Lin et al., 2002].

To approach this problem we have first considered the performance of the fully supervised SVM classifier with a Gaussian RBF kernel was applied to the labelled part of the data. It has produced 600 and 1400 Support Vectors for the "dangerous" and "safe" classes correspondingly. It means that most of the generated "safe" samples are lying far from the decision surface and do not contribute to the discrimination. Thus, the problem is reasonably balanced and the modification of the costs of labelled samples can be avoided. Moreover, the semi-supervised implementation of the SVM foresees the cost given to *unlabelled* samples as a specific parameter. It was tuned with cross-validation resulting in a lower value of C^{unlab}=0.1 considered to the labelled samples misclassification cost, which was found to be C^{lab}=3.

Note that in the semi-supervised setting the model selection is a non-trivial task. Concerning the fully supervised models, the cross-validation is usually applied. However, the choice of unlabelled sample acts as a user-defined parameter of the semi-supervised algorithm, and this issue is not yet properly explored.

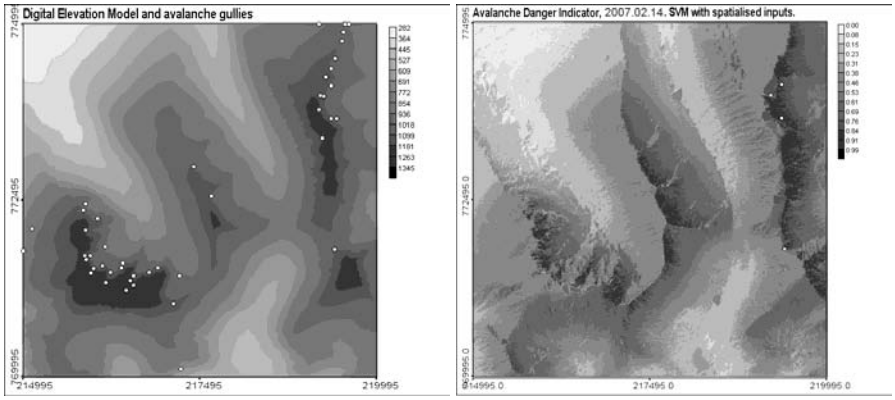


Figure 4. Left: Digital Elevation Model of the region and the locations of avalanche gullies (marked with circles). Right: The output of the semi-supervised SVM. The actual observed validation events are shown with circles.

An example of the produced forecast is presented in Figure 4, right. The performance curves obtained on the validation data (the data of the 2006-2007 seasons were reserved for the latter) are presented in Figure 5. While the overall behaviour of the systems is different, one can notice just slight improvement in the performance obtained with the use of unlabelled data. More efforts need to be done to properly validate the system.

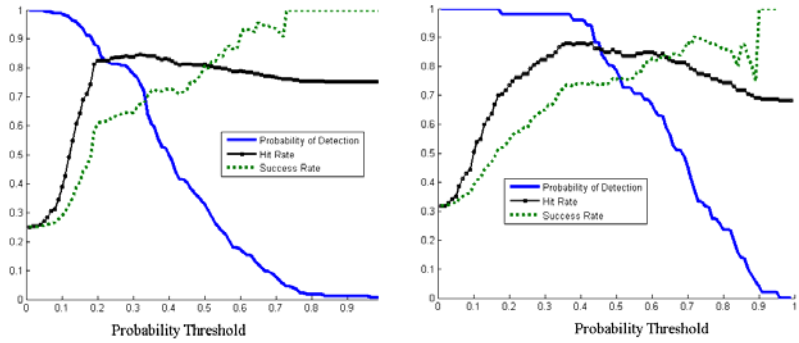


Figure 5. Performance curves of the baseline (left) and semi-supervised (right) systems, obtained on validation data for years 2004-2007. The hit rate of the semi-supervised system is just slightly higher.

5. DISCUSSION AND CONCLUSIONS

The recent developments in semi-supervised and transductive machine learning find exciting applications in natural hazards forecasting. The uncertain nature of these phenomena and availability of information make it possible to formulate the problem as a data-driven prediction based on labelled and unlabelled data. Particularly, in this paper we have illustrated the application of a semi-supervised Support Vector Machine to the spatio-temporal forecasting of the snow avalanche activity in Lochaber region of Scotland.

Bringing such systems into operational use for real-life decision support is a difficult undertaking, though the temporal forecasting with nearest neighbour models is an already accepted practice. Concerning the spatially varying forecasting, the output of physics-based models, such as the Safran-Crocus-Mepra tool [Durand et al., 1999] appears to be useful. Though, with the growing amounts of available data at finer spatial resolutions the data-driven modelling may be more advantageous in regional avalanche forecasting, at least due to the lower computational costs. The aim of the presented research was to make the first steps in this direction by considering whether the latest achievements from machine learning are suited to provide a predictive engine for this problem.

The focus of the further work will be the development of the proper validation schemes in order to investigate whether the semi-supervised approaches can produce useful improvements to the spatio-temporal forecasting of natural hazards. Then, the descriptive forecasts, highly required in an operational use, will be approached by the exploration of the set of support vectors responsible for the predictions.

ACKNOWLEDGEMENTS

This work was partially supported by Swiss National Science Foundation through the project N°200021-113944, “GeoKernels: kernel-based methods for geo and environmental sciences”.

REFERENCES

Belkin, M. (2003). Problems of Learning on Manifolds. Ph.D. thesis, Chicago, 2003.

Chapelle O., Scholkopf B., Zien A., (Eds.) Semi-Supervised Learning. MIT Press, 2006.

Collobert R., Sinz F., J. Weston, and L. Bottou. Large scale transductive SVMs. Journal of Machine Learning Research, 7:1687–1712, 2006.

Durand, Y., Giraud, G., Brun, E., Me´rindol, L., Martin, E., 1999. A computer-based system simulating swowpack structures as a tool for regional avalanche forecast. J. Glaciol. 45 (151), 466–484.

Lin Y., Lee Y., and Wahba G. Support vector machines for classification in non-standard situations. Machine Learning, 46:191–202, 2002.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues, Journal of the Atmospheric Sciences, 26, 636-646.

McCollister C., Birkeland K., Hansen K., Aspinall R. and Comey R. 2003. Exploring multi-scale spatial patterns in historical avalanche data, Jackson Hole Mountain Resort, Wyoming. Cold Regions Science and Technology, 37, 3, 299-313.

Platt J., Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In Advances in Large Margin Classiers, MIT Press, 1999.

Pozdnoukhov A., Purves R., and Kanevski M. Applying Machine Learning Methods to Avalanche Forecasting, to appear in Annals of Glaciology, 2008.

Purves R.S., K.W. Morrison, G. Moss and D.S.B. Wright. 2003. Nearest neighbours for avalanche forecasting in Scotland – development, verification and optimization of a model. Cold Regions Science and Technology, 37, 343-355.

Scholkopf, B. and Smola, A. 2002., Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge, MA.

Schweizer, J. and Föhn, P. M. B. 1996. Avalanche forecasting – an expert system approach. Journal of Glaciology, 42, 141, 318-332.

Vapnik V. N., Statistical Learning Theory. John Wiley & Sons, Inc., New York, 1998.

Wilks D.S. 1995. Statistical methods in the atmospheric sciences. (International Geo-physics Series 59), Academic Press, New York.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Using Artificial Neural Networks to Predict Local Disease Risk Indicators with Multi-Scale Weather, Land and Crop Data

A. Pechenick and D.M. Rizzo

College of Engineering & Mathematical Sciences, University of Vermont, Burlington, VT 05405 USA ({apecheni,drizzo}@cems.uvm.edu)

Abstract: The risk of fungal and bacterial crop disease can be predicted using risk models with specific environmental parameters such as temperature, relative humidity, solar radiation, wind speed, and leaf wetness duration (LWD). LWD has long been recognized as key in the management of crop disease. Air temperature and wetness influence the majority of fungal plant diseases. Wetness also impacts insect populations, as well as pollution deposits. Many parameters are well understood, readily defined, and easily measured. Unfortunately, LWD is a complex phenomenon, due to its spatial and temporal variability within a crop canopy. The inconvenience and uncertainty associated with monitoring LWD at the local leaf scale and the complexity of upscaling to the crop level prevent existing disease risk models from being used with reliability. In spite of their imprecision, LW projections are already included in a number of online weather products. One non-parametric statistical approach receiving scant attention for the modeling of LWD is that of artificial neural networks (ANNs). In this work, two previously untried ANNs estimate this key environmental variable at local crop scales, using local and regional weather station data and site-specific sensing data. The first ANN combines two statistical methods to accomplish this spatial mapping (a K-nearest means classifier and a Bayesian classifier), while the recurrent nature of the second ANN provides a means of leveraging the temporal property of the data. The ultimate goal is to embed the ANN into a highly-portable tool, designed to predict leaf wetness duration as an SOC (system on a chip) in conjunction with local weather stations, and as input to real-time decision support systems.

Keywords: Artificial Neural Networks; Leaf and Surface Wetness; Multi-scale Data; Crop Disease Risk; Decision Support Systems.

1. INTRODUCTION

Disease risk modeling for crop management has been shown to reduce disease incidence and severity [Campbell and Madden, 1990, Funt et al., 1990, Gleason et al., 1994]. Pathogens, pests and air pollution deposits are influenced by several environmental variables including temperature, relative humidity, net solar radiation, wind speed, and surface, or more commonly leaf, wetness duration (LWD) [Huber and Gillespie, 1992; Sharma 1976; Schuepp 1989; Getz 1991]. Among the most critical [Huber and Gillespie, 1992], as well as the most difficult [Sentelhas et al., 2007] of these variables to quantify and forecast are 1) canopy surface wetness (W_{obs}), defined as the observed fraction of plant parts or organs that are wet in a canopy, and 2) canopy surface wetness duration (SWD), defined as the sum of the continuous hours where W_{obs} is greater than 0.1 [Magarey et. al., 2006a, 2006b]. Despite these constraints, decision support systems incorporate LWD

estimates into their products, using both historical, and increasingly, forecast weather data as parameters for their LWD prediction models [Kim et al., 2006]. Model designs include physics-based, empirical and statistical, as well as hybrids of the above; performance trade-offs include complexity, accuracy, convenience and portability [Sentelhas et al., 2008].

2. MODELING SURFACE WETNESS DURATION

Converting W_{obs} into SWD requires classifying the data over a given time period (e.g., one hour) into binary categories of wet (1) and and dry (0). According to Magarey and Seem [2001], W_{obs} is classified as 0 if its value is less than 0.2, or 1 if its value is at least 0.2. Consider the small dataset for a single moisture event, illustrated by Figure 1. Predicted canopy surface wetness (obtained from an artificial neural network described below) and leaf wetness sensor data (obtained from Campbell Scientific leaf wetness sensors [Model 237, Campbell Scientific, Inc., Logan, UT 84321]) are tabulated at left and plotted versus time at right. The table also shows the percentage and binary SW classification for each hour.

	LW Sensor		ANN	
Hr.	W_{obs}	SWhr.	W_{obs}	SWhr.
284	0.0	0	0.0	0
285	0.2	1	0.4	1
286	0.5	1	0.3	1
287	0.9	1	0.8	1
288	0.9	1	0.8	1
289	0.8	1	0.7	1
290	0.77	1	0.7	1
291	0.77	1	0.9	1
292	0.9	1	0.9	1
293	0.8	1	0.9	1
294	0.6	1	0.6	1
295	0.2	1	0.9	1
296	0.07	0	0.0	0
297	0.0	0	0.0	0
	Σ = 12		Σ = 12	

Duration (SWD) = 12 hrs for both the sensor and ANN (RMS=0.0).

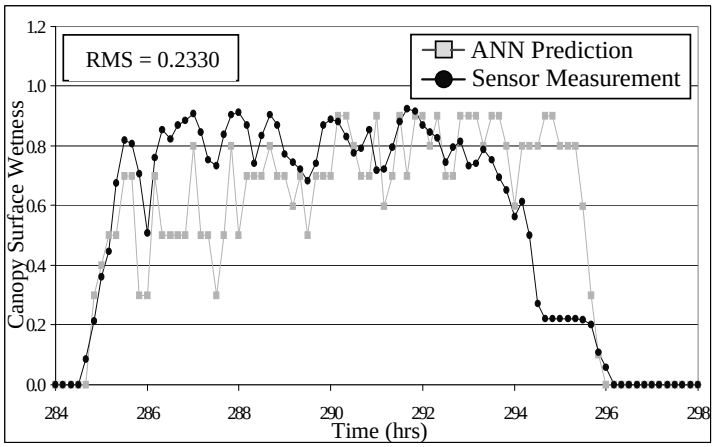


Figure 1. Example calculation of processing W_{obs} to SWD for a given moisture event.

Two important measures of error may be used to describe these data. One is the difference between canopy surface wetness sensor measurements and a model prediction *prior* to classification over each hour, and the second is the difference between sensor measurements and model predictions after classification to 0 or 1, over each hour. The root-mean-square (RMS) error is defined as $\sqrt{\frac{\sum_j (T_j - \hat{W}_{obsj})^2}{n}}$, where T_j is the measurement data (either W_{obs} visually recorded by an observer or, in the case of Figure 1, estimated by a Campbell Scientific leaf wetness sensor), $\hat{\mathbf{W}}_{\mathbf{obs}}$ is estimated by a model, and n is total number of data records over the moisture event. It is important to note that for the moisture event of Figure 1, the RMS error is computed as 0.2330 when comparing canopy surface wetness predicted by a model (indicated by squares) to sensor measurements (circles). However, when W_{obs} is post-processed into SWD (up-scaled through time), the RMS error for the same moisture event is 0. In an effort to provide conservative estimates of error (and

because the post-processing of W_{obs} into SWD is straightforward), we will present model forecasts in terms of W_{obs}.

Overall, an accurate visual representation of W_{obs} is required for successful model validation and sensor calibration. While temperature, relative humidity, wind speed, and rainfall have become standard measurements in a weather station set-up, surface wetness remains without a standard of measurement [Seem and Magarey, personal communication, 2002; Sentelhas et al., 2007]. Although leaf wetness sensors are presently used as surrogates, experts in quantifying canopy surface wetness rely on visual observations on some number of individual leaves to reflect the spatial aspect of the proportion of leaves wet in an entire canopy. Unfortunately, it is impossible to collect sufficient visual measurements at the same snapshots in time over the entire crop space, and the individual leaf wetness measurements must be averaged. If greater than 10% of the leaves in a canopy appear to have moisture for at least 12 minutes in an hour, the hour is considered "wet". W_{obs} is then calculated by averaging the binary values of the leaves over the time period of observation.

3. ARTIFICIAL NEURAL NETWORKS (ANNs)

ANNs have been used to predict wheat leaf wetness [Francl and Panigrahi, 1997], to estimate moisture occurrence and duration [Chtioui et. al, 1999], to forecast treatment for *Plasmopara viticola* infection [Dalla Marta et al. 2005] and to distinguish the type of infection and classify the infection period in a wheat field environment [de Wolf and Francl, 1997; 2000]. ANNs are not programmed; they are data-driven, and learn by example. Typically, an ANN is presented with a training set of examples (training patterns, typically represented as vectors) from which the network can learn. The two ANNs selected for this work utilize supervised learning, during which the ANN is also presented with a target output pattern: the *known* answer (classification) for the corresponding input pattern. The ANN methodology involves training the ANN to iteratively determine the hidden weights that will accurately map appropriate environmental parameters to W_{obs}. During training, examples of the mapping are presented to the network, and the weights of the hidden and output layers are adjusted over a series of iterations until the network has satisfactorily mapped the training set inputs to the training set outputs. Once a satisfactory mapping has been obtained, training ends and the weights are fixed. These fixed weights are then used during the interpolation phase to map new inputs (that the network has never seen) to predict $\hat{\mathbf{W}}_{\mathbf{obs}}$. In this work, the initial composition of input vector consisted of local weather variables (specifically temperature, relative humidity, wind speed, net canopy radiation, and leaf area index), with output being the canopy wet surface area, W_{obs}, for the given canopy elevation.

4. MODELING WETNESS DURATION USING COUNTERPROPAGATION ANN

4.1 What Distinguishes This Approach

Two ANNs have been developed, trained, and tested for the prediction of temporal trends associated with local crop scale SWD using multiple types of data measured from local scale and regional weather station data and site specific sensing data. The first is a counterpropagation network (CP), based on the work of Hecht-Nielsen [1987; 1988], the second a recurrent backpropagation network (RBP), independently generalized from standard backpropagation by Pineda [1987] and Almeida [1987]. The majority of earlier ANN attempts to model LWD use standard backpropagation, whose weaknesses, unlike CP, include a tendency towards local minima, the required optimization of multiple parameters, and a "black box" result [Frasconi, et al., 1993; Cristianini, 2001]. As shown in Figure 2,

CP is a combination of two classifiers – K-nearest means (Kohonen) and Bayesian (Grossberg) – and is guaranteed to converge. For details of the method see Hecht-Nielsen [1987]; for application of the method and pseudo-code see Rizzo and Dougherty [1994].

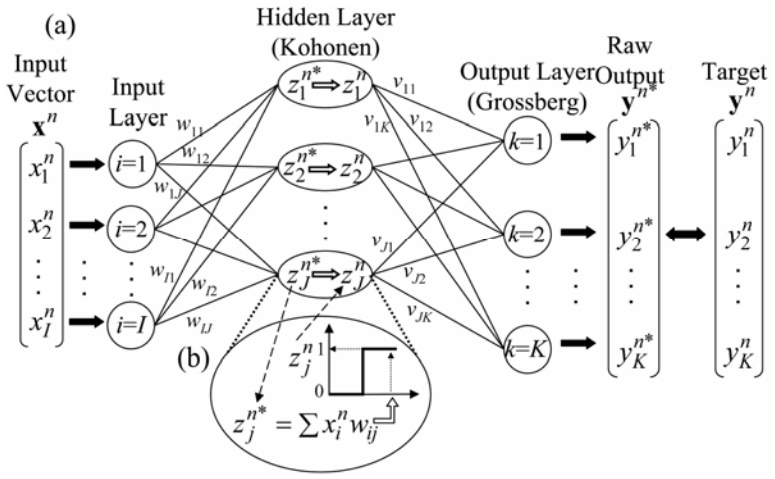


Figure 2. General schematic showing (a) architecture and notation and (b) activation function of the counterpropagation ANN. From Besaw and Rizzo [2007].

It is important to distinguish this method from the traditional feedforward backpropagation ANN (used in every software package), which requires stochastic training to select/optimize the number of hidden nodes. With the exception of feedforward backpropagation, the majority of the 50 or so existing ANN algorithms found in the literature perform better when large numbers of data are available. This custom counterpropagation algorithm does not suffer from some of the limitations associated with the feedforward backpropagation in that it cannot be over-trained. The hidden layer is a Kohonen self-organizing-map (SOM) used to cluster the data. The output layer maps the clusters to a known *a priori* classification (turning the Kohonen unsupervised ANN into a surpervised ANN). The more training data available, the better the classifications/predictions. This is true of most ANN algorithms, and explains why the majority of most commercial and proprietary data mining applications now use ANNs.

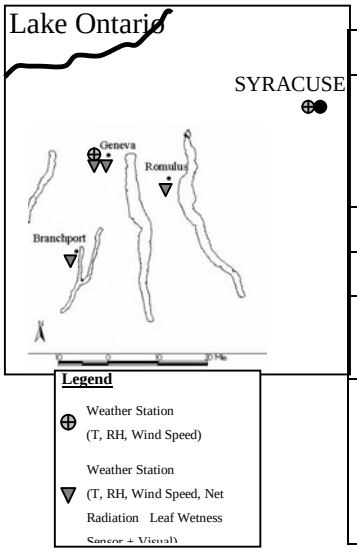
4.2 Counterpropagation Preferred over Recurrent Backpropagation

Comparison of the root-mean-square (RMS) error values between canopy surface wetness data (both visual observations and measurements), W_{obs}, and the ANN predictions, $\hat{\mathbf{W}}_{\mathbf{obs}}$, using the two ANNs indicates slightly better predictions using CP over RBP for the moisture events shown in this work (Figures 4 and 5). However, comparisons of predictions to observations (both visual and sensor) over 28 of the moisture events comprising two of the four datasets in Geneva N.Y., indicate no significant advantage in terms of the prediction capability of either network over the other. A significant difference, however, was observed in the time required for training each of the ANNs. RBP takes on average 12 to 13 times longer to train the same number of training patterns to the required RMS error value of 1×10^{-6}.

[Note: For our largest training set, consisting of 2450 records of multiple types of weather data, and corresponding sensor data measurements at approximately 10 minute intervals, the recurrent backpropagation network never did converge to the required RMS error value. As a result, only predictions using the counterpropagation model will be presented in this paper.]

4.3 Site and Data

A variety of datasets were used for training, testing, and validation. Roger Magarey, while at Cornell University, collected an extensive dataset from four grape cultivars growing at the Climatological Reference Station (NWS) in Geneva, New York. Drs. Seem and Magarey have graciously provided the data needed to train the ANN models for predictions of canopy surface wetness (see Figure 3).



Site	Site I.D.	Cultivar	Data Type	Data Frequency
Geneva	GV97	Chardonnay	Ag. Station†; Visual	1 hr
	LO98	Concord	Ag. Station; Visual	10 min
	NWS	Ag. Monitoring Site	Ag. Station	1 hr
Branchport	DE97	Concord	Ag. Station; Visual	1 hr
Romulus	SW97	Cabernet Franc	Ag. Station; Visual	1 hr
Syracuse	—	———	Weather Station† (no net radiation)	1 hr

†Ag station = Temperature, relative humidity, wind speed, LW sensors and net radiometers with measurements of each at separate elevations.

††Weather Station= Typical of data collected at a regional weather station, *i.e.* wet and dry bulb temperature, relative humidity, wind speed and direction.

Figure 3. a) Approximate location of the four grape canopy data sets and (b) description of cultivars and types of data collected at each New York data site.

The four cultivar data sets contain temperature, relative humidity, wind speed, canopy net radiation, rainfall, soil heat flux, soil moisture, and surface wetness (both visual and sensor data). Leaf surface wetness measurements were collected at three vines within a canopy. Each vine supported five sensors. Visual canopy surface wetness, W_{obs}, was estimated from the proportion of observed wet leaves. Visual measurements were replicated for three leaves observed at five canopy positions for three vines per site totaling 45 visual measurements at any given time. Sensor measurements of W_{obs} were also estimated from the proportion of wet sensors. Painted and unpainted Campbell scientific sensors were placed in the same five canopy positions along the same 3 vines. The visual estimates of W_{obs} provide validation of the forecasts of canopy surface wetness, $\hat{\mathbf{W}}_{\mathbf{obs}}$, from the ANN-based models. Observations of canopy surface wetness included both rain and dew moisture events. Temperature and relative humidity were collected at four elevations within the canopy. (Note: Surface wetness measurements (both visual and sensor) were collected at three of these elevations (bottom, middle and top of the canopy).) The weather stations logging the local data at all sites (4 sites in total) were located away from the edge of the field, near the middle of a row. Analysis of the W_{obs} data, (both visual and sensor), indicates a significant dependence on elevation within the canopy, an observation echoed by Jacobs et al. [2005] and Batzer et al. [2008]. As a result of these analyses, it was decided that 1) visual observations (rather than sensor data) would be considered "ground truth" and used throughout this research time frame to train the ANNs, 2) W_{obs} should not be averaged over the three canopy elevations, and 3) dew events and rains events would be treated separately (i.e., the ANN will be a function of elevation and moisture event-type).

4.4 Methodology

The CP was developed and trained using ten classes of surface wetness ranging from 0 to 1 in increments of 0.1. For the preliminary results shown in Figure 1, only 12 visual observations (along with corresponding measurements of temperature, relative humidity, wind speed, and canopy net radiation) were used as training patterns. Although not discussed here, the selection of training data is crucial to whether or not the network "learns" a particular task: how well a network "learns" depends on the examples presented during training. For this particular moisture event consisting of only 12 training patterns, the visual canopy surface wetness data (each point being an average over all vines, and elevations) contained no instances/classifications of 0.2 or 0.4. As a result, during the testing phase, when input vector patterns to the network consist of temperature, relative humidity, wind speed, and canopy net radiation collected at 10-minute intervals, it is not possible for the ANN to classify the output as 0.2 or 0.4 (*i.e.*, the CP ANN cannot predict outside the data range that it has been trained on). One of the first datasets used for training, testing and validation was Geneva N.Y. (LO98). The growing season consisted of 173 days and contained 116 visual W_{obs} measurements. This training dataset used the visual measurements of W_{obs} from the first 8 moisture events (75 records) spanning 40 days (in addition to temperature, relative humidity, wind speed, and net canopy radiation) as input training patterns, and then predicted $\hat{\mathbf{W}}_{\mathbf{obs}}$ over the remaining four moisture events (23 days). Comparisons of the ANN predictions to the observed visual data (along with the corresponding RMS error values and sensor measurements for the same time frames) are plotted in Figure 4. Sensor measurements are plotted on the figure, but only to indicate the beginning and end of moisture events. Similar results were obtained for the top and bottom canopy elevations; however, only the middle canopy elevations are shown.

Initially, a sensitivity analysis was performed on an energy-based surface wetness model (SWEB28) developed by Drs. Magarey and Seem at Cornell University. The results of this analysis indicate that of the four parameters tested (temperature, relative humidity, wind speed, and net canopy radiation), the physics-based model is the most sensitive to small changes in relative humidity, followed by net canopy, and, to a much less extent, wind speed and temperature.

A suite of sensitivity analyses were also performed on the ANN-based models using the two Geneva datasets to determine the most influential parameters needed for forecasting W_{obs} using the ANN. Analysis revealed that two of these 6 inputs (net canopy radiation and leaf area index) did not significantly improve predictions of W_{obs}**.** Comparison of the RMS error values over all moisture events indicates no significant increase in the RMS error values when net canopy radiation and leaf area index are omitted from the list of variables. In fact, in some cases the error was reduced. As a result, both were removed from the input training sets. A review of the literature corroborates these findings [Kim et al., 2006; Sentelhas et al., 2008].

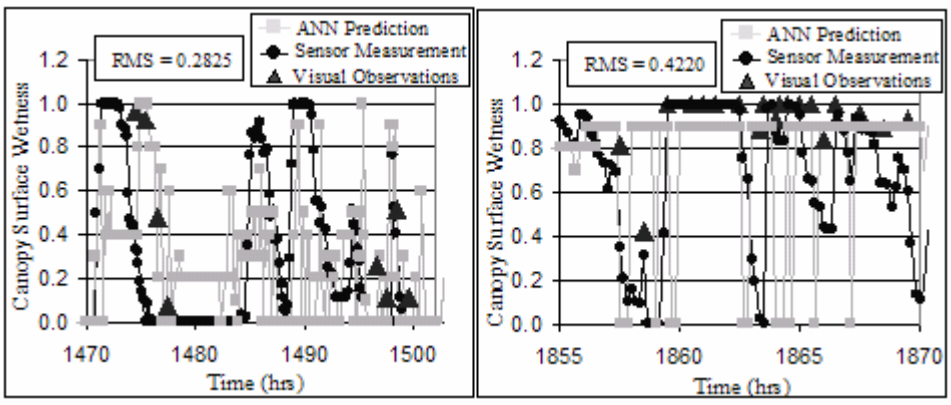


Figure 4. ANN predictions (squares) of canopy surface wetness and visual W_{obs} measurements (large triangles) obtained using moisture event records of local weather data for the Geneva, N.Y. (LO98) data.

4.5 Results and Scalability

The overall RMS error for all moisture events when 75 visual records (8 moisture events) are used in training is 0.2512. The overall RMS error for all moisture events when trained on 2450 records (15 sensor events) is 0.1981. Similar reductions in the RMS error values were found for the GV97 data set (RMS = 0.3014 and 0.1613 respectively). Therefore, significant decrease in the overall RMS error value is found when the training data are increased. Unfortunately, it was only possible to perform this test using the sensor data as the training target outputs; the collection of visual data over the same number of moisture events and time intervals is too labor intensive and companion data sets do not exist.

Hewitson and Crane [1996] describe techniques and applications of climate downscaling. Magarey et. al [2002] describe some emerging technologies, the limitations, and the prospects for future improvements for obtaining site-specific weather data without on-site sensors. Weather data may not be available at local spatial and temporal scales and the errors associated with spatial interpolation may be too large to provide accurate estimates of SWD using local models. To test and validate the ANN for local forecast errors (associated with the use of weather data that has been interpolated or downscaled from regional weather station data), the following test was performed: Crude local estimates of temperature, relative humidity, and wind speed data for Branchport, N.Y. were obtained using the ordinary kriging package available in ArcGIS Spatial Analyst and regional data from three weather stations located in Geneva, Romulus, and Syracuse, N.Y. (see Figure 3a). The data was not adjusted for elevation; the intent was simply to test the robustness and errors associated with the ANN's ability to forecast local canopy surface wetness given weather data interpolated from regional weather stations. Validation of the ANN forecast was achieved using visual data gathered at the local scale over a trial prediction period.

For training purposes, we used locally gathered Branchport temperature, relative humidity, wind speed measurements, and corresponding leaf wetness sensor measurements provided at 1-hour intervals during the month of July 1997. Estimates of canopy surface wetness were forecast by the ANN for the month of August 1997. Comparison of the ANN forecasts trained on both locally recorded and spatially interpolated weather data (see Figure 5, squares and diamonds respectively) indicate a significant decrease in the RMS error values when local weather data are used. We note that other companies such as SkyBit Technologies, Inc., have experience in interpolating spatial data such as that gathered at weather stations, and deliver weather and disease forecasts in the form of data or maps at the crop scale. Improving local weather estimates of temperature, relative humidity, wind speed, and net canopy radiation using more sophisticated downscaling interpolation methods and shorter prediction intervals will greatly improve ANN forecasts and reduce error estimates. However, since there was no significant decrease in the prediction capabilities of the ANN when the net canopy radiation was omitted from the input training set, downscaling of local weather variables may not be warranted since temperature, relative humidity, and wind speed are relative easy to measure by local growers.

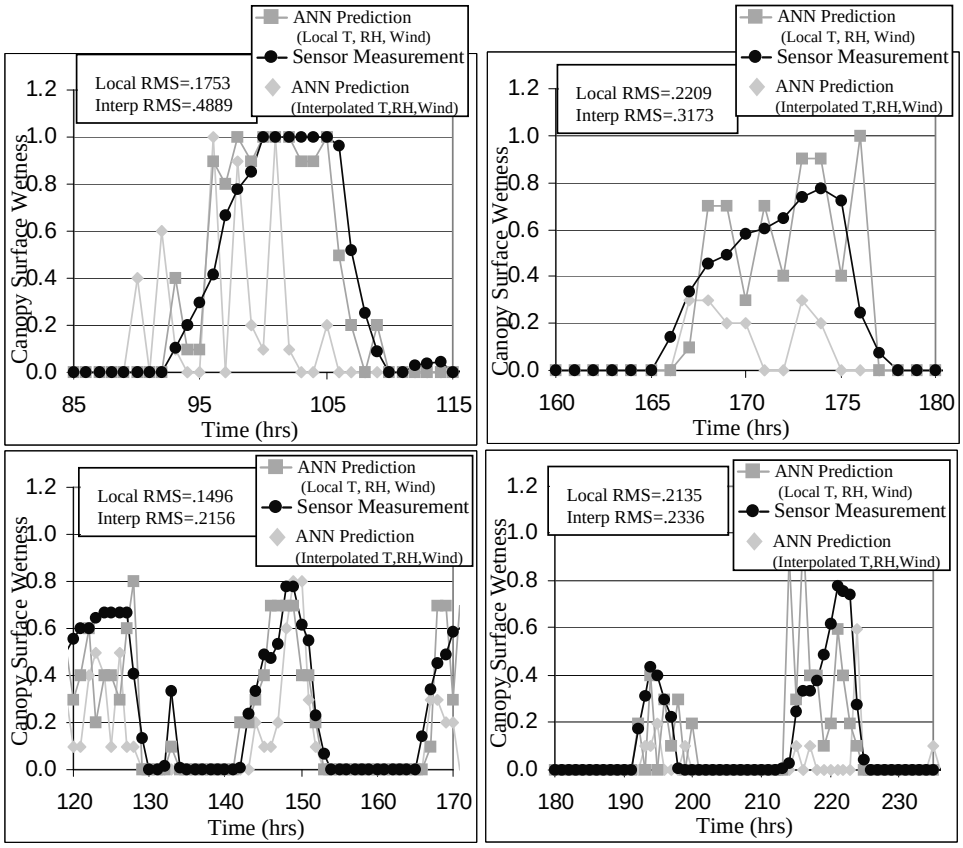


Figure 5. Trained on 3 rain events and 12 dew events LO98 sensor data, interpolated on 5 rain events, where visual and sensor data exist.

5. CONCLUSIONS

The work described above leads to several conclusions:

1. Comparison of the ANN forecasts using both the visual and sensor W_{obs} measurements as target outputs shows a significant increase in the accuracy of prediction when the number of training examples increases. That is, the ANN is better able to *learn* or *generalize* the relationship between input patterns (locally measured weather data) and output patterns (locally observed W_{obs} measurements) when the number of training patterns increases. One advantage of incorporating the statistical power of ANNs into existing risk disease models is that they "learn" the empirical relationships directly from the measured data. These relationships may be nonlinear and a neural network may model physical relationships with high degrees of nonlinearity.
2. We expect the ANN methodology to be easily transferable to other crops. However, the validation of model predictions with sensor and visual observations over a range of crop types and climates using standard protocol will be mandatory.
3. A standard for surface wetness measurements is required to validate models used to predict W_{obs} and SWD. Without standardized measurements to accomplish proper validation, these models will face barriers to entry into disease management strategies.
4. Visual observations and sensor measurements of leaf wetness represent point measurements of wetness and, as a result, cannot measure the proportion of leaves wet in a canopy (W_{obs}). Measurements are therefore averaged to provide information that is appropriate for the larger, canopy scale.

5. There was no significant increase in the prediction capabilities of the ANN when net canopy radiation and leaf area index were omitted from the input training set, and in some cases, the error was actually reduced. This is a favorable outcome, as net radiation measurements and estimates are not as commonly available as wind, rain, relative humidity, and temperature measurements.
6. Local collection of temperature, relative humidity, and wind speed data (as opposed to obtaining those weather variables from more complicated downscaling interpolation models) may be warranted.

ACKNOWLEDGEMENTS

We are grateful to Dr. Robert Seem, Professor of Plant Pathology at Cornell University and the New York State Agricultural Experiment Station, and Dr. Roger Magarey, Plant Pathologist at USDA-APHIS and North Carolina State University, for fostering a productive collaboration and for providing an extensive dataset and preliminary copy of their canopy water budget model (SWEB Version 0.28).

REFERENCES

Almeida, L.B., A learning rule for asynchronous perceptrons with feedback in a combinatorial environment, *Proc. IEEE First Int. Conf. on Neural Networks*, 609-18, IEEE Press, New York, 1987.

Batzer, J.C., M.L. Gleason, S.E. Taylor, K.J. Koehler, J.E.B.A. Monteiro, Spatial heterogeneity of leaf wetness duration in apple trees and its influence on performance of a warning system for sooty blotch and flyspeck, *Plant Disease, 92*(1), 164-170, 2008.

Besaw, L.E., and D.M. Rizzo, Stochastic simulation and spatial estimation with multiple data types using artificial neural networks, *Water Resources Research, 43*, W11409, 2007.

Campbell, C.L., and L.V. Madden, *Introduction to Plant Disease Epidemiology*, John Wiley & Sons, New York, 1990.

Campbell Scientific, *www.campbellsci.com/237-l*, 2008.

Chtioui, Y, L.J. Francl, and S. Panigrahi, Moisture prediction from simple micrometeorological data. *Phytopathology*, *89*(8), 668-672, 1999.

Cristianini, N., Support vector and kernel machines, *www.support-vector.net/icml-tutorial.pdf*, ICML 2001.

Dalla Marta, A., M. DeVincenzi, S. Dietrich, S. Orlandini, Neural network for the estimation of leaf wetness duration: application to a Plasmopara viticola infection forecasting, *Physics and Chemistry of the Earth, 30*, 91-96, 2005.

de Wolf, E. D. and L. J. Francl, Neural network that distinguish infection periods of wheat tan spot in an outdoor environment, *Phytopathology*, *87*(1), 83-88, 1997.

de Wolf, E.D. and L. J. Francl, Neural network classification of tan spot and *Stagonospora* blotch infection periods in a wheat field environment, *Phytopathology, 90*, 108-113, 2000.

Francl, L. J. and S. Panigrahi, Artificial neural network models of wheat leaf wetness, *Agricultural and Forest Meteorology*, *88*, 57-65, 1997.

Frasconi, P., M. Gori, A. Tesi, Successes and failures of back-propagation: a theoretical investigation, in Omid Omidvar_Eds., *Progress in Neural Networks*, Ablex Publishing, 1993.

Funt, R. C., M. A. Ellis, and L. V. Madden, Economic analysis of protectant and disease-forecast-based fungicide spray programs for control of apple scab and grape black rot in Ohio, *Plant Disease*, *74*(9), 638-642, 1990.

Getz, R.R., Report on the measurement of leaf wetness, *World Meteorological Organization, Commission for Instruments and Methods of Observation*, 1991.

Gleason, M. L., S.E. Taylor, T.M. Loughin, and K.J. Koehler, Development and validation of an empirical model to estimate the duration of dew periods *Plant Disease, 78*(10), 1011-1016, 1994.

Gleason, M.L., S.K. Parker, R.E. Pitblado, R.X. Latin, D. Speranzini, R. V. Hazzard, M.J. Maletta,W.P. Cowgill Jr., and D.L. Biederstedt, Validation of a commercial system for remote estimation of wetness duration, *Plant Disease*, *81*, 825-829, 1997.

Hecht-Nielsen, R., Counterpropagation networks, *Applied Optics*, *26*(23), 4979-4984, 1987.

Hecht-Nielsen, R., Applications of counterpropagation networks, *Neural Networks*, *1*, 131-139, 1988.

Hewitson, B.C., and R.G. Crane, Climate downscaling: Techniques and application, *Climate Research*, *7*, 85-95, 1996.

Huber, L., and T.J. Gillespie, T. J., Modeling leaf wetness in relation to plant-disease epidemiology, *Annual Review of Phytopathology*, *30*, 553-577, 1992.

Jacobs, A.F.G., B.G. Heusinkveld and E.J. Klok, Leaf wetness within a lily canopy, *Meteorological Applications*, *12*, 193-198, 2005.

Kim, K.D., M.L. Gleason, and S.E. Taylor, Forecasting site-specific leaf wetness duration for input to disease-warning systems, *Plant Disease*, *90*(5), 650-656, 2006.

Magarey, R.D. and R.C. Seem, Technologies for Estimating Wetting Periods, *http://www.nysaes.cornell.edu/pp/faculty/seem/magarey/leafwet/multexamp.html*, 2001.

Magarey, R. D., J.W. Travis, J.M. Russo, R.C. Seem, and P.A. Magarey, P. A., Decision support systems : Quenching the thirst, *Plant Disease*, Vol. *86*(1), 2002.

Magarey, R. D., R.C. Seem, A. Weiss, T. Gillespie, and L. Huber, Estimating surface wetness duration on plants, *Micrometeorology of Agricultural Systems*, Hatfield, J.L and J.M. Baker (eds.), American Society of Agronomy Monograph Series No. 20, ASA-CSSA-SSSA, Madison, WI, 199-226, 2005.

Magarey, R. D., J.M. Russo, and R.C. Seem, Simulation of surface wetness with a simple water budget and energy balance, *Agricultural and Forest Meteorology*, *139*(3-4), 373-381, 2006a.

Magarey, R. D., R.C. Seem, and J.M. Russo, Grape canopy surface wetness : Simulation versus visualization and measurement, *Agricultural and Forest Meteorology*, *139*(3-4), 361-372, 2006b.

Oke, T.R., Boundary Layer Climates, 2nd edition, Routledge, New York, 1987.

Pineda, F.J., Generalization of back-propagation to recurrent neural networks, *Physical Review Letters*, *59*(19), 2229-2232, 1987.

Rizzo, D. M., and D. E. Dougherty, Characterization of aquifer properties using artificial neural networks: Neural kriging. *Water Resources Research*, *30*(2), 483-455, 1994.

Schuepp, P. H., Microstructure, density and wetness effects on dry deposition to foliage, *Agricultural and Forest Meteorology*, *47*(2-4), 179-198, 1989.

Sentelhas, P. C., T.J. Gillespie, and E.A. Santos, Leaf wetness duration measurement: comparison of cylindrical and flat plate sensors under different field conditions, *International Journal of Biometeorology*, *51*(4), 265-273, 2007.

Sentelhas, P.C., A. Dalla Marta, S. Orlandini, E.A. Santos, T.J. Gillespie, M.L. Gleason , Suitability of relative humidity as an estimator of leaf wetness duration, *Agricultural and Forest Meteorology*, *148*(3), 392-400, 2008.

Sharma, M.L., Contribution of dew in the hydrologic balance of a semi-arid grassland, *Agricultural Meteorology*, *17*, 321-331, 1976.

iEMSs 2008: International Congress on Environmental Modelling and Software
Intelligent Environmental Decision Support Systems 4th Biennial Meeting of iEMSs,
http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Incorporating Likelihood information into Multi-objective Calibration of Conceptual Rainfall-Runoff Models

Alireza Nazemi[a], **Andrew H. C. Chan**[a], **Xin Yao**[b]

[a] *Department of Civil Engineering, The University of Birmingham, Birmingham, UK, B15 2TT (a.nazemi@cs.bham.ac.uk, a.h.chan@bham.ac.uk)*

[b] *School of Computer Science, The University of Birmingham, Birmingham, UK, B15 2TT (x.yao@cs.bham.ac.uk)*

Abstract: In this paper, the incorporation of parametric likelihood information into NSGA-II algorithm has been attempted in order to preserve solutions with more overall likelihood. The crowded comparison operator in NSGA-II, which is used to select the potential solutions for the next generation, is substituted by likelihood comparison operator which includes the consideration of the likelihood information about the potential solutions rather than their distance from each other. As a result the potential solution with higher overall likelihood measure has more chance to be selected in the next generation. Three different scenarios for the estimation of overall likelihood measure are presented. The modified algorithm is used for calibration of two different conceptual rainfall-runoff models in 24 USA MOPEX catchments. The results show that the new modification results to different searching process which can be compared with NSGA-II from different perspectives.

Keywords: Rainfall-Runoff models; Multi-objective calibration; NSGA-II; Parametric Likelihood, Likelihood comparison operator.

1. INTRODUCTION

The conceptual modelling paradigm assumes that rainfall-runoff process can be described by a simple sequential structure formed from a soil moisture accounting unit which converts the total rainfall into effective rainfall, followed by a routing unit which translates the amount of effective rainfall to its corresponding runoff quantity. All the conceptual models have some internal parameters which should be approximated using the observed measurements. Although the majority of previous calibration studies have been focused on single objective optimization for the calibration of conceptual rainfall-runoff models, recent investigations suggested that single objective function, no matter how carefully chosen, is often inadequate to properly represent all of the characteristics of the observed data and simulated values [Vrugt et al., 2003a]. Also, Boyle et al. [2000] showed the existence of different response modes on catchment's hydrograph, during the wetting up and drainage periods. It can be also shown that there are different response modes during high and low flows in both the wetting and drainage periods. These observations have opened new research direction toward multi-objective calibration of rainfall-runoff models, in which each particular part of the hydrograph can be described by a separate objective function.

Considering more than two conflicting objective functions in the search process, it may result in a set of solutions which might be quite diverse in both parametric and objective spaces. However, many of the solutions might not be useful in practice because of the considerable amount of sensitivity to small parametric changes. This behaviour, in many cases, is a disadvantage particularly in dealing with real world modelling problems, such as rainfall-runoff process, which contains several sources of uncertainties that can affect the performance of overall model output [Nazemi et al., 2006a]. Looking at this problem from another perspective, it can be questioned that if it is more promising to use other aspects of

the solutions such as parametric likelihood in the search process which may be able to give more chance to more likely solutions to be included in the final calibration results.

In this paper, an intuitive framework is introduced to incorporate likelihood information into the multi-objective calibration of rainfall-runoff models.

2. MULTI-OBJECTIVE CALIBRATION OF RAINFALL-RUNOFF MODELS

Let assume that the observed hydrograph and the model structure are Q_{obs} and ξ, respectively. By considering the model parameter set θ, the simulated hydrograph $Q(\xi \mid \theta)$ can be calculated. In the case of multi-objective calibration, the difference between Q_{obs} and $Q(\xi \mid \theta)$ is measured by a set of M objective functions $\{f_1, f_2, ..., f_M\}$ and calibration can be described as:

$$\min_{\theta \in \Theta} \{f_1, f_2, ..., f_M\} \tag{1}$$

The result of this optimization problem will be a set of Pareto parametric values Φ^* in which there is no solution better than the other regarding to all M objective functions. From the optimization point of view, the task of multi-objective optimization is not only assigning the best parametric values regarding each objective function; but also locating all possible Pareto optimal solutions and store them in the non-dominated set Φ^*.

Before applying multi-objective optimization framework to the calibration of rainfall-runoff models, a set of objective functions should be assigned. One possible approach would be to divide the whole hydrograph into different response modes and try to find a behavioural error measure related to each segment. We applied a heuristic segmentation approach introduced by Wagener and Wheater [2002] which uses the gradient of the hydrograph and an additional threshold as segmentation criteria to divide the whole hydrograph into different response modes. The flow gradient separates the hydrograph into wetting up and drainage periods in a way that positive gradient shows the rainfall periods, i.e. wetting up times and negative gradient shows the drainage periods. A threshold is used to separate period of high and low flow, which is the mean flow for wetting periods and 50% of the mean flow for drainage periods. In order to quantify the behaviour of the model for each segment, the Root Mean Squared Error (RMSE) in each mode was used. For instance for Driven High flows (FDH), the following error measure can be defined.

$$RMSE_{FDH} = \sqrt{\frac{\sum_{i=1}^{nFDH} (Q_i - \hat{Q}_i)^2}{nFDH}}, i \in FDH \tag{2}$$

Non-dominated Sorting Genetic Algorithm II (NSGA-II) is a multi-objective evolutionary algorithm introduced by Deb et al. [2002]. NSGA-II has been used in rainfall-runoff model calibration such as Tang et al [2005] and Nazemi et al [2006b]. In brief, NSGA-II can be considered as an ordinary evolutionary multi-objective algorithm, coupled with three new operators. The first operator is the fast non-dominated sort which can efficiently rank a set of solutions based on the multi-objective dominancy principle. The second operator is the crowding distance calculation operator which gives each solution a distant measure indicating its diversity and it is designed in a way that gives ultimate amount of distant to the edges of the Pareto-front. The third operator, i.e. the crowded comparison operator, is a selection rule designed in a way that not only consider the rank of the solutions but also select the solutions with more diversity distant measures based on the information obtained from crowding distance calculation. By this rule, the diversity of solution would be preserved during the search process. The algorithm will be run for a number of generations and the best non-dominating front will be saved in an archive. The solutions in this archive are the non-dominated solutions of the calibration.

Although these solutions have the greatest importance from the theoretical point of view, they might be unsuitable for the implementation of the model in practice. Comprehensive

preliminary study done by authors showed that, in most of the cases, the usage of the whole Pareto set is practically impossible for model implementation because of the high variation in the model output and the diversity of non-dominated solutions in the parametric space. As a result, a solution or a region of solutions should be picked. However, the parametric selection is not an easy task due to the existence of different parametric and objective clusters that can be linked to each other in a quite irregular way. It means that similar (close) solutions in either space can map to different (distant) solutions in the other space, which shows large amount of multi-objective multi-modality in the models landscape.

3. INCORPORATING LIKELIHOOD INFORMATION INTO MULTI-OBJECTIVE CALIBRATION OF RAINFALL-RUNOFF MODELS

As suggested by Vrugt et al [2003b], in Bayesian statistics, the parametric solutions are treated as probabilistic variables having joint probability density functions, which show the belief about the parametric solutions in the light of measured observation. Considering the model parameter as θ and the measured observation as y, the probability density function $p(\theta \mid y)$ is proportional to the product of likelihood function and the initial probability density function. Beven and Binley (1992) proposed the imposition of a uniform distribution on the feasible parametric space as the prior density function. Assuming a non-informative prior distribution, it has been shown that the posterior probability density function $p(\theta^{(t)} \mid y)$ is proportional to the sum square error of the model in the form of:

$$p(\theta^{(t)} \mid y) \propto \left[\sum_{i=1}^{N} \left(e(\theta^{(t)})_i \right)^2 \right]^{-\frac{1}{2}N} \tag{3}$$

The above formulation can be easily extended to each segment of the hydrograph. As an illustration, the posterior density distribution of parameter set $\theta^{(t)}$ in the light of observed driven high flow (FDH) segment of the hydrograph can be described as:

$$p(\theta^{(t)} \mid y_{FDH}) \propto \left[\sum_{i=1}^{N_{FDH}} \left(e(\theta^{(t)})_i \right)^2 \right]^{-\frac{1}{2}N_{FDH}} \tag{4}$$

The crowded comparison operator of NSGA-II is an intuitive rule: A solution *i* dominates another solution *j*, if and only if the solution *i* has a better rank or both solutions have the same rank but the solution *i* has a higher crowding distance measure. Based on Deb *et al* [2002], this rule can be formulated as following:

$$i \prec_n j \text{ if } (i_{rank} < j_{rank}) \text{ Or } [(i_{rank} = j_{rank}) \text{ and } (i_{dist} > j_{dist})] \tag{5}$$

This formulation can be easily reconstructed in order to preserve the solutions with higher likelihood. Considering $i_{Likelihood}$ and $j_{Likelihood}$ as the representatives of the overall likelihood estimations for solutions *i* and *j*, respectively; the *Likelihood Comparison Operator* can be defined as following:

$$i \prec_n j \text{ if } (i_{rank} < j_{rank}) \text{ Or } [(i_{rank} = j_{rank}) \text{ and } (i_{Likelihood} > j_{Likelihood})] \tag{6}$$

For each solution, the row measures of total and segmental posterior density measure be calculated using Equation 3 and Equation 4 respectively. It should be noted that these equations allocate a lower posterior measures to more likely solutions; as a result, in order to express the more likely solutions with higher measures of likelihood, the row measures can be considered with a minus sign. In order to eliminate the effect of different scales, all likelihood measures should be normalized to the same scale, i.e., [0,1] using the maximum and minimum of the minus row posterior density measures regarding each segment. Therefore, considering *m* as the number of segments considered and *n* as the number of solutions in each selecting pool, it would be a matrix $L_{n \times m}$ indicating the relative likelihood

measures of the solutions in a particular selecting pool. Here, three different scenarios are suggested in order to estimate the overall likelihood measure for each solution.

Scenario I. The maximum likelihood operator: In this scenario, the overall likelihood for each solution *i* can be estimated using following equation:

$$i_{Likelihood} = \max(\bigcup_{k=1}^{m} \{i_k\}) \tag{7}$$

Scenario II. The average likelihood operator: In this interpretation, the overall likelihood can be estimated using following equation:

$$i_{Likelihood} = \frac{1}{m}(\sum_{k=1}^{m} i_k) \tag{8}$$

Scenario III. The minimum likelihood operator: The overall measure can be described as

$$i_{Likelihood} = \min(\bigcup_{k=1}^{m} \{i_k\}) \tag{9}$$

By applying one of these scenarios to estimate the overall likelihood measure, the crowded comparison operator of NSGA-II (Equation 5) can be replaced by the likelihood comparison operator (Equation 6) in order to preserve the solutions with higher amount of overall likelihood measure. The new modification can be considered as the Likelihood non-dominated Sorting Genetic Algorithm II or LNSGA-II.

4. CASE STUDIES AND MODELS APPLIED

4.1 MOPEX catchments

The Model Parameter Estimation Experiment (MOPEX) is an international project aimed at developing enhanced techniques for the a priori estimation of parameters in hydrologic models and in land surface parameterization schemes of atmospheric models [Duan et al. 2005]. In this study we considered daily time series data for 24 USA catchments provided by MOPEX. These catchments can represent different hydrological mechanism. The key physical characteristics of the catchments used in this study are shown in Table 1.

Table 1. The variation of key physical properties of twenty four USA MOPEX catchments used in this study (http://www.nws.noaa.gov/oh/mopex/)

Area (km^2)	Main channel slope (m/km)	Stream length (km)	Mean basin elevation (m)	Area of storage (%)	Forest area (%)	Mean annual rainfall (mm)
966 – 9889	0.35 – 3.43	79.2 – 306.7	192 – 664	0 – 0.5	0.39 – 98	571.5 – 1397

4.2 Models applied

The conceptual structures used here are based on the Rainfall Runoff Modelling Toolbox (RRMT) [Wagener et al., 2001]. Both structures have a soil moisture accounting unit and a routing module. The two soil moisture accounting units have been used here are the catchment wetness index [Jakeman and Hornberger, 1993] and the probability distribution of stores [Moore, 1999]. These units have been combined with a two parallel linear reservoirs representing quick and slow responses of the system producing two conceptual models. Here, we labelled the two models used as model 1 and Model 2. Both models have five free parameters. Table 2 and Table 3 summarize the free model parameters and their feasible parametric space.

Table 2. Description and feasible interval for free parameters of Model 1 applied in this study [Wagener et al., 2001]

Model parameter	Definition	Feasible interval
M1-1	Soil moisture accounting parameter, time constant of the catchment losses	[1 – 40]
M1-3	Soil moisture accounting parameter, modulation factor	[0 – 5]
M1-6	Routing parameter, time constant of quick flow reservoir	[1 – 15]
M1-8	Routing parameter, time constant of slow flow reservoir	[15 – 200]

M1-10	Routing parameter, fraction of flow through quick reservoir	[0 – 1]

Table 3. Description and feasible interval for free parameters of Model 2 applied in this study [Wagener et al., 2001]

Model parameter	Definition	Feasible interval
M2-1	Soil moisture accounting parameter, maximum storage capacity	[0 – 500]
M2-2	Soil moisture accounting parameter, shape of Pareto distribution	[0 – 2.5]
M2-4	Routing parameter, time constant of quick flow reservoir	[1 – 15]
M2-6	Routing parameter, time constant of slow flow reservoir	[15 – 200]
M2-8	Routing parameter, fraction of flow through quick reservoir	[0 – 1]

4.3 Calibration framework

Three scenarios of LNSGA-II were introduced along with original NSGA-II algorithm for calibration of free parameters of Model 1 and Model 2 in MOPEX catchments. For each catchment, five years daily data have been considered in the calibration. Scatter crossover, Gaussian mutation and binary tournament selection have been fixed as the genetic operators in this experimental study. For the calibration of Model 1, five objective functions have been used including total and segmental RMSE. Also the same values for internal parameters have been used for NSGA-II and LNSGA-II scenarios. For Model 1, probability of crossover, probability of mutation, number of population and number of generations were assumed to be 0.8, 0.01, 100 and 100 respectively. In the case of Model 2 only four segmental error measures were used and the internal parameters were assumed to be 0.8, 0.001, 200 and 100 for the probability of crossover, probability of mutation, number of population and number of generations, respectively. In order to compare the objective optimality of solutions, the non-dominated solutions resulted from applied algorithms were compared with the single objective calibration solution of SCEM [Vrugt et al., 2003b]. If the archived solutions can reach relatively to the same level of optimality that single objective calibration does, it can be confirmed that the multi-objective sampler is at least as optimal as the single objective optimization but more time efficient because a set of objective functions were considered simultaneously. Considering the results of single objective optimization achieved by SCEM as the benchmark for the comparison, the level of optimality can be defined as $\frac{F_S}{F_M}\times 100$ in which F_S and F_M are the objective function convergence achieved by the single and multi-objective procedures, respectively.

5. RESULTS AND DISCUSSIONS

Table 4 shows the average level of optimality achieved by considered algorithms in MOPEX catchments using applied models in the case of Model 1, it is possible for LNSGA-II scenarios to converge to some optimal objective function values that can dominate the results of NSGA-II or even single objective solutions assigned by SCEM. In the case of Model 2, it can be concluded that LNSGA-II scenarios are not as successful as they were in the case of Model 1; however, it is still possible for LNSGA-II scenarios to result in better level of objective function optimality in comparison to NSGA-II and SCEM for some of the MOPEX catchments.

Considering the relationship between the Pareto fronts achieved from LNSGA-II scenarios and NSGA-II algorithm, three different outcomes can be observed and they are shown in Figure 1. In the first case, the LNSGA-II front can dominate the solutions achieved from NSGA-II. In the second one, LNSGA-II and NSGA-II can converge to more-or-less same front whereas in the third situation, the LNSGA-II solutions were dominated by NSGA-II. Based on this observation, it can be concluded that the applying likelihood comparison operator can result in an entirely different searching mechanism which can result in dominating, similar, or dominated solutions when compared to the NSGA-II solutions.

Figure 2 shows the histograms for the number of archived calibration solutions assigned by NSGA-II and scenarios of LNSGA-II considered in the MOPEX catchments. Analysing Figure 2, it can be concluded that for the LNSGA-II algorithm the diversity preservation

capability is lower than NSGA-II. This characteristic can be addressed by fundamental difference between the crowded comparison operator and the likelihood comparison operator for which the aim is to preserve the more likely solutions rather than providing a spread-out Pareto front. However, as Figure 3 shows regarding the calibration of Model 2, LNSGA-II can converge quite fast toward the neighbourhood of the non-dominated solutions. This property can have a particular importance in the calibration of rainfall-runoff models in which in some cases, each model simulation can be quite computationally expensive. Perhaps, the advantage of this fast convergence can be further highlighted for the regionalization purposes in which a model should be calibrated using several catchments. However, it should be noted that there is a trade-off between speed and robustness of convergence in LNSGA-II. The same observation can be made for Model 1.

Table 4. The average level of optimality achieved by considered algorithms in MOPEX catchments using applied models

Model	Calibration method	RMSE	RMSE(FDH)	RMSE(FDL)	RMSE(FNQ)	RMSE(FNS)
Model 1	NSGA-II	98.75%	98.31%	98.15%	98.42%	96.51%
	LNSGA-II, S1	100.60%	96.55%	104.22%	97.49%	96.27%
	LNSGA-II, S2	99.85%	95.45%	101.78%	97.26%	95.67%
	LNSGA-II, S3	99.04%	94.43%	102.98%	96.21%	94.15%
Model 2	NSGA-II	---	99.95%	96.77%	99.59%	96.64%
	LNSGA-II, S1	---	96.17%	99.75%	99.47%	92.76%
	LNSGA-II, S2	---	96.15%	96.45%	99.61%	89.66%
	LNSGA-II, S3	---	96.12%	94.73%	99.93%	88.93%

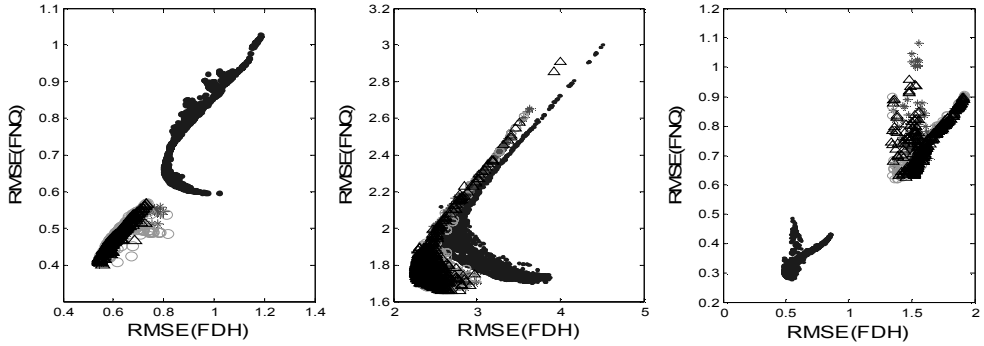
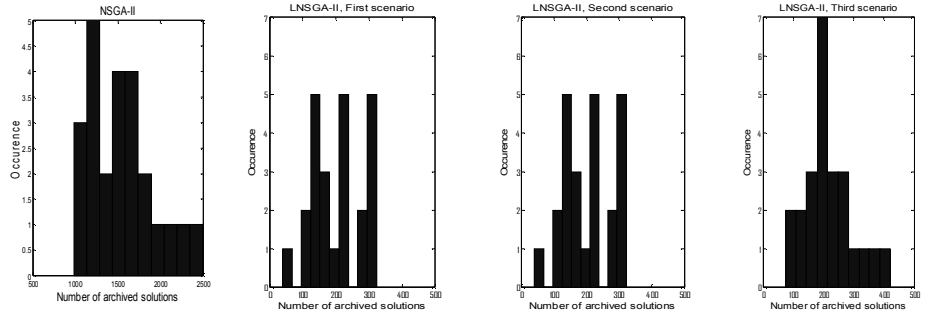
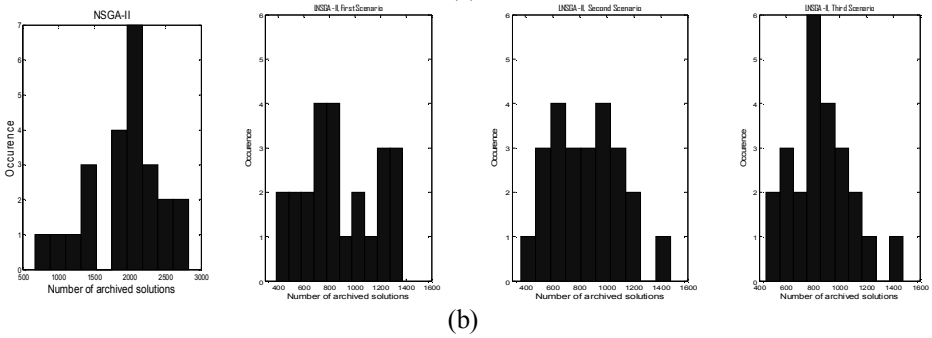


Figure 1. Different conditions between the Pareto front achieved by NSGA-II and LNSGA-II scenarios. Dots are representing NSGA-II solutions and the results of LNSGA-II scenarios have been plotted using stars, circles and triangular respectively. Left) Results of Model 2 in catchment ID 07141200; Middle) Results of Model 2 in catchment ID 07186000; Right) Results of Model 2 in catchment ID 07311700



(a)



(b)

Figure 2. Comparison among the numbers of archived solutions for the applied models in MOPEX catchments using NSGA-II and LNSGA-II scenarios; (a) Model 1 (b) Model 2

Using LNSGA-II, a more compact and clustered solutions in the parametric space can result which makes the parameter selection less complex. As an illustration, Figure 4 compares the MDS plots [Buja and Swayne, 2002] of parametric solutions resulted from NSGA-II and three scenarios using LNSGA-II for the calibration of Model 2 in the catchment ID 07187000. As can be observed, NSGA-II results in wide spread solutions in the parametric space; whereas the solutions are more compact and clustered in MDS plots for LNSGA-II scenarios.

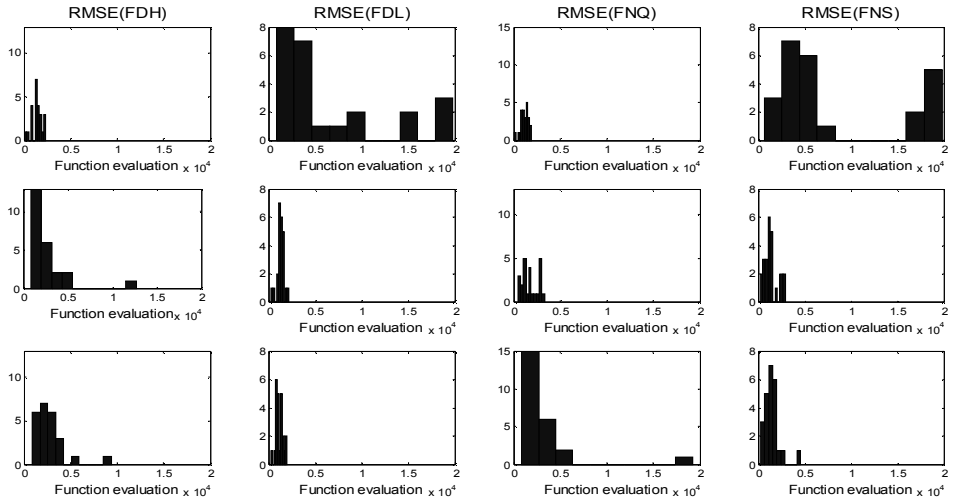


Figure 3. Comparison among the speed of convergence related to calibration of Model 2 in MOPEX catchments using three different scenarios of LNSGA-II. The first row indicates the number of required function evaluations in the first scenario and the other two rows are related to second and third scenarios, respectively.

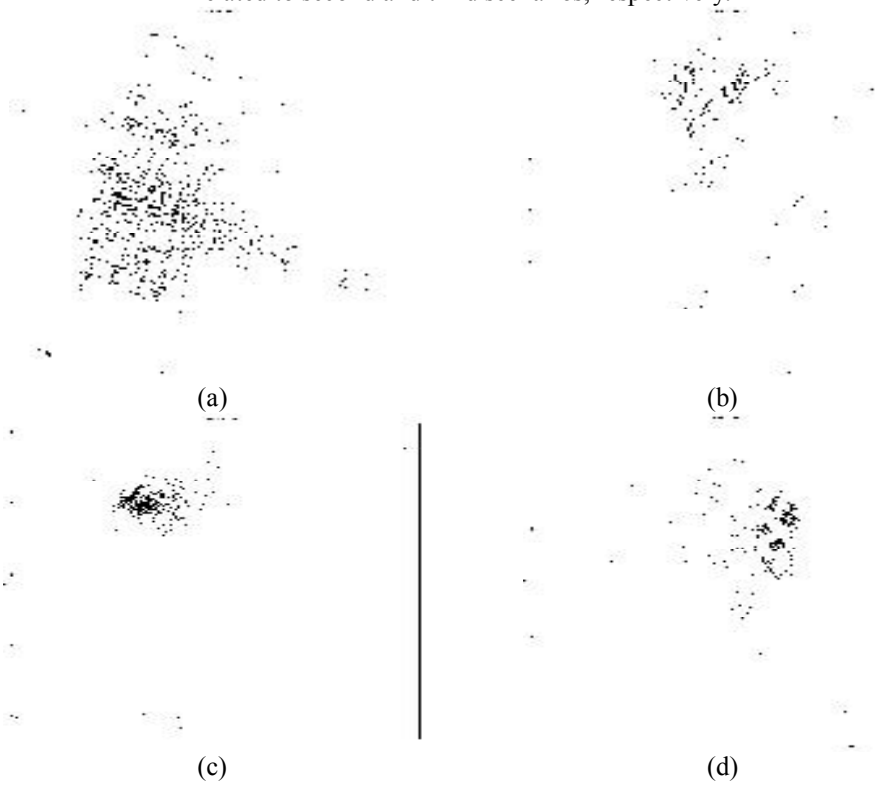


Figure 4. Comparison among MDS plots of non-dominated solutions resulted from calibration of Model 2 in catchment ID 07187000 using NSGA-II and three different scenarios for LNSGA-II; (a) NSGA-II (b) LNSGA-II, first scenario (c) LNSGA-II, second scenario (d) LNSGA-II, third scenario

6. CONCLUSIONS

Applying the whole Pareto calibration solutions of rainfall-runoff models is practically impossible due to the considerable amount of diversity in the assigned solutions of NSGA-II. This paper has explored the possibility of tailoring NSGA-II in order to capture more likely solutions instead of providing spread out Pareto optimal solutions. The results have shown that the solutions assigned by the tailored algorithm, i.e. LNSGA-II, contain less diversity (due to the less number of archived questions) and can locate more compact solutions in the parametric space (due to the results of MDS plots). In addition, changing the selection rule establishes a new searching mechanism which can result in a better or worse Pareto-front in comparison to the result of NSGA-II. However, it was found that LNSGA-II scenarios can converge much faster than NSGA-II.

REFERENCES

Beven, K. J. and Binley, A. M. "The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6, pp. 279-298, 1992.

Boyle, D. P., Gupta, H. V., and Sorooshian S., "Toward improved calibration of hydrological models: Combination the strengths of manual and automatic methods", Water Resources Research, VOL. 36(12), 3663-3674, 2000.

Buja, A. and Swayne, D. "Visualization Methodology for Multidimensional Scaling", Journal of Classification, (19), pp 7–43, 2002

Deb, K., Pratap, A., Agarwal, S., Mayarivan, T., "Fast and elitist multi-objective genetic algorithm: NSGA-II", IEEE Transactions on Evolutionary Computation, VOL. 6, NO. 2, pp.182-196, 2002.

Duan, Q., Schaake, J., Andreassian, V., et al., Model Parameter Estimation Experiment (MOPEX): An Overview of Science Strategy and Major Results from the Second and Third Workshops, J. Hydrol., 2005, vol. 320, nos. 1–2, pp. 3–17.

Jakeman, A. J., Hornberger, G. M., "How much complexity is warranted in rainfall-runoff model", Water Resources Research, VOL. 29, NO. 8, pp. 2637-2649, 1993.

Moore, R.J., Real-time flood forecasting systems: Perspectives and prospects. In: Floods and landslides: Integrated risk assessment (Eds. R. Casale and C. Margottini), Springer, Berlin, 147-189. 1999.

Nazemi, A.-R., Yao, x., Chan A. H., " Extracting a Set of Robust Pareto-Optimal Parameters for Hydrologic Models using NSGA-II and SCEM", in the proceeding of IEEE 7th World Congress on Computational Intelligence, Vancouver, Canada, July 2006.

Nazemi, A.-R., Yao, X., Chan A. H., " Multi-objective calibration of conceptual rainfall-runoff models using NSGA-II", in the proceeding of 7th International Conference on Hydroinformatics, Nice, France, September 2006.

Tang, Y., Reed, P., Wagener, T., "How effective and efficient are multi-objective evolutionary algorithms at hydrologic model calibration?" Hydrology and Earth Sciences Discussion, (2), pp 2465-2520, 2005.

Vrugt, J. A., Gupta H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S., "Effective and efficient algorithm for multi-objective optimization of hydrologic models", Water Resources Research, VOL. 39 (8), 2003.

Vrugt, J. A., Gupta H. V., Bouten, W., and Sorooshian, S., "A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters", Water Resources Research, VOL. 39 (8), 2003.

Wagener, T., Lee, M. J., Wheater, H. S., Rainfall-runoff modelling toolbox user manual (RRMT v3.0), Civil Engineering and Environmental Sciences Department, Imperial College, London, 2001.

Wagener, T., Wheater, H. S., "On the evaluation of conceptual rainfall-runoff models using multiple-objectives and dynamic identifiability analysis", In Littlewood, I. (ed.), Continuous river flow simulation: methods, applications and uncertainty, British Hydrological Society, Occasional paper, No. 13, Wallingford, UK, pp.45-51, 2002.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Environmental data inference through fuzzy clustering

S. Marsili-Libelli[a] and E. Giusti[a]

[a] *Department of Systems and Computers, University of Florence, Via S. Marta 3, 50139 Firenze, Italy (marsili/giusti@dsi.unifi.it)*

Abstract: Retrieving the relevant information from noisy environmental data is a difficult task, which requires advanced filtering techniques. This paper presents a method for extracting representative prototypes from a set of data containing daily and seasonal fluctuations. It is based on a combination of wavelet filtering and fuzzy clustering. Discriminating features are extracted with a Fuzzy Maximum Likelihood Estimates (FMLE) clustering algorithm, which was selected for its variable metric enabling the adaptation of the cluster shape and volume to the data. The results show that the discriminating power of this algorithm is considerable, as demonstrated by the application to two differing domains: the discrimination of dissolved oxygen circadian cycles in the Orbetello lagoon and the daily and seasonal fluctuations in photosynthesis in the Arno river. In both cases the isolated patterns have a clear ecological meaning and reveal the relevant ecosystem variations on differing time-scales.

Keywords: Fuzzy modelling, pattern recognition, fuzzy clustering, Water quality, photosynthesis, dissolved oxygen.

1. INTRODUCTION

Extracting the information from environmental data may become a very complex process for a number of reasons: the noise affecting the data is non-Gaussian and difficult to model, the information is contained in the data rate of change rather than in the data itself, and complex patterns may be hidden in the data. This paper describes an information retrieval procedure developed with fuzzy tools, in order to extract information from environmental noisy data.

An efficient fuzzy inferential system may be composed of a large number of IF-THEN predicates, so that practical application may easily reach unmanageable proportions. To avoid the "curse of dimensionality" data are grouped into fuzzy clusters depending on the extent to which they share some common features. In this way the antecedents are replaced by partition clusters, whose degrees of truth are used to activate the fuzzy inference engine inferring the current environmental condition

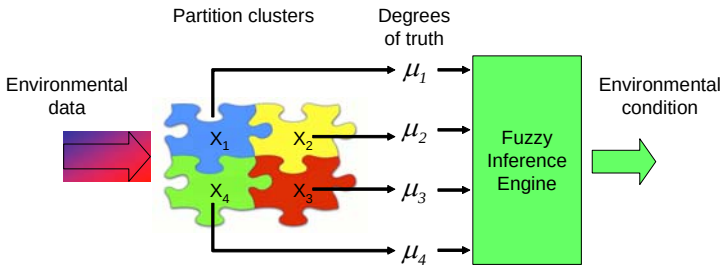


Figure 1. General structure of the Sugeno fuzzy inference algorithm with clustered antecedents**.**

The generic i-th fuzzy rule of the Sugeno fuzzy inference algorithm of Figure 1 has the form

$$R_i : IF\ x \subset X_1\ AND\ ...\ x \subset X_n\ THEN\ y = y_i \,. \tag{1}$$

and the resulting environmental condition is obtained by weighted average defuzzification

$$y = \frac{\sum_{i=1}^{n} \mu_i y_i}{\sum_{i=1}^{n} \mu_i} \tag{2}$$

where n is the number of rules and μ_i is the degree of activation of each cluster.

1.1. The pattern recognition algorithm

The algorithm described in this paper is aimed at extracting meaningful behavioural patterns from a set of data representing circadian cycles. It is based on a combination of wavelet filtering and fuzzy clustering.

1.1.1. Wavelet denoising

The data are first processed with a discrete wavelet decomposition in order to isolate the relevant circadian cycle, and it is based on a combination of wavelet filtering and fuzzy clustering. The wavelet filtering represents a useful data pre-processing tool to separate noise from the relevant deterministic information allowing a more discriminating features extraction. This is later accomplished by applying a fuzzy clustering algorithm to the denoised data. The Fuzzy Maximum Likelihood Estimates (FMLE) clustering method was preferred for its variable metric, which adapts the cluster shape and volume to the data. The results show that the discriminating power of this combined technique is much higher than that of each technique used separately.

Wavelet denoising [Strang and Nguyen, 1996; Percival and Walden, 2000] is performed to alleviate the noise problem and provide a stable numerical derivative, if required. As thoroughly described in [Marsili-Libelli and Arrigucci, 2004] wavelet denoising is obtained by a Multi Level Digital Wavelet Decomposition (MLDWD) where the scale factor is halved at each stage. The decomposition is based on powers of two of the circadian period T_s, providing a signal decomposition into an *approximation*, grouping the low-frequency components of the signal, and a *detail*, retaining the high-frequency components. From extensive data experimentation, it was concluded that the Level 2 approximation is a good denoised reconstruction of the true circadian cycle, as thoroughly discussed in [Marsili-Libelli and Arrigucci, 2004; Marsili-Libelli, 2004]. The Meyers [Meyers et al., 1993] was selected for its properties of symmetry, orthogonality and finite support.

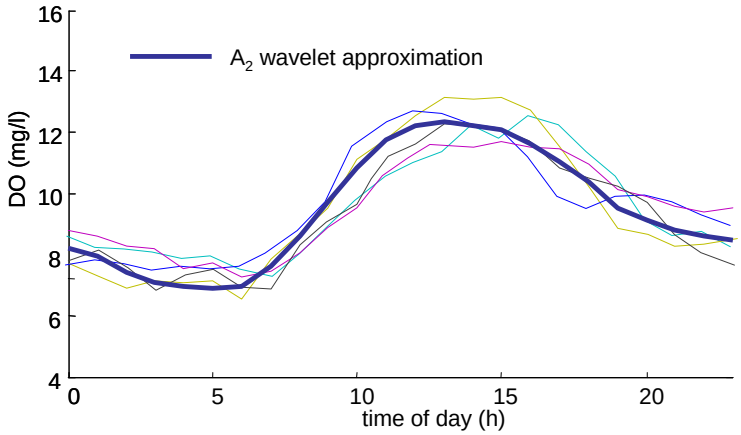


Figure 2. Meyers second approximation A_2 (thick line) of several daily patterns (thin lines) for the daily DO variations in the Orbetello lagoon.

1.1.2. Fuzzy clustering

Once the basic features are denoised, pattern recognition is applied to group significant behaviours a number of meaningful fuzzy clusters. Of the many techniques available [Bezdek, 1981; Babuska, 1998; Abonyi, 2003] the Fuzzy Maximum Likelihood Estimates (FMLE) algorithm is used here for its ability to produce clusters of varying volume and shape, thus adapting to the dimension of the data. This flexibility is a consequence of basing the partition on an adaptive distance norm derived from the fuzzy covariance matrix, as shown in Eq. (4).

The FMLE fuzzy clustering algorithm arranges the data $\boldsymbol{x}_k$ into c clusters through the constrained minimisation of the partition functional

$$J(\mu_{i,k},m)=\sum_{i=1}^{c}\sum_{k=1}^{N}\left(\mu_{i,k}\right)^{m} d_{ik}^{2}\ , \tag{3}$$

where m is the fuzzy exponent and the distance d_{ik} depends on the norm-inducing matrix, which in the FMLE case is the fuzzy covariance matrix $\boldsymbol{\Sigma}$, defined as

$$\Sigma_i=\frac{\sum_{k=1}^{N}\mu_{ik}\left(\boldsymbol{x}_k-\boldsymbol{v}_i\right)\left(\boldsymbol{x}_k-\boldsymbol{v}_i\right)^{T}}{\sum_{k=1}^{N}\mu_{ik}} \qquad i=1,\ldots,c\ . \tag{4}$$

The distance in Eq. (3) then becomes

$$d_{ik_F}=\frac{\left(det\ \Sigma_i\right)^{\frac{1}{2}}}{\left(\frac{1}{N}\sum_{k=1}^{N}\mu_{ik}\right)}\ exp\left(\frac{1}{2}\left(\boldsymbol{x}_k-\boldsymbol{v}_i\right)^{T}\Sigma_i^{-1}\left(\boldsymbol{x}_k-\boldsymbol{v}_i\right)\right) \tag{5}$$

and the partition membership functions are obtained as

$$\mu_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik_\Sigma}}{d_{ij_\Sigma}}\right)^{\frac{2}{m-1}}} \qquad i=1,\ldots,c \qquad k=1,\ldots,N \tag{6}$$

The resulting membership μ_{ik} in Eq. (6) represents the degree of membership of the *k-th* data point $\boldsymbol{x}_k$ to the *i-th* cluster with center $\boldsymbol{v}_i$.

The choice of the best number of clusters (c) was performed by applying three different cluster validity measures:

- *Fuzzy hypervolume*:

$$V_h=\sum_{i=1}^{c}\left[det\left(\Sigma_i\right)\right]^{\frac{1}{2}}, \tag{7}$$

where $\boldsymbol{\Sigma}_i$ are the cluster covariance matrices. The first minimum of this measure yields the best cluster partition;

- *Average Partition Density*:

$$D_A=\frac{1}{c}\sum_{i=1}^{c}\frac{S_i}{\left[det(\Sigma_i)\right]^{\frac{1}{2}}}, \tag{8}$$

where S_i is the sum of the membership degrees of the data that lie within a hyper-ellipsoid with radius equal to the standard deviations of the cluster features. The first maximum of this measure yields the best cluster partition;

- *Average Cluster Flatness* :

$$t_A=\frac{1}{c}\sum_{i=1}^{c}\frac{\lambda_{in}}{\lambda_{i1}}\,. \tag{9}$$

It is based on the ratio between the smallest and largest eigenvalue of the fuzzy covariance matrices $\boldsymbol{\Sigma}_i$. The first minimum of this measure yields the best cluster partition.

2. APPLICATIONS OF THE ALGORITHM

The combined wavelet denoising and fuzzy clustering algorithm is used as a pattern recognition procedure and applied to the detection of variations in two water quality environmental time-series: the patterns in the daily dissolved oxygen in the Orbetello lagoon and the seasonal fluctuations in the river water quality parameters measured in the Arno river, both in Tuscany, central Italy, as shown in Figure 3.



Figure 3. Location of the two case studies described in the paper.

2.1. Daily dissolved oxygen patterns in the Orbetello lagoon

This problem has already been treated in the past [Marsili-Libelli and Arrigucci, 2004; Marsili-Libelli, 2004], thus only the new results will be presented here. Starting with the problem of circadian pattern recognition for eutrophication early warning [Marsili-Libelli, 2004] the interest in dissolved oxygen dynamics has shifted to the completion of a complex ecological model of the lagoon [Giusti and Marsili-Libelli, 2006], encompassing all the lagoon dynamics, except the dissolved oxygen. In adding this last feature to the model, it was recognized that the operating conditions of the lagoon varied so widely over the years that no single model would be able to accommodate such large variations [Marsili-Libelli and Giusti, 2007], hence the need to develop a bank of differing models and use the most appropriate of them for the current conditions, which are detected by applying the procedure outlined in the previous section. Using the same clustering parameters for discriminating the diurnal cycle (minimum and maximum Dissolved Oxygen concentration and the slope of the line connecting these two extremal points) as in [Marsili-Libelli and Arrigucci, 2004; Marsili-Libelli, 2004] the new pattern recognition results are now used to select the most appropriate dissolved oxygen dynamical model, developed separately, for a long-term simulation [Marsili-Libelli and Giusti, 2007]. In fact it was found that no single model (i.e. parameter set) yields an acceptable agreement between data and the DO model, thus the problem arises of choosing the best combination of parameters in order to adapt the model to the long-term DO variations caused by environmental changes induced by the seasonality and varying physico-chemical conditions. The pattern recognition algorithm was used for model parameters patching to obtain the best response on a long time horizon. After obtaining four meaningful clusters, this number being given by the partition indicators previously mentioned in Eqs. (7 – 9), four model parameter sets were obtained by calibrating the model with respect to the prototypical pattern of each cluster, obtaining the model-data agreement on Figure 4. The actual parameter set to be used in the model for each day is obtained by fuzzy implication of the four basic prototypes, as shown in Figure 5. The degree of truth of each pattern is obtained by fuzzy comparison between the current DO pattern and the prototypes. This is obtained by using again Eq. (3), but this time in a one-pass way (i.e. dropping the k index) and not iteratively because the prototypes are now known quantities [Marsili-Libelli and Müller, 1996]

$$\mu_i = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{i_\Sigma}}{d_{j_\Sigma}} \right)^{\frac{2}{m-1}}} \qquad i = 1,\ldots, c \,, \tag{10}$$

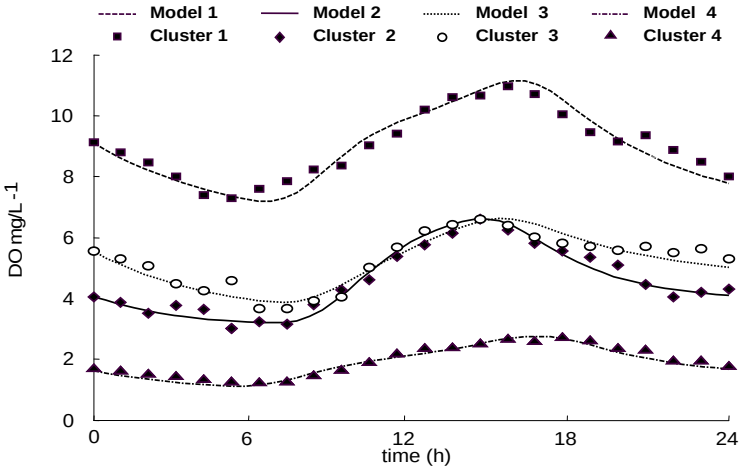


Figure 4. Calibrated response of the four basic DO models, each compared with the corresponding prototypical data sets, representing the four prototypical daily DO patterns in the lagoon.

where the $d_{\bullet_\Sigma}$ are the Σ-norm distances of the current DO pattern parameters (DO_{min}, DO_{max}, α) from the cluster prototypes, previously computed with Eq. (6), where the norm is induced by the fuzzy covariance matrix $\boldsymbol{\Sigma}_i$ defined by Eq. (4).

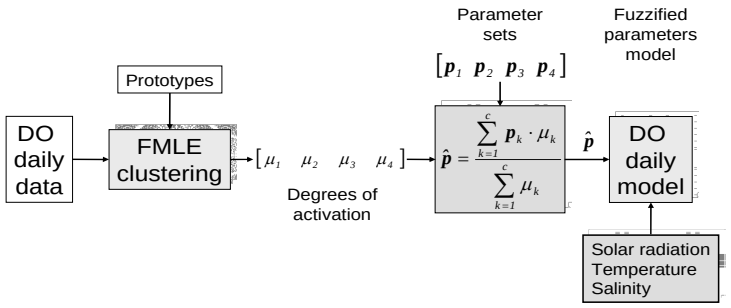


Figure 5. Model patching with fuzzy parameters. The model parameter set results from the fuzzy combination of the four parameter sets through the degrees of activation of each pattern.

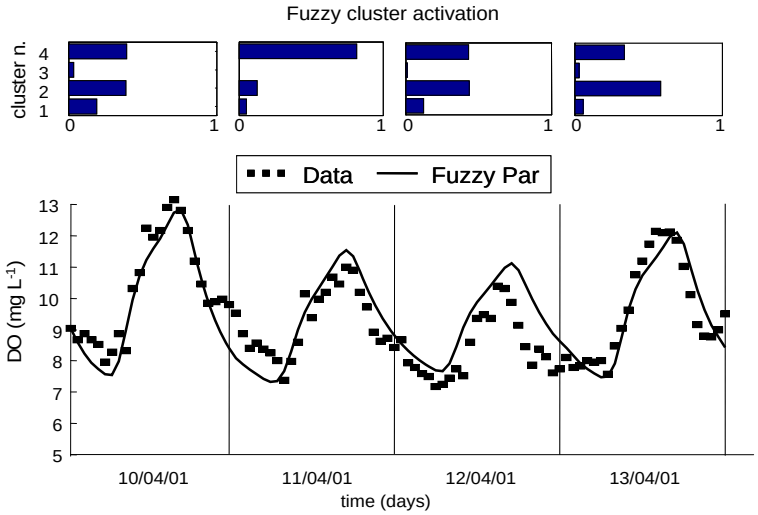


Figure 6. Performance of the patched DO model of Figure 5. The upper boxes represent the degree of activation of the four prototypes.

Figure 6 shows the performance of the patched model of Figure 5. Thus, in this application the pattern recognition algorithm is used to select the best combination of parameter values to be used in a dynamical model, in order to adapt to the varying environmental conditions in a complex ecosystem.

2.2. Daily and seasonal analysis of water quality data in the Arno river

As a second application of the pattern recognition algorithm of sect. 1.1, the detection of significant daily and seasonal trends in river quality was considered. This problem arises from the need to validate the data produced by the many automated monitoring stations deployed along the main rivers of Tuscany and their interpretation in terms of environmental changes. The basic water quality parameters (Dissolved oxygen, pH, Oxido-Reduction Potential and temperature) continuously measured by automated monitoring stations along the river. Significant variations of these parameters occur with daily and seasonal frequency. An example of these patterns is shown in Figure 7, where all the four main quality indicators follow a very clear diurnal cycle superimposed to a seasonal drift. The DO daily cycle is due to photosynthesis and is synchronous with the temperature variations.

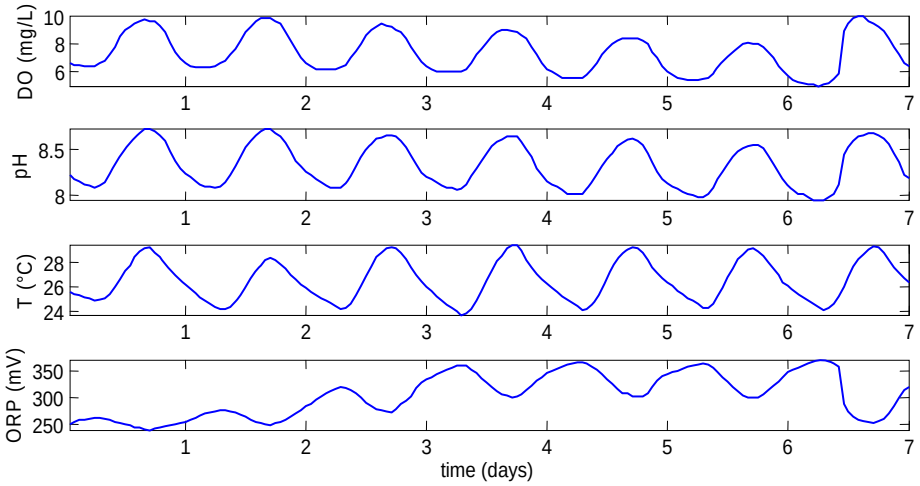


Figure 7. Typical patterns in the main four quality variables recorded in the Arno river in the week 17 – 23 July 2003.

During the night DO decreases because the photosynthetic contribution is absent, whereas the degradation processes (night respiration) are still active. Photosynthesis is active during daytime and induces a CO_2 depletion resulting in an increase in daytime pH [Chapra, 1997]. The Oxido-Reduction-Potential follows a similar path, but shows more emphasis on a long-time variation due to alkalinity changes.

The purpose of this study is to detect anomalies in the natural cycle and provide a pre-emptive discriminating power for the detection of potentially dangerous eutrophication levels. Hourly samples covering the time span of 2003 – 2007 were analysed and four clusters were selected to represent the information contained in the data.

Analysing the daily patterns, all the validity measures of Eq. (6 – 8) indicated an optimal partition of four clusters, two of which can be related to the daily photosynthesis in warm and temperate periods and two to the night respiration in the same thermal conditions, as shown in the example of Figure 8, where the differing level of photosynthesis are clearly discernible, with an exchange between the warm and cool clusters.

In addition to isolating steady state steady patterns, like the ones in Figure 8, the fuzzy clustering algorithm can detect transition induced by a change in environmental variable, such as temperature. Figure 9 shows two changes from a warm to a comparatively cooler period, shown by a decreasing temperature. However, the change in membership reveals the differing patterns of the two transitions. In the July 2003 change the cluster analysis shows a marked shift from warm to cool period behaviour with limited day photosynthesis

due to cloudy weather, whereas in the August 2006 sample the shift is much less sharp, with a sustained warm night respiration, due to the mature algal population and massive daytime radiation.

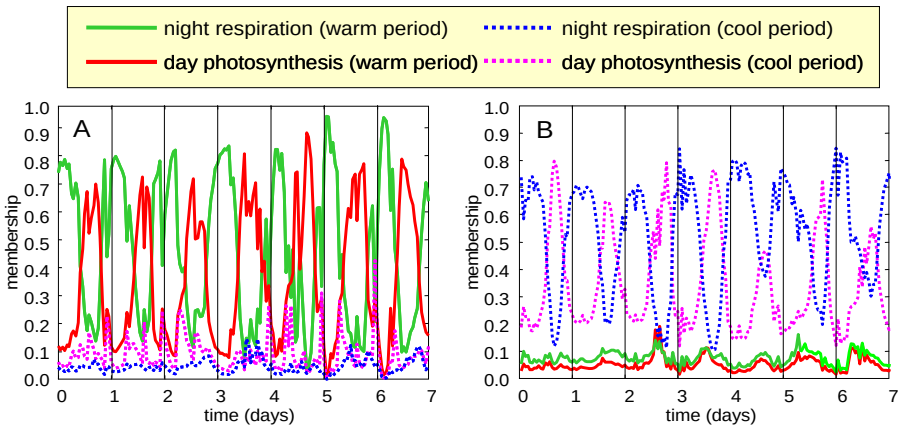


Figure 8. Steady-state alternating photosynthesis/respiration daily patterns during a warm (A) and a temperate (B) week. The vertical lines indicate the midnight of each day.

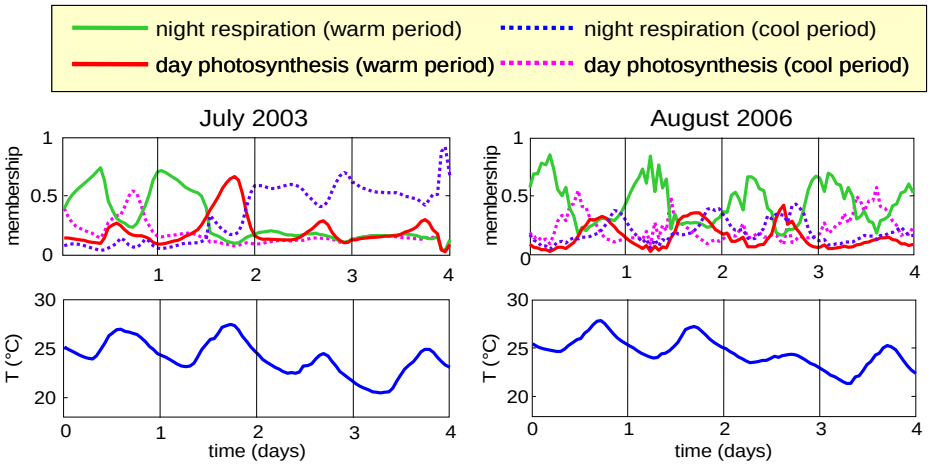


Figure 9. Transition from a warm period to a cool one. In both cases the temperature (shown in the bottom graphs) is the most evident indicator, whereas the memberships (top graphs) show differing transition modes. The vertical lines indicate midnight of each day.

Further to short-term analysis, more information can be extracted from the composite time-series covering the whole 2003 – 2007 period. Centroid displacements could be used to detect climatic anomalies reflected in the water quality parameters.

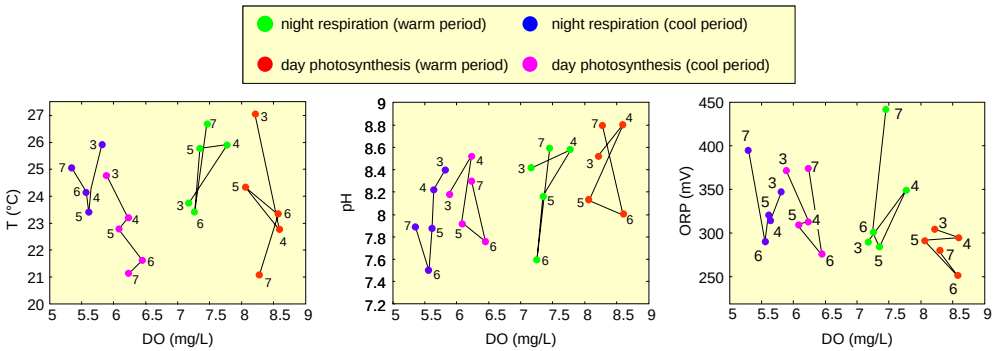


Figure 10. Centroid displacement showing the anomaly of 2004, obtained by analysing the composite time series. The numbers stand for the year after 2000.

As an example, Figure 10 shows the centroid variations in the various parameter combinations, clearly showing a continuing trend over the five years which have been considered and can possibly be related to climate changes.

Inspecting the centroid displacements over the years, shown in Figure 10, it appears that the most discriminating variable is Dissolved Oxygen (DO), with the four clusters maintaining the same relative position in all the figures, coherent with the daily trends shown in Figure 8. Clearly, the warm behaviour (day and night) correspond to the highest DO values, whereas the DO values of the cool behaviour is consistently lower. The middle plot (DO vs. pH) also reveals that an increased photosynthesis increases the pH as a consequence of CO_2 depletion, whereas the opposite is true for ORP.

3. CONCLUSION

A fuzzy pattern recognition algorithm has been presented, composed of a denoising wavelet filter and a fuzzy clustering procedure based on the Fuzzy Maximum Likelihood Estimates (FMLE) method. This composite procedure provided robust partitions and could be used in a number of environmental applications. Two of them have been presented here: the detection of daily patterns in the dissolved oxygen data from the Orbetello lagoon, and the analysis of water quality time-series from the automated measuring stations in the Arno river. In the first case, the algorithm was used to mechanize the necessary parameter-patching procedure to feed a long-term DO dynamical model for the lagoon. In the second example, the algorithm was used on two different time-scales: on a daily basis it provided a discrimination between daytime photosynthesis and night time respiration in differing climatic conditions, whereas on a yearly horizon, it could discriminate significant environmental anomalies.

As to the generality of the algorithm, in principle the two-stage procedure (wavelet denoising followed by fuzzy clustering) can be applied to any environmental problem, provided that proper indicators are preliminarily selected. The required level of wavelet decomposition should then be checked with a frequency analysis. Lastly, the effectiveness of the fuzzy partition should be checked with the indicators Eqs. (7 – 9). If these waypoints are observed, the algorithm can be a plied to a wide class of environmental problems.

4. REFERENCES

Abonyi J., *Fuzzy model identification for control*. Birkhäuser, pp. 273, Boston, 2003.

Babuska, R., *Fuzzy modeling for control.* Kluver Publ. Co., Amsterdam, 260 pp., 1998.

Bezdek J.C., 1981. *Pattern recognition with fuzzy objective function algorithms.* Plenum Press, New York, pp. 256, 1981.

Chapra, S., *Surface Water-quality Modeling*. McGraw Hill Int. Ed., New York, pp. 844, 1997.

Marsili-Libelli, S. and A. Müller, Adaptive fuzzy pattern recognition in the anaerobic digestion process, *Pattern Recognition Lett.* 17, 651 – 659, 1996.

Marsili-Libelli S. and S. Arrigucci, Circadian patterns recognition in ecosystems by wavelet filtering and fuzzy clustering. In Pahl C., Schmidt S. and T. Jakeman (eds.) *iEMSs 2004 International Congress*, Osnabrück, Germany, 2004.

Marsili-Libelli S., Fuzzy pattern recognition of circadian cycles in ecosystems. *Ecol. Model.*, **174**, 67 - 84, 2004.

Giusti, E. and S. Marsili-Libelli, An integrated model for the Orbetello Lagoon ecosystem. *Ecol. Model.* 196, 379 – 394, 2006.

Marsili-Libelli, S. and E. Giusti, Spatio-temporal modelling of the dissolved oxygen dynamics in the Orbetello lagoon by pattern recognition and fuzzy patching. Submitted to *Ecological Modelling*, 2007.

Meyers, S.D., Kelly, B.G., and J.J. O'Brien. An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Mon. Weather Rev.* 121, 2858–2866, 1993.

Percival, D.B., and A.T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, pp. 594, 2000.

Strang, G. and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Bayesian Challenges in Integrated Catchment Modelling

V. Kumar[a]*, A. Holzkaemper[a], B. Surridge[a], P.I. Rockett[c], M. Niranjan[b], D. N. Lerner[a]

[a]*Catchment Science Centre, Kroto Research Institute, University of Sheffield, Broad Lane, Sheffield S3 7HQ, UK*

[b]*Department of Computer Science, University of Sheffield, Portobello Street, Sheffield, S1 4DP, UK*

[c]*Department of Electronic and Electrical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK*

**email: vikas.kumar@sheffield.ac.uk*

Abstract: Bayesian Networks (BNs) are increasingly being used as decision support tools to aid the management of the complex and uncertain domains of natural systems. They are particularly useful for addressing problems of natural resource management by complex data analysis and incorporation of expert knowledge. BNs are useful for clearly articulating both the assumptions and evidence behind the understanding of a problem, and approaches for managing a problem. For example they can effectively articulate the cause-effect relationships between human interventions and ecosystem functioning, which is a major difficulty faced by planners and environment managers. The flexible architecture and graphical representation make BNs attractive tools for integrated modelling. The robust statistical basis of BNs provides a mathematically coherent framework for model development, and explicitly represents the uncertainties in model predictions. However, there are also a number of challenges in their use. Examples include i) the need to express conditional probabilities in discrete form for analytical solution, which adds another layer of uncertainty; ii) belief updating in very large Bayesian networks; iii) difficulties associated with knowledge elicitation such as the range of questions to be answered by experts, especially for large networks; iv) the inability to incorporate feedback loops and v) inconsistency associated with incomplete training data. In this paper we discuss some of the key research problems associated with the use of BNs as decision-support tools for environmental management. We provide some real-life examples from a current project (Macro Ecological Model) dealing with the development of a BN-based decision support tool for Integrated Catchment Management to illustrate these challenges. We also discuss the pros and cons of some existing solutions. For example, belief updating in very large BNs cannot be effectively addressed by exact methods (NP hard problem), therefore approximate inference schemes may often be the only computationally feasible alternative. We will also discuss the discretisation problem for continuous variables, solutions to the problem of missing data, and the implementation of a knowledge elicitation framework.

Keywords: Bayesian networks; Integrated Catchment Management; Decision support for natural resource management.

1. INTRODUCTION

The European Water Framework Directive (WFD) sets out an integrated perspective to water management where the river catchment and river basin district are the scales of focus. It encourages the active involvement of all affected parties within the planning process [Giupponi, 2007]. At these large spatial scales, there are always competing and often conflicting management challenges that need to be addressed, such as demands for water of good quality, the integrity of aquatic ecosystems and flood risk management. In turn, the water related objectives must be pursued within the broader context of economic, social, cultural and other environmental objectives for the catchment as a whole. Given these challenges, decision making must be supported by new tools which evaluate the potential impacts of planned management interventions on multiple objectives.

These new tools are often equated with integrative models. Integrative models are typically developed for five main reasons: i) prediction; ii) forecasting; iii) management and decision making; iv) social learning; v) developing system understanding and experimentation. The types of models suited to integrated catchment management include system dynamics, Bayesian networks, metamodels, coupled complex models, agent based models and expert systems. In this paper we focus on Bayesian networks.

Bayesian networks provide a useful tool to assist in the structuring and analysis of decision problems [Watthayu and Peng, 2004]. A Bayesian network is a decision analysis framework, based on Bayesian probability theory, which allows the integration of scientific and experiential knowledge, and the uncertainty associated with this knowledge [Castelletti and Soncini-Sessa, 2007]. The approach involves describing a system in terms of variables and linkages, or relationships between variables, at a level appropriate to the decision making. This is achieved through representing linkages as conditional probability tables and propagating probabilities through the network to give the likelihood of variable outcomes [Murphy, 2001]. Therefore, the approach ensures that the treatment of risks and uncertainties is an intrinsic part of the decision-making processes [Borsuk et al., 2004]. The Bayesian network is flexible and interactive, and hence if a previously developed network does not fit a user's conceptual understanding of the system, it can be adapted quickly and simply to the cognitive understanding of the user.

The scope and feasibility of BNs in integrated catchment modelling are currently being investigated by the Catchment Science Centre (CSC) at the University of Sheffield. The model, termed the Macro-Ecological Model (MEM), is being designed to model the links between the technical, economic and social processes which interact within any given catchment. The aim is to develop a decision support tool, combining a simulation model of the causal relationships within catchment processes with plausible scenarios of management options. The tool will estimate a set of indicators which stakeholders can use to inform their decisions. In this paper we discuss some of the challenges of BNs in developing this model.

2. BN CHALLENGES IN INTEGRATED CATCHMENT MODELLING

2.1 Building large networks

Developing an integrated catchment model like the MEM involves many components which, when translated into a BN become a large network. From a practitioners' point of view, the process of compiling and executing a BN, using the latest software tools and improved computational power, is relatively easy. Also, the accuracy and speed of current algorithms makes it feasible for an application in integrated catchment management. However, the problems of building a complete BN for a particular "large" problem remain a complex task. Designing the right network structure is a prerequisite for meaningful elicitation of any probabilities. It needs proper software and knowledge engineering practice during the development phase. Knowledge engineers work with experts to decompose the system, recognise patterns at the macro and micro level [Shaw and Garland, 1996] and continually change the model as both sides' understanding increases. The

benefits of constructing software systems from components or modules are well known and the properties that modular systems must contain were articulated as early as 1972 [Parnas, 1972]. However, little work has been done on applying modular structures in BN design. Laskey and Mahoney recognised that BN construction required a method for specifying knowledge in larger, semantically meaningful, units they called network "fragments" [Laskey and Mahoney, 1997]. Laskey and Mahoney also argued that current approaches lacked a means of constructing BNs from components. The available development software for BNs lack sophisticated software engineering tools compared to the modern software development tools available for the mainstream software industry where these forms parts of an integrated development environment. In the MEM project, we follow a "bottom up" approach decomposing the model into fragmented modules that, when joined together, form the complete system. A fragmented module, which we call a sub-network, is a set of related system variables that could be constructed and reasoned about separately from other sub-networks. However consistency must be maintained in defining the common variables across different sub-networks. Ideally, sub-networks must make sense to the expert who must be able to supply some underlying motive or reason for the variables belonging together. In Figure 1 a simplified subset of the MEM network has been depicted showing how the developed sub-networks are linked together. The focus in Figure 1 is on the 'connector' variables of "PO_4 concentration" and "River discharge" that enable the different sub-networks to be connected. Another approach could be to use an object-oriented methodology where fragmented modules become classes, both variables (nodes) and instantiated BN fragments become objects and encapsulation is implemented via an interface and private variables [Koller and Pfeffer, 1997]. Connector variables can be used to specify interfaces and intermediate variables can encapsulate private data. However this can be difficult to follow strictly because different sub-networks can contain the same connector node as an intermediate variable. When creating an object of a sub-network we must know whether it shares intermediate nodes with other sub-network objects in order to define the influence combination rule. Clearly this is not a problem when the influence combination rule treats all parent nodes equally irrespective of type and value, as a form of dynamic polymorphism, but such a combination rule would be very difficult to conceive and implement [Neil et al., 2000].

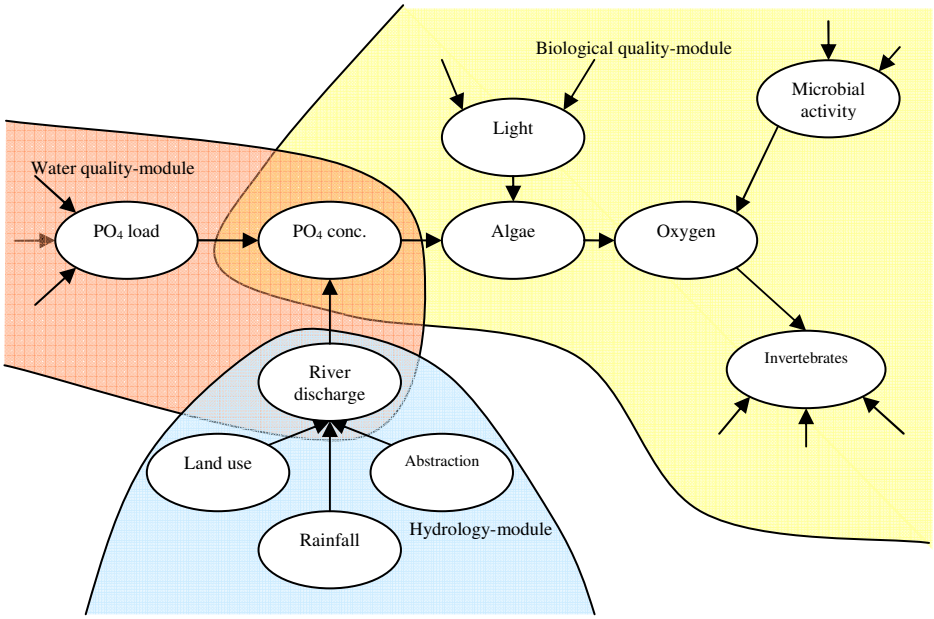


Figure 1. Subset of the MEM network showing how the developed sub-networks are linked together

2.2 Inventing hidden node

One of the major steps in large domain modelling is structural simplification. Such simplification is necessary for greater compactness and to reduce the computational complexity. A more interesting problem is inventing hidden nodes. Hidden nodes can make a model much more compact (see Figure 2). Introducing hidden nodes to the network structure may reduce the computational complexity of a Bayesian network by reducing its dimensionality, and may also help to capture non-trivial (higher order) correlations between observed events. However hidden nodes may also be included in a Bayesian network structure to reflect expert domain knowledge regarding hidden causes and functionalities which impose some structure on the interaction among the observed variables.

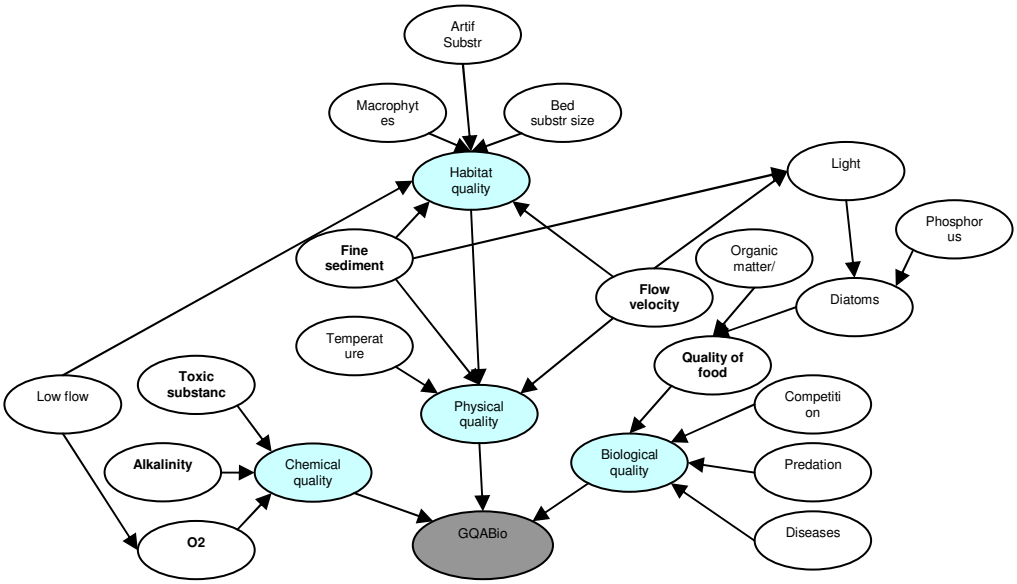


Figure 2. Conceptual biological quality network where light blue bubbles represent hidden nodes, white bubbles represent observed variables and the grey bubble represents the index variable (GQABio = General Quality Assessment score for Biology).

Generally hidden nodes should be added in consultation with experts or they should be added during structural learning. One problem is choosing the cardinality (number of possible values) for the hidden node, and its type of Conditional Probability Distribution. Another problem is choosing where to add the new hidden node. There is no point making it a child, since hidden children can always be marginalized away, so we need to find an existing node which needs a new parent, when the current set of possible parents is not adequate [Heckerman, 1995]. Furthermore, interpreting the "meaning" of hidden nodes is always tricky, especially since they are often unidentifiable. One way is to follow fully automated structure discovery techniques which can be useful as hypothesis generators and which can then be tested by experiment. In the MEM the simplification process has been done in consultation with domain experts and hidden nodes are generally some proxy indices used by the Environment Agency.

2.3 Learning with distributed datasets

Training datasets have been classified as fully observable (which means that the values of all variables are known) and partially observable (meaning that we do not know the values of some of the variables) [Heckerman, 1995]. Partial observability might occur because variables are measured at different spatial points (we call this a distributed dataset), through a change in survey strategy (introducing or excluding some of the variables), due to missing

records for certain observation (called missing variables), or simply because certain variables cannot be measured (hidden variables). Learning approaches from partial observability are confined to dealing with the issue of missing values. When some variables are missing, the likelihood surface becomes multimodal. The problem of simple missing values in the dataset has been addressed very simply in Bayesian networks, because likelihoods can be computed using normal iterative methods (such as gradient decent or Expectation-Maximization (EM)) [Heckerman, 1995]. These iterative methods try to find a local maximum of the Maximum Likelihood (ML) or Maximum A Posterori (MAP) function [Murphy, 2001]. This capability of BNs has been cited as very useful as there is no minimum data requirement to perform the analysis and BNs take into account all the data available [Myllymaki et al., 2002]. However, there are many practical cases where the observed dataset is distributed among different sites which make a distributed heterogeneous dataset scenario, where each site has observations corresponding to a subset of the attributes. In some cases we have only a small set of overlapping data points from which it is not sufficient to derive casual relationships. It will be difficult or impossible (depending on the scenarios) to find local maxima in the ML/MAP function for such distributed datasets. Iterative algorithms need to use an inference algorithm to compute the expected sufficient statistics which will be difficult for an extreme missing value problem. So learning from distributed datasets is not possible with conventional learning methods.

In the MEM project we have tried to tackle this problem using an Artificial Neural Network (ANN) approach. The goal was to build a constant predictor by using distributed datasets. Different ANN predictors were built mapping subsets of predictor variables with response variables which are combined into a single model over a weighted average. The weights are determined by training based on the small overlapping dataset and they represent the prediction capability of each sub-model.

2.4 Discretisation of continuous variables

A large number of the observation data required for the development of the MEM are continuous data (e.g. phosphate concentration). Bayesian networks can deal with continuous variables in only a limited manner. If continuous variables are to be incorporated in BNs some means must be found of *optimally* partitioning the values into sub-ranges which can then be treated as discrete categories. The way this discretisation is performed affects the performance of the subsequently derived BN model.

In order to maximise the predictive power of BNs, it is necessary to discretise each of the continuous input variables in some optimal manner. Every continuous variable could require differing numbers of bins which might well also be of non-uniform width. The problem is thus one of identifying the optimal set of binning parameters. Genetic algorithms (GAs) have proved very successful at solving this sort of discrete search problem and we propose employing them here. GAs work by evolving a population of possible solutions in a manner analogous to Darwinian survival of the fittest. For N continuous input variables to the BN, we require some set of bin widths spanning each variable's range. The number of bins and their widths can differ between and indeed within variables. We propose constructing a GA individual (a chromosome) as a set of N lists of binning schedules, one list per BN input variable. At each iteration of the GA, stochastically-selected pairs of individuals will be crossed-over and mutated using specially adapted genetic operators; the objective function 'driving' the evolution will be the predictive power of a BN employing the binning schedule specified by a given chromosome. In this way, the set of bin specifications yielding the (near-) optimal performance will be generated. We envisage significantly improved predictive power for BNs. Figure 3 shows some results of this exercise where we show performance comparison of a GA approach with the classical 1R method proposed by Holte et al. [1989]. There is a marked increase in predictability of the target variable with the GA approach. The technique should also be applicable to decision trees which are a very widely used learning paradigm operating on nominal variables.

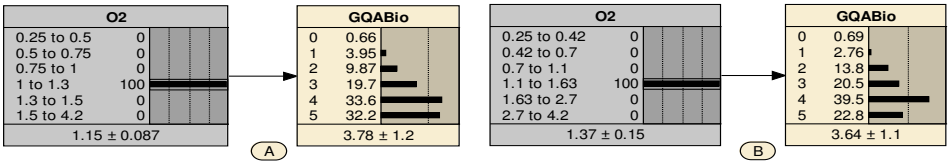


Figure 3. Performance of discretisation algorithms: (A) 1R method [Holte et al., 1989]; (B) Optimization using Genetic algorithm.

2.5 Expert knowledge elicitation and structuring

Bayesian networks allow for casual relationship (probabilities) to be specified based on subjective assessments ("expert opinion"), empirical evidence, or a combination of both. Incorporating expert opinion is important because, often, there are insufficient data to learn and model relationships between casual factors and outcomes using population-based data ("machine learning"). When there is a paucity of data, domain expert opinion can be used to create Bayesian networks. Expert-derived probabilities can be improved over time with observational data from multiple sources, obviating the need for a single data repository that contains all casual relationships. However the number of probability distribution required to populate a Conditional Probability Table (CPT) grows exponentially with the number of parent nodes. The sheer volume of questions to be answered by the experts poses a considerable cognitive barrier. Also extracting knowledge in a form that can be converted into probability distributions may prove difficult in real life situations. This is because many field experts are used to working with real sampling or experimental data, and may find it difficult to provide any numbers without relying on data. Also they may be used to classical statistical analyses and feel uncertain when trying to think about their knowledge in terms of probability distributions rather than point estimates and confidence intervals. This uncertainty together with only superficial knowledge about the methodology may also lead to distrust towards the BNs, which easily leads to reluctance to provide the estimates.

Some well-known methods in this area are the Noisy-OR model [Pearl, 1998] and some improvement and generalisation by others [Henrion, 1989; Srinivas, 1993]. These models can compute the distributions needed for the CPT from a small set of questions elicited from the expert. All these models, however, are constrained by the assumption that parents act independently without synergy [Srinivas, 1993].

For the MEM project we are implementing a modified version of the relative weight and compatible probability method proposed by [Das, 2004]. The input to the algorithm consists of a set of weights that quantify the relative strengths of the influences of the parent nodes on the child node and a set of compatible probability distributions. The number of questions grows linearly with the associated parent nodes. We introduced a special case when certain parent nodes are critical and have some thresholds, above or below which the effect of other parent nodes is none or minimal. For example in the biological quality network shown in Figure 4, oxygen can be a critical variable. We asked a few more questions to the experts to elicit the critical variables and their critical states. The same set of questions has also been used for cross validation and intra consistency check of expert knowledge. Figure 4 shows a simple outcome comparing results from the original approach of Das with those from the modified approach we have used in the MEM project. Because the variable 'oxygen' is a critical variable under approach B, its existence in a low state has a far stronger influence on the outcome of GQABio than under approach A where oxygen has not been identified as a critical variable.

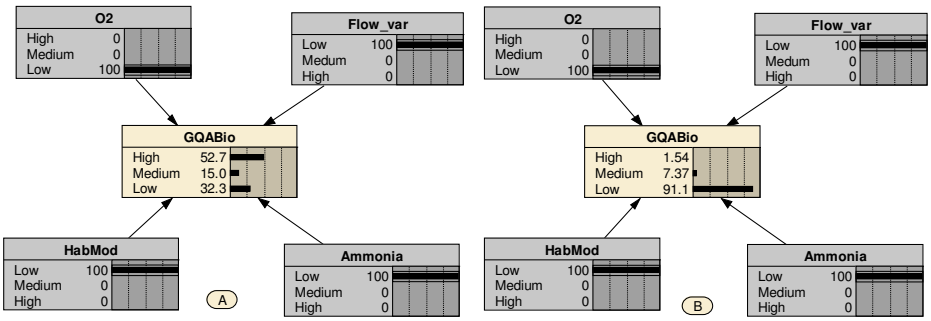


Figure 4. Results for a critical case (Oxygen low) (A) implemented with [Das, 2004] knowledge elicitation approach (B) Modified version of [Das, 2004] developed for MEM.

2.6 Dynamic model structure and feedback loops

BNs are static models and as such not able to integrate system dynamics and feedback loops. However, there are different ways of dealing with this issue when applying BNs for modelling dynamic systems. The simplest approach is to assume that the system under consideration is in equilibrium. For example in the MEM sub-network of water quality, phosphate concentration in the river has been taken as annual mean PO_4 concentration. A more sophisticated approach to incorporate temporal and ambient aspects is the Dynamic Bayesian Network (DBN) approach. The predominant DBN literature [Korb and Nicholson, 2004; Murphy, 2002] deals with contiguous time slice models based on the Markov assumption that the current state of the model depends on its previous time state and action taken in its current state. However a known impediment of this approach is the exponential increase in network size apart from the knowledge acquisition problem. Figure 5 shows a simple two-node feedback loop for 3 time steps (A0-A3 and B0-B3). Many solutions are being proposed to solve feedback loops in DBNs but most of them still need to be tested in practical applications. For example, Gossink et al. [2007] have proposed an extended version of DBN to ease the "intractability of knowledge acquisition" problem. The inference algorithm used for this approach is likelihood weighted approximation. Exact algorithms are still not applied with this approach and modification of the extended DBN may be needed [Gossink et al., 2007].

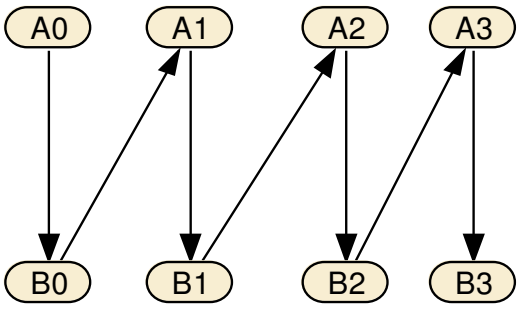


Figure 5. Dynamic Bayesian Network representing a feedback loop.

Even if we decided to build a time sliced DBN, inference is another problem at a later stage. It needs to unroll the network to represent dependencies between nodes from two consecutive time slices. Further, it needs to preserve conditional relationships between nodes as well as influence that evidence nodes have on hidden nodes (if there are any). Parameter learning can be difficult and the main reason is unquantified uncertainty in the elicited probabilities. Further, a dynamic network can have a static node which creates a problem of dependency and representation. [Schafer and Weyrath, 1997] claimed that the problems in this type of model can be differentiated in two classes: (a) the dependency between dynamic child node and static parent node, and (b) the initial value of static node is not known with certainty, and we can only estimate this node's value which could be changed as time progresses. Schafer and Weyrath claimed that the dependency can be interpreted by incorporating the effect of the static node in the dynamic node, so the dynamic node will

convey all of the information needed for inference. Problem (b) can be solved if we compromise accuracy for efficiency in a way that can delay rollup until we receive the exact probability of the static node.

2.7 Belief updating

BNs for a large domain like ICM are implemented as a modular network consisting of many subnets joined together to form the complete system. Consistency among subnets in an integrated BN is achieved by communication. When a subnet updates its belief, it communicates with an adjacent subnet to maintain the consistency across the system. Large-scale models generally have high-order stochastic dependencies. The computational complexity of treating these dependencies exactly can be characterized in terms of the size of the maximal clique of the "moralized" graph [Murphy, 2002]. Existing inference algorithms [Xiang, 1995] require repeated belief propagations (which is proportional to the number of linkages between the subnets) within the receiving subnet. It has been stated that singly connected BNs are tractable and have time algorithm linear in the number of nodes in a network or the size of the network for exact inference [Cooper, 1990]. On the other hand, multiply connected Bayesian networks are intractable and do not admit efficient algorithms for exact inference in the worst case which makes the exact inference in BNs to be NP-hard [Cooper, 1990]. Others [Murphy, 2002; Xiang, 2000] believe that inference in a BN can be performed effectively using its junction tree (JT) representation. However, the junction tree propagation method cannot compute p(X|e) when X is not contained in a node of the junction tree. Also, the local propagation procedure has to be applied whenever new evidence is observed. [Xiang, 2000] has worked on extending junction trees beyond single BNs and tried to prove that two local propagations are sufficient for propagating evidence from one JT to an adjacent one no matter how many linkages there are between the two JTs. This results in big computational savings which are particularly significant for a large domain with subnets comprising multiple linkage JTs.

The complexity of the exact inference algorithm does not mean that we cannot solve inference in large BNs. It implies that we cannot find a general procedure that works for all networks. However, depending on the network complexities, we can have an efficient algorithm either exact or approximate. Various algorithms e.g. approximate, heuristic, hybrid or special case algorithms should be taken into consideration for a large domain network. A future research goal should be that of identifying effective approximate algorithms that work well for large domain problems. Also the integration of various kinds of exact and approximate algorithms exploiting the best of each can be an interesting area of research.

There are other problems which we have not discussed in this paper but which are relevant for ICM modelling. In the missing value problem there is a particular case reported by [Rubin, 1978]: failure to observe a variable may in itself be informative about the true state of the system. For example, in habitat survey data, failure to report a population of a certain species may suggest that system state is not suitable for survival of that species. In medical science, researchers have worked on this problem and proposed methods for dealing with such situations [Chickering and Pearl, 1996; Robins, 1986; Rubin, 1978].

3. CONCLUSION

Bayesian networks are still a relatively new approach but they are gaining popularity in many application areas, including environmental applications. Many easy-to-use BN development software packages are available but they are still lacking the sophistication and robustness of a standard development environment. Also they are not updated with the ongoing developments in the field of BN research. Integrated Catchment Management modelling provides its own set of challenges to BNs. Questions that need to be addressed in the context of applying BNs for ICM are: Do we have data of sufficient quantity and quality? Is expert knowledge sufficient for model characterisation? Which algorithm should we use for inference, in order to extract valuable information and keep the computational complexity low? There are number of promising research avenues, challenges, and needs

that directly concern the usefulness of Bayesian networks and their impact on real world environmental applications.

ACKNOWLEDGEMENTS

The research was funded by the Environment Agency's Science Project (SC060039) and the University of Sheffield. Additional funding was provided by the European Union through the Marie Curie Early Stage Training project 'CatSci' at the University of Sheffield.

REFERENCES

Borsuk, M. E., C. A. Stow, and K. H. Reckhow, A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling, 173*(2-3), 219-239,2004.

Castelletti, A., and R. Soncini-Sessa, Bayesian Networks and participatory modelling in water resource management. *Environmental Modelling & Software, 22*(8), 1075,2007.

Chickering, D., and J. Pearl, *A clinician's tool for analyzing noncompliance.* Paper presented at the In Proceedings of the Thirteenth National Conference on Artificial Intelligence, Portland,1996.

Cooper, G. F., The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence, 42*(2-3), 393-405,1990.

Das, B., Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem. *CoRR, cs.AI/0411034*,2004.

Giupponi, C., Decision Support Systems for implementing the European Water Framework Directive: The MULINO approach. *Environmental Modelling & Software, 22*(2), 248,2007.

Gossink, D., M. Shahin, J. Lemmer, and I. Fuss, *A practical approach to Dynamic Bayesian Networks.* Paper presented at the Information, Decision and Control, Adelaide, Australia,2007.

Heckerman, D., *A Tutorial on Learning With Bayesian Networks* (No. MSR-TR-95-06),1995.

Henrion, M., *Some practical issues in constructing belief networks* (Vol. 3). North Holland: Elsevier Science Publishers,1989.

Holte, R. C., L. Acker, and B. W. Porter, *Concept learning and the problem of small disjuncts.* Paper presented at the Proceeding of Eleventh International Joint Conference on Artificial Intelligence, Detroit,1989.

Koller, D., and A. Pfeffer, *Object-oriented Bayesian networks.* Paper presented at the Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence,1997.

Korb, K. B., and A. E. Nicholson, *Bayesian Artificial Intelligence*: Chapman & Hall/CRC,2004.

Laskey, K. B., and S. M. Mahoney, *Network fragments: representing knowledge for constructing probabilistic models.* Paper presented at the Proceedings of Thirteenth Annual Conference on Uncertainty in Artificial Intelligence,1997.

Murphy, K. P., A Brief Introduction to Graphical Models and Bayesian Networks, 2001.

Murphy, K. P., *Dynamic Bayesian Networks: Representation, Inference and Learning.* University of California, Berkeley,2002.

Myllymaki, P., T. Silander, H. Tirri, and P. Uronen, B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis. *International Journal on Artificial Intelligence Tools, 11*(3), 369-387,2002.

Neil, M., N. Fenton, and L. Nielsen, Buidling large-scale bayesian networks. *The Knowledge Engineering Review, 15*(3), 257-284,2000.

Parnas, D. L., On the criteria to be used in decomposing systems into modules. *Communications of the ACM, 15*(12), 1052-1058,1972.

Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*. San Francisco, California: Morgan Kaufmann Publishers,1998.

Robins, J., A new approach to causal inference in mortality studies with sustained exposure results. Mathematical Modelling, 7, 1393-1512,1986.

Rubin, D., Bayesian inference for causal effects: The role of randomization. *Annals of Statistics, 6*, 34-58,1978.

Schafer, R., and T. Weyrath., *Assessing Temporal Variable User Properties with Dynamic Bayesian Networks.* Paper presented at the Proceeding of the sixth International Conference on User Modeling,1997.

Shaw, M., and D. Garland, *Software Architecture*: Prentice Hall,1996.

Srinivas, S., *A generalization of the noisy-OR model.* Paper presented at the Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, San Mateo, CA,1993.

Watthayu, W., and Y. Peng, *A Bayesian network based framework for multi-criteria decision making.* Paper presented at the MCDM 2004, Whistler, British Columbia, Canada.,2004.

Xiang, Y., *Optimization of inter-subnet belief updating in multiply sectioned Bayesian networks.* Paper presented at the Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal,1995.

Xiang, Y., Belief updating in multiply sectioned Bayesian networks without repeated local propagations. *International Journal of Approximate Reasoning, 23*(1), 1,2000.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Developing Bayesian network models within a Risk Assessment framework

Carmel A. Pollino[1] and Barry T. Hart[2]
[1]The Fenner School of Environment and Society, The Australian National University, Canberra, Australia, [2]Water Studies Centre, Monash University, Melbourne, Australia

Abstract: The risk assessment framework is increasingly being applied to examine both human and non-human stressors on ecological systems. Risk-based decision-making aims to quantify the likelihood of a threat occurring, the consequences of this to an ecological system, process or value, and the associated uncertainty in the predictions. Until recently, the ability to predict changes in dynamic ecosystems due to stressors was limited by both the poor understanding of the drivers of ecological processes and structure, and the lack of modelling tools that could represent such complexity with associated uncertainties. However, the recent growth in the use of Bayesian network tools for ecological risk assessments has resulted in major advances in better understanding and managing ecosystems despite their inherent complexity. Bayesian networks have the advantage of being able to investigate the impacts of multiple stressors in complex environments, while explicitly acknowledging the associated uncertainties resulting from inherent variability and lack of knowledge of ecological systems within an adaptive framework. Bayesian networks have the flexibility to incorporate diverse knowledge systems, ranging from 'gut feel' to quantitative process-based or simulation models. In this paper, we discuss the relationships between the risk assessment framework and Bayesian network building process, and will illustrate the main concepts with a series of Bayesian network models.

***Keywords**:* Bayesian network; risk assessment; ecology

1. INTRODUCTION

The development of quantitative models to support decision-making in environmental management is considered to be of high priority. Recently, ecological risk assessment and risk management approaches have arisen as an approach to improve integrated decision-making processes. The ecological risk assessment framework offers a formal approach to decision-making, allowing for a greater understanding of decision processes and how they relate to ecological endpoints to be achieved. Bayesian networks are being utilized to produce complex quantitative models that are both pragmatic and scientifically credible. In this paper, we overview the risk assessment and Bayesian networks cycles, and discusses the relationship between these cycles using two case studies.

1.2 Framework for undertaking a Risk Assessment

Risk is defined as a state of uncertainty where some of the possibilities involve a loss, catastrophe, or other undesirable outcome [Hubbard, 2007]. Consequently, uncertainty is a key component in assessing risk. In an ecological context, the most common sources of uncertainties include poor or incomplete understanding in the linkages between ecological processes within a system, and the inherent variability associated with such processes. As risk is the product of likelihood/hazard and consequences/effects, it is not a directly measurable attribute.

Risk assessment is the initial step in risk management. It involves the determination of risk related to a past or current situation and a recognised threat. It also provides a robust process that incorporates a transparent, scientific, precautionary and sustainable approach to the management of environmental risks.

The ecological risk assessment framework, involves a number of key steps:

- *Defining the problem* –Defining assessment objectives and scope, identifying and engaging stakeholder groups.
- *Defining important ecological assets, and identifying hazards to these assets* - hazards are identified in the context of ecological assets, activities that mediate hazards are also identified. A conceptual model/influence diagram is constructed, preferably in collaboration with stakeholder groups.
- *Analysing the risks to the ecological values* – the risk analysis process used needs to be appropriate for the situation in order to provide adequate information for decision-making. Both qualitative and quantitative information types should be incorporated in the risk analysis step. The assumptions, strengths and limitations of the analysis should be documented.
- *Characterising the risks* - the technical details of risk analyses needs to be made accessible to decision-makers and broader stakeholders. In particular, the uncertainties and assumptions associated with analyses require careful and transparent documentation.
- *Making decisions* – selection of the best management option or strategy will be the one that results in the effective minimisation of the ecological risks, while also being cost-effective and acceptable to the stakeholders. Guidance is provided on a number of multi-criteria methods for assisting this process.
- *Managing the risks* – a risk management plan provides recommendations on managing or mitigating all high or unacceptable risks. The risk management plan should include a robust program to monitor progress to ensure the strategies are working, and a review and feedback process for making changes if needed.

These steps are shown in Figure 1.

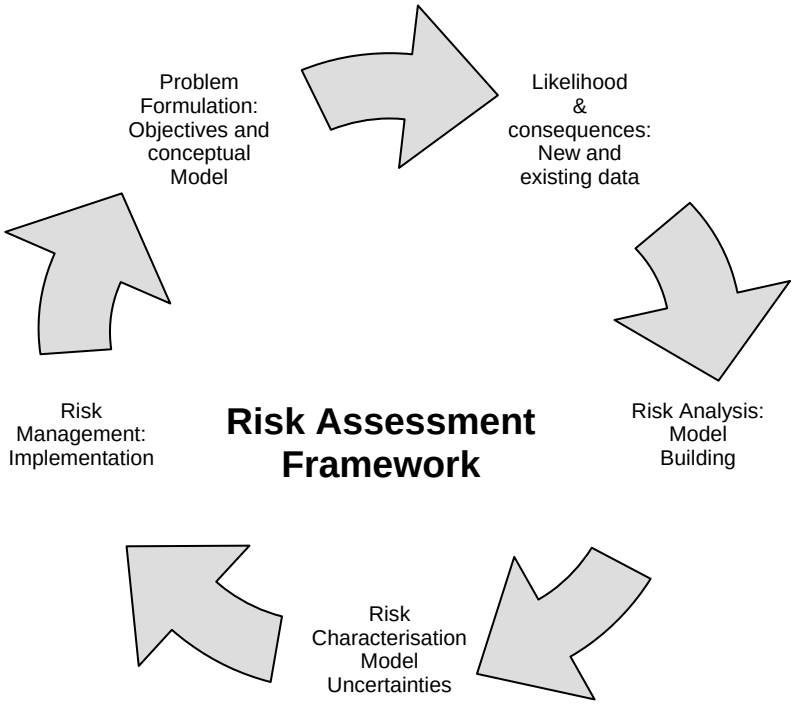


Figure 1: Risk Assessment Framework

1.1 Framework for constructing Bayesian networks

Bayesian networks (BNs), also known as Bayesian Belief Networks (BBNs) and Belief Networks, are increasing being used in ecological applications as they offer a pragmatic and scientific approach to modelling complex ecological systems where high uncertainties exist. Unlike many other ecological modelling approaches, BNs can utilise prior knowledge and data to model systems. Furthermore, BN models are particularly useful for analysing and communicating causal assumptions not easily expressed using mathematical notation, and for analysing multivariate and complex relationships among variables. Of particular note, as BNs promote stakeholder involvement, they fit into the participatory decision-making approach advocated in environmental decision-making. They also provide a platform in which disciplines can work together in a more integrative fashion.

BNs are graphical models that use probabilistic expressions to describe the relationships among variables. They are able to explore and display causal relationships between key factors and final outcomes of a system in a straightforward and easily understood manner. A *prior probability* represents the likelihood that an input parameter will be in a particular state. The *conditional probability* calculates the likelihood of the state of a variable given the states of input variables affecting it. And the *posterior probability* is the likelihood that a variable will be in a particular state, given the input variables, the conditional probabilities, and the rules governing how the probabilities combine. BNs use Bayes' Theorem to update or revise the beliefs of the probabilities of system states taking certain values, in light of new evidence.

As BNs are causal, they can also be used to calculate the effectiveness of interventions, such as management decisions, and system changes, such as those predicted for climate change. Importantly, the uncertainties associated with these causal relationships can also be explored at the same time. BNs are able to maintain clarity by making causal assumptions explicit [Stow & Borsuk, 2003] and are often used for modelling when relationships to be described are not easily expressed using mathematical notation [Pearl, 2000]. BNs use the network structure to calculate the probability certain events will occur, and how these probabilities will change given subsequent observations or a set of external (management) interventions.

1. Structure of a Bayesian network

The first step in constructing a BN is to develop a causal structure with relevant variables (nodes) and dependencies, which forms the model framework. Criteria for inclusion of variables in a BN are that the variable be: (a) manageable, (b) predictable, or (c) observable at the scale of the management problem [Borsuk *et al.*, 2004]. Any processes or factors not included become part of the uncertainty of the network, forming the predictive uncertainty described in probability distributions. Within a Bayesian network, sub-networks can be nested to describe physical or chemical processes relevant to the spatial scale specified.

2. Discretization of nodes (assigning states)

States or condition of the variables can be categorical, continuous or discrete. In order to represent continuous relationships in a Bayesian network, a continuous variable must be divided or discretized into states. The states of a variable can be numerical ranges (< 3, >3) or expressions (that can also represent data if appropriate, e.g. acceptable ≤ 3, unacceptable >3). If relevant, these states can represent targets, guidelines, existing classifications or percentiles of data.

3. Specification of prior probabilities

After defining states, linkages between nodes need to be described. Parent nodes lead into child nodes, the outcome of child nodes are conditional on how the parent variables combine. This is relationship is defined using conditional probability tables (CPTs). CPTs can be derived via one or a combination of methods:

- Direct elicitation of scenarios from expert;
- Parameterisation from datasets;
- Equations that describe relationships between variables.

4. Calculating posterior probabilities

Data or new knowledge can be incorporated into BNs and used to calculate posterior probabilities. Data sources can be entered into the network as a series of 'cases'. Cases can represent data collected during a monitoring exercise, undertaken as part of a research study, and so on.

5. Model evaluation

A range of validation tools can be used for BNs. Evaluation can involve data or technical experts, or both. Quantitative evaluation with data is preferable. Such measures include predictive accuracy and sensitivity analyses.

Predictive accuracy tests are used to determine model error rates, which are quantified using data (although not the same data used for model parameterisation). This method measures the frequency with which the predicted node state (that with the highest probability) is observed, relative to the actual value. Outcomes can be used to identify weaknesses in the model, and where more effort can be targeted in order to improve model accuracy.

Sensitivity analysis is used to identify the key drivers in the model and major knowledge gaps in our understanding. Sensitivity analysis of mathematical models can be used to investigate the uncertainties and inaccuracies in model structure, relationships and outputs, and subsequently identify where priority knowledge and data gaps exist. Thus, based on these results, recommendations for targeted monitoring and research studies can be made.
Sensitivity analyses provide a ranking of importance of variables, relative to the variable of interest (usually the endpoint). These variables indicate where better quantification in the network should be investigated and identify the most influential variables on model endpoints. Subsequently, these are the variables that should be given greater attention. In a management context, it is these variables that may represent key management actions or knowledge gaps. As sensitivity findings can differ for different spatial areas of interest or scenarios tested, key knowledge gaps and priority risks can also differ.

6. Knowledge gaps and priority risks

Having established the structure of the model, and the relationships used to drive the model, the key knowledge gaps in our understanding and priority risks can be identified. Sensitivity analysis can be used to examine where key uncertainties in model linkages exist.

7. Testing management scenarios

Management scenarios can be tested by entering new information into the network as evidence, directly changing the distribution of probabilities on the node itself.

2. BAYESIAN NETWORKS IN A RISK ASSESSMENT FRAMEWORK

Ecological risk assessment frameworks are increasingly being used as they add rigor to traditional natural resource management decision-making [Burgman, 2005; EPA, 1998] addressing limitations of traditional approaches to natural resource management, such as decisions being based on expert judgements alone, or on single focus assessments. Expert opinion (and weight of evidence approaches) are subject to cognitive and knowledge-based bias [Newman & Evans, 2002; Pollino & Hart, 2006]. Whereas assessments that only have a single focus or hazard relationships, also introduce bias by not integrating other system processes. Although such assessments are limited by the information available, they still undergo little review or updating.

Many natural resource models for decision-making still only poorly represent uncertainties associated with predictive outputs. Without characterisation of uncertainties and focus on ecological systems as a whole, the prioritisation of hazards and subsequently prioritisation of management actions has been problematic. The recent growth in the use of Bayesian network tools [see Henderson *et al.*, 2008] for ecological risk assessments has resulted in major advances in better understanding and managing ecosystems despite their inherent complexity.

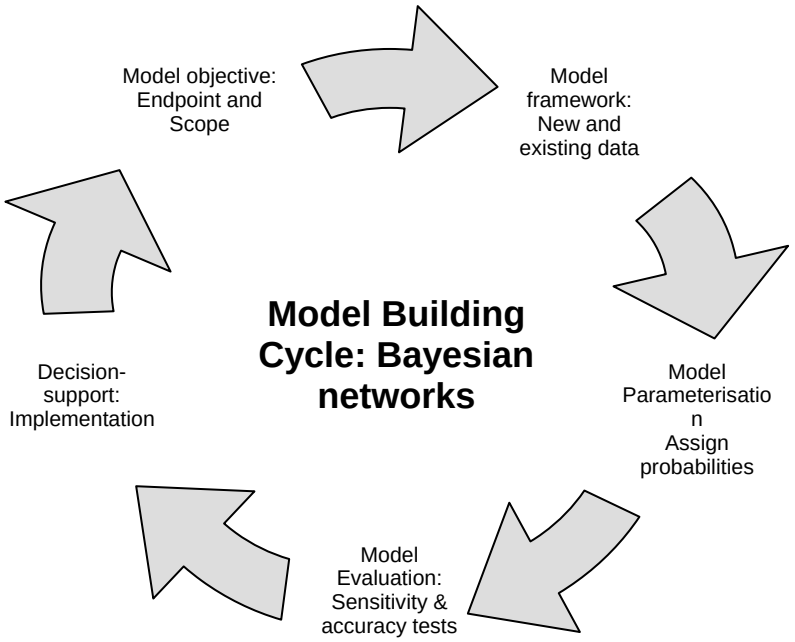


Figure 2: Steps followed in building a Bayesian network.

Abductive inference and its quantification by means of Bayes's theorem can also further reduce bias and provide a framework for the efficient accumulation and use of evidence [Newman & Evans, 2002]. Via abductive inference, Bayesian networks also offer a means of examining attributable risk or probability of causes, and for combining data sources (e.g. experiential and non-experiential) to yield information that neither study alone can do [Pearl, 2000]. Tools and techniques (including applying weightings) for combining information sources have been trialled and reviewed elsewhere [e.g. Pollino & Hart, 2006; Pollino *et al*., 2007].

At the core of an ecological risk assessment is the assessment of causality [Newman & Evans, 2002]. Bayesian networks can be used to formulise quantification of risk, representing likelihood/hazards as system variables that are linked to consequences/effects (such as environmental values that can act as model endpoints) within a risk assessment framework. As such, Bayesian networks readily enable relationships between multiple stressor/hazards and multiple endpoints to be modeled in a holistic way. The steps taken in building a BN facilitate the ERA process and provide a tangible outcome that can be used to examine and prioritise key risks, including uncertainties. As both processes are iterative, they also fit within an adaptive management context. BNs can also be extended for risk management planning and decision-making, where other non-scientific factors (e.g. cultural, economic, etc.) can be included in models.

Two case studies where Bayesian networks were developed in a risk assessment context are briefly described below. For an in-depth description of the case study methods used for expert elicitation and model parameterisation and evaluation, please refer to our previous papers [Pollino & Hart, 2006, Pollino *et al*., 2007].

2.1 Case Study 1: Goulburn Catchment, Australia

An ecological risk assessment was undertaken in the Goulburn Catchment (Victoria, Australia). Water from the Goulburn River, and its tributaries, have a number of uses, including: irrigation, native and trout fishery, recreational boating and town supply.

In the problem formulation phase, stakeholders identified the native fish abundance and diversity in the Catchment as an asset under threat. A risk assessment was designed to support future decision-making in the catchment, by:

- Quantifying linkages between threats and native fish communities throughout the catchment, considering both site and reach scales;
- Prioirtising risks to native fish communities;
- Communicating uncertainties in predictions and identifying key knowledge gaps.

A conceptual model of the catchment was constructed, linking system variables and endpoints. This was developed in association with catchment managers and ecologists, and formed the structure of the BN. Model endpoints were Future Abundance and Future Diversity of native fish. The model consists of five interacting components, water quality, hydraulic habitat, structural habitat, biological potential and species diversity [Pollino *et al.*, 2007].

A BN was constructed to represent the 23 sites in the Goulburn Catchment, which were aggregated into 7 priority reaches. Prediction can be made per site or reach, or as aggregates. Model parameters were derived in 3 ways: expert only; data only; and a combination of expert and data. Data parameterization used the expectation maximisation (EM) algorithm, where missing data was predominately due to an absence of monitoring data. For details of model parameterisation and evaluation, see [Pollino *et al.*, 2007]. The relationships in the BN had not been characterized previously.

Despite the high uncertainties associated with the lack of knowledge of the relationships between variables and lack of available data, the BN developed has the ability to predict the abundance and diversity of native fish communities based on existing and predicted changes to environmental conditions. Consequently, it also has the capability to assist in determining what management options are most favourable for maintaining and rehabilitating fish communities at multiple spatial scales. However, it is recognised that the model requires further testing in the field to determine its accuracy pre- and post-management interventions or system changes.

The Bayesian network endpoints are directly measurable and results are easily interpreted by stakeholder groups, including system managers and ecologists. Management scenarios or system changes can be investigated to inform decision-making, where top-down and bottom-up reasoning can be applied. As model predictions are likelihoods, they are directly applicable to risk management. Sensitivity analyses allowed prioritisation of threats to environmental value, and key data and knowledge gaps identified [Pollino *et al.*, 2007], allowing areas needing further research to be identified and prioritised.

2.2 Case Study 2: Ok Tedi-Fly River, Papua New Guinea

The mining industry has a long legacy of causing major environmental impacts. Although most mining activities today have improved environmental operations, unacceptable changes to the environment still occur. Increasingly, the mining industry is using risk-based approaches to better predict and manage unacceptable changes to downstream ecosystems.

The Ok Tedi copper and gold mine, operated by Ok Tedi Mining Limited (OTML), is located in the mountains of Western Province of Papua New Guinea. The mine has been operating since 1984 and discharges approximately 164,000 tonnes of waste rock and 82,000 tonnes of tailings per day to the Ok Tedi, with these mine-derived sediments being transported downstream to the Fly River. As a result, channel aggradation, up to 70 m in the upper sections of the catchment, has occurred. Studies investigating the impacts of mine-derived materials in the Ok Tedi-Fly River system have raised concerns that acid rock drainage (ARD) and the associated liberation of increased concentrations of heavy metals, in particular copper, may cause additional environmental damage.

In order to assist decision-making regarding future mine operations, OTML decided to undertake an ecological risk assessment, with the aim:

- To assess the risks to environmental values in the Ok Tedi and Fly River system from ARD; and
- To test the effectiveness in mitigating these risks by removing 85% sulphur from the tailings before they are discharged to the river system.

Assessment endpoints in this study were: drinking water standards; the availability and edibility of fish; the access to forest products and availability and edibility of terrestrial foods; and primary productivity (focussing on algae). Five Bayesian network models were constructed [Pollino & Hart, 2005; Pollino & Hart, accepted].

The model causal structures were based on conceptual models developed by mine managers, ecologists, chemists, geologists and ecologists. BN models represent two major riverine processes, water chemistry and sediment processes. Water chemistry is under the influence of ARD processes (ARD and no ARD scenarios can be tested). Channel aggradation is influenced by sediment transport. The outcomes of sub-models are integrated into endpoint variables. Activities at the mine and their subsequent influence on system processes and model endpoints are the focus of management activities in models. The Bayesian network represented 5 reaches over a number of time intervals.

Two physical models were used to parameterise water quality and channel aggradation processes. The physical models used were OkChem-OkARD [EGi, 2005], a purpose build water quality model and HEC-6 [Pickup & Cui, 2003]. Expert opinion and/or research and monitoring data were used to parameterise ecological endpoints. Sensitivity analysis was conducted on all models and predictive accuracy tests were used, where possible. For details of model parameterisation and evaluation, see [Pollino *et al.*, 2007] and Pollino *et al*. [accepted]. The models were used in risk management processes, where a series of mine operation scenarios, and the resultant ecological outcomes, were investigated. The findings were used to assist in scenario-based risk management and to guide investments in improved mining operations and further research.

3. CONCLUSION

As embodied in many existing regulatory guidelines [e.g. USEPA, 1998; ANZECC, 2000], the ecological risk assessment framework has been advocated as a ‘reductionalist’ approach to management. This type of approach has been highly criticised as not being representative or applicable to ecosystems and the needs of natural resource mangement. To broaden the relevance of risk analysis techniques to ecosystems, a ‘holistic’ approach to risk assessment is now advocated [Assmuth & Hilden, 2008; Hart *et al.*, 2007; Macleod *et al.*, 2007], where multiple stressors and their interactions across an ecosystem are taken into account [Leuven & Poudevigne, 2002]. To achieve this, a nested hierarchical framework is often used to break down complex ecosystem processes into simpler levels of ecological organisation, which can be more easily reassembled using a bottom up process [Leuven & Poudevigne, 2002; Levin, 1999].

This bottom up process for a risk assessment is widely advocated in modern risk assessment frameworks [Hart *et al.*, 2007]. Bayesian networks can also be used to represent and integrate knowledge of processes, or existing models, from other fields (e.g. hydrology, sediment transport and contaminant life cycle models), and consequently, they are not intended to ‘reinvent the wheel’. As models are iterative, this knowledge can be regularly updated, fitting into an adaptive management framework of continual learning. Wider application of Bayesian methods can reduce problems associated with causality assessments, reduce conflicts emerging from less formal integration of available evidence, and more effectively use limited resources for risk assessments, and their use in risk management [Newman & Evans, 2002].

Finally, BNs are also ideal for scenario based risk management, where scenario analysis is a key element required for assessing risk [Hart *et al.*, 2007]. The BNs allowed for greater

analysis of a decision, where uncertainty was a key component of the analysis, and risk management outcomes, where decisions are liable to be open to scrutiny.

REFERENCES

Assmuth, T. & Hilden, M. (2008) The significance of information frameworks in integrated risk assessment and management. *Environmental Science & Policy*, **11**, 71-86.

Borsuk, M.E., Stow, C.A., & Reckhow, K. (2004) A Bayesian network of eutrophication models for synthesis, prediction and uncertainty analysis. *Ecol. Modelling*, **173**, 219-239.

Burgman, M. (2005) *Risks and decisions for conservation and environmental management* Cambridge University Press, Melbourne.

EPA (1998). Guidelines for Ecological Risk Assessment. United States Environmental Protection Agency, Washington D.C.

Hart, B., Pollino, C., White, A., Grace, M., Burgman, M., Fox, D., Carey, J., Chee, Y.E., Henderson, B., & Bui, E. (2007). Risk-based approaches for managing contaminants in catchments. In *Salt, Nutrient, Sediment and Interactions: Findings from the National River Contaminants Program* (eds S. Lovett, P. Price & B. Edgar), pp. 127 - 140. Land and Water Australia, Canberra.

Henderson, C., Pollino, C.A. and Hart, B.T. (2008). Bayesian modelling as a basis for risk-based Environmental Flow assessments: Literature Review. Report by iCAM, ANU and Water Science Pty Ltd for the National Water Commission, Canberra

Hubbard, D. (2007) *How to Measure Anything: Finding the Value of Intangibles in Business.* John Wiley and Sons, New York.

Leuven, R.S. & Poudevigne, I. (2002) Riverine landscape dynamics and ecological risk assessment. *Freshwater Biol.*, **47**, 845-865.

Levin, S.A. (1999) Towards a science of ecological management. *Conservation Ecology*, **3**, 6-9.

Macleod, C.J.A., Scholefield, D., & Haygarth, P.M. (2007) Integration for sustainable catchment management. *Science of the Total Environment*, **373**, 591-602.

Newman, M.C. & Evans, D.A. (2002). Enhancing Belief during Causality Assessments: Cognitive Idols or Bayes's Theorem. In *Overview of Ecological Risk Assessment in Coastal and Estuarine Environments* (eds J. M.H. Roberts, M.C. Newman & R.C. Hale), pp. 73 - 96. Lewis Publishers, New York, New York, USA.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference.* Cambridge University Press, Cambridge, UK.

Pollino, C.A. & Hart, B.T. (2005). Assessment of the risks to the environment and human health from mine-derived heavy metals in the Ok Tedi and Fly Rivers, PNG. Unpublished report by Water Science Pty. Ltd. and Water Studies Centre, Monash University for Ok Tedi Mining Limited dated August 2005.

Pollino, C.A. & Hart, B.T. (2006) Bayesian decision networks - going beyond expert elicitation for parameterisation and evaluation of ecological endpoints. In Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modelling and Software Society (eds A. Voinov, A. Jakeman & A. Rizzoli), Burlington, USA, July 2006.

Pollino, C.A. & Hart, B.T. (accepted) Assessing the ecological risks of a mine operation in Papua New Guinea: Part 1. Background and Methodology. *Risk Analysis*.

Pollino, C.A., Woodberry, O., Nicholson, A.E., Korb, K.B., & Hart, B.T. (2007) Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software*, **22**, 1140-1152.

Stow, C.A. & Borsuk, M.E. (2003) Enhancing causal assessment of fishkills using graphical models. *Ecosystems*, **6**, 11-19.

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Spatially Distributed Identification of Debris Flow Source Areas by Credal Networks

A. Salvetti [a], **A. Antonucci**[b], **and M. Zaffalon**[b]

[a]*Politecnico di Milano, Dipartimento di Elettronica e Informazione,*
Via Ponzio 34/5, 20123 Milano, Italy.
Dipartimento del Territorio, Cantone Ticino,
Viale Franscini 17, 6500 Bellinzona, Switzerland.
salvetti@elet.polimit.it, andrea.salvetti@ti.ch

[b]*Istituto Dalle Molle di Studi sull'Intelligenza Artificiale,*
Galleria 2 CH-6928, Manno-Lugano, Switzerland.
{alessandro,zaffalon}@idsia.ch

Abstract: Debris flows represent a very destructive natural hazard, affecting buildings, transport infrastructures, and, very often, causing human losses in mountain regions. That makes the identification of potential source areas of debris flows inside a watershed particularly important. In this paper we present a general identification procedure based on the *credal network* (that is an imprecise probabilistic graphical model generalizing Bayesian networks) originally introduced by Antonucci et al. [2004]. That model is significantly improved by a more refined description of the meteorological and hydrological processes contributing to the debris flow initiation. As a counterpart of such improvement, the model pays a slight increase in terms of computational time for identifications. That does not prevent its extensive, spatially distributed, application to whole basins, thanks to a preliminary deterministic analysis that rejects local areas where the triggering of a debris flow cannot take place. The overall procedure is tested for a debris flow prone watershed in Southern Switzerland. The model detects the areas in the basin more prone to debris flow initiation and also shows that different rainfall return periods produce different patterns of hazard in the basin. That makes it possible with this procedure to determine the return period of the critical rainfall that triggers debris flow as a result of channel-bed failure in a specific point along the drainage network.

Keywords: Debris Flow; Geomorphologic Theory; Geographic Information System; Imprecise Probabilities; Credal Networks.

1 INTRODUCTION

Debris flows (Section 2) represent a very destructive natural hazard, affecting buildings, transport infrastructures, and, very often, causing human losses in mountain regions. As recently pointed out by Berti and Simoni [2005], the triggering mechanisms and the causal relationships for the whole process are still partially unknown. Thus, human expertise together with an analysis of historical data are still necessary to support any deterministic model for the identification of potential source areas of debris flow.

A *credal network* (which is an imprecise probabilistic graphical model as described in Section 3) has been introduced by Antonucci et al. [2004] in order to fuse into a single coherent framework the model of Takahashi [1991] with expert qualitative judgments and historical data. In this paper we first improve this probabilistic model with a more refined description of the *drainage network*

delineation and the *maximum peak runoff* formation (see Section 4). We also allow for more freedom in the description of the observation of the triggering factors by showing how sets of probability mass functions can model the so-called *soft evidence* provided by vague observations.

Overall, such improved flexibility in the modeling phase has a counterpart when we compute the level of risk corresponding to the available evidence by appropriate updating algorithm. The required computational time slightly increases. Nevertheless, we show that our approach can be properly employed for spatially distributed identifications on extensive areas. This is made faster by the rejection of the points of the basin where the geomorphological conditions prevent any debris flow initiation. Finally, with the support of a detailed GIS analysis, we test this procedure for a debris flow prone watershed in Southern Switzerland (Section 5). The results indicate that the model detects the areas of the basin more prone to debris flow initiation and produces different hazard patters according to different rainfall events.

2 DEBRIS FLOWS

Debris flows are gravity-induced mass movement intermediate between landslides and water floods. They are composed of a mixture of water and sediment with a characteristic mechanical behavior varying with water and soil content. According to Costa [1984], prerequisite conditions for most debris flows include an abundant source of unconsolidated fine-grained rock and soil debris, steep slopes, a large but intermittent source of moisture (rainfall or snowmelt), and sparse vegetation. As mentioned in Griffiths et al. [2004], several investigation have focused on debris flows initiation and frequency. Benda and Dunne [1997] approached the modeling of spatial and temporal variability of sediment yields, Glade [2005] focused on existing links between debris-flow hazard and geomorphology. Several hypotheses have been formulated to explain mobilization of debris flows and this aspect still represents a research field. The triggering mechanism of the identification procedure presented in this paper is based on the theoretical model proposed by Takahashi [1991], although a more sophisticated explanation of the triggering of debris flow by channel-bed failure has been recently proposed by Armanini and Gregoretti [2005], which consider the exposure of a single particle to the stream flow and, explicitly, the flow velocity profile. For the purposes of this study the Takahashi's theory is regarded as appropriate and this triggering theory is further coupled with geological, hydro-meteorological and topographic factors, which all contribute to the definition of channel-bed failure. Unfortunately, not all the triggering factors considered by this model can be directly observed in the field. Takahashi's theory will therefore offer the deterministic skeleton for our model that will be integrated with probabilistic knowledge according to the methods described in the following section.

3 CREDAL SETS AND CREDAL NETWORKS FOR UNCERTAIN REASONING

If a complete deterministic model of the relations between some variables cannot be provided, probabilistic approaches should be considered instead. Probabilistic graphical models like Bayesian nets seem to be particularly suited for situations where some conditional independence relations hold between the different variables. Here, we consider *credal nets* (Cozman [2000]), which are a generalization of Bayesian nets based on the fundamental notion of *credal set*.

Given a categorical random variable X,[1] we denote by $\mathcal{X}$ the set of the possible values of X, while x denotes a generic element of $\mathcal{X}$. The notation $(X = x)$ denotes an event that is true if and only we know that X is in the state x. This is clearly the most informative scenario we can consider for X. Otherwise, our knowledge about the actual state of X can be modeled by a probability mass function $P(X)$. There are also situations where a single *precise* numerical value for the probability $P(X = x)$ cannot be easily assessed. In these cases, a more realistic model of our knowledge about X could be a *credal set* $K(X)$, i.e., a closed convex set of probability mass functions over X. As noted by Reichert [1997], such *imprecise probability* (Walley [1991]) approach offers important advantages for environmental studies, where a vague prior knowledge

[1] All the quantities considered in this paper are regarded as random variables which assume only finitely many values. Therefore, in the case of continuous variables, a discretization should be preliminary done.

of processes is very usual, due to incomplete observation of the process or to the impossibility of gathering enough data. Furthermore, expert knowledge and qualitative beliefs can be more easily considered in the general framework of sets of probability distributions.

A credal set $K(X)$ can be specified by a set of *probability intervals* $I_X := \{I_x : I_x := [l_x, u_x], 0 \leq l_x \leq u_x \leq 1, x \in \mathcal{X}\}$, from which $K(X)$ is clearly obtained from $\{P(X) : P(x) \in I_x, x \in \mathcal{X}, \sum_{x \in \mathcal{X}} P(x) = 1\}$. Accordingly, we can ask experts to assess sets of probability intervals (and more generally linear constraints on probabilities). Probability intervals can be also inferred from data by the *imprecise Dirichlet model* according to Walley [1996], a generalization of Bayesian learning from multinomial data based on an imprecise probability model of prior ignorance.

Inference over a credal set is intended as the computation of lower and upper expectations over all the mass functions of a credal set. For example the lower probability is defined as $\underline{P}(x) := \min_{P(X) \in K(X)} P(x)$, and similarly for the upper probability $\overline{P}(x)$. Remarkably, a set of mass functions, its convex hull and its *extreme mass functions* produce the same lower and upper expectations and probabilities. Conditioning with credal sets is done by element-wise application of Bayes' rule. The posterior credal set is the union of all posterior mass functions. Denote by $K(X|y)$ the set of mass functions $P(X|Y = y)$, for generic variables X and Y.

A *credal network* is a graphical model where each node (or variable) of a directed acyclic graph is associated with a credal set for any configuration of the variable's parents (see Cozman [2000]); informally, credal nets are equivalent to sets of Bayesian networks with the same graph. The graph codes strong dependencies by the so-called *strong Markov condition*: every variable is strongly independent of its non-descendant non-parents given its parents. A generic variable, or node of the graph, X_i holds the collection of conditional credal sets $K(X_i|\pi_{X_i})$, one for each possible joint state π_{X_i} of its parents Π_{X_i}. We assume that the credal sets of the net are *separately specified* Walley [1991]: this implies that selecting a mass function from a credal set does not influence the possible choices in others. A credal network defines a joint credal set $K(\mathbf{X})$, which is called the *strong extension* of a credal network. This is the convex hull of the set of joint mass functions $P(\boldsymbol{X}) = P(X_1, \ldots, X_t)$, over the t variables of the net, that factorize according to:

$$P(x_1, \ldots, x_t) = \prod_{i=1}^{t} P(x_i \,|pa\,(X_i)\,) \quad \forall (x_1, \ldots, x_t) \in \times_{i=1}^{t} \mathcal{X}_i \tag{1}$$

Here $pa\,(X_i)$ is the assignment to the parents of X_i consistent with $(x_1, \ldots, x_t)$; and the conditional mass functions $P(X_i \,|pa\,(X_i)\,)$ are chosen in all the possible ways from the respective credal sets.

3.1 Updating Credal Networks

Credal networks are often used as *expert systems*. The available evidential information x_E about the variables X_E that have been observed is first gathered. Then, we compute the posterior lower (and upper) probabilities for a queried variable X_q, say $\underline{P}(X_q|x_E)$ (and similarly for the upper) with respect to the network strong extension. The evidence $X_E = x_E$ is said to be *strong* as it provides completely informative information about the state of their variables. Updating of credal networks with strong evidence is a hard task, but a number of approximate (and even exact in some special cases) algorithms can be employed.

Moreover, it is possible to consider situations where the result of the observation of a variable is vague, and we cannot obtain a strong evidence about the state of the observed variable. Nevertheless, such *soft evidence* about the state of the variable could be expressed by a credal set. This is a generalization of Jeffrey's updating to imprecise probability. This soft evidence can be easily embedded in the structure of a credal network if the observed variable is a root node: it suffices to replace the unconditional credal set for the variable with that corresponding to the observation. This approach will be applied to the credal network for debris flow evaluation presented in this paper in order to model the vague observation of *Granulometry*.

4 SPATIALLY DISTRIBUTED IDENTIFICATION OF SOURCE AREAS

Antonucci et al. [2004] proposed a credal network for a single-point analysis of debris flow initiation. We present a significantly improved version of that model obtained with a refined description of the meteorological and hydrological processes contributing to the debris flow initiation. The new credal network is based on the directed acyclic graph in Fig. 1 and expresses the causal relationships between the topographic and geological characteristics, and hydrological preconditions, which are recognized as triggering factors. The key nodes (denoted as shadowed nodes in Fig. 1) of the network are the *Effective Soil Water Capacity*, which reflects the influences of the soil and the geological characteristics of the area, the *Basin Response Function*, related to the topographic properties of the watershed and whose footprint can be detected in the hydrologic response of the basin, the *Peak Flow*, which summarizes the interactions among the hydro-meteorological factors and topographic and geologic preconditions, the *Theoretical Debris Thickness*, obtained by the Takahashi's theory, and the *Available Debris Thickness*, which considers the influence of the topography on the debris availability. The leaf node *Movable Debris Thickness*, which is defined as the depth of debris likely to be transported downstream during a flood event, is our proxy for the risk level in the specific point along the drainage network where we have collected evidence about the triggering factors. The ranges 0–$10\mathrm{cm}$, 10–$30\mathrm{cm}$, $>30\mathrm{cm}$ for that thickness are assumed to indicate respectively a low, medium and high level of risk. In this section only topics related to the novel improvements of the credal network regarding the *Stream Power Index* and the *Maximum Peak Runoff* are described in detail, while for a more comprehensive description of each node the reader is referred to Antonucci et al. [2004].

Stability of a Debris Cluster The effect of a water depth on the movable debris quantity is based on the equilibrium of forces acting on a debris cluster under different conditions. According to Takahashi [1991], the local slope θ for which debris-flow formation can take place obeys the following constraint:

$$\frac{c^*\Delta}{\frac{4}{3}+c^*\Delta}\tan\phi \leq \tan\theta \leq \frac{c^*\Delta}{1+c^*\Delta}\tan\phi, \tag{2}$$

where ϕ is the *Friction Angle* corresponding to the actual level of *Granulometry*, c^* is the volume concentration of the particles, and Δ is the ratio between the mixture and the water density. For the points of the basin whose values of ϕ and θ do not satisfy the constraint in Eq. (2), either the cluster is not completely saturated and, if unstable at high slope angles, produces a landslide or the process that takes place is the ordinary solid transport (Dietrich et al. [2006]), and therefore we drop the relative point from the potential source areas of this hazard without any further analysis.

Drainage Network Delineation and Debris Availability Many authors have dealt with the capability of *Local Slope* and *Upstream Contributing Area* to account for topographic control on erosion and deposition potential in complex terrain and with the use of slope and contributing area for channel network extraction, based on critical area and slope-area threshold (e.g., Prosser and Abernethy [1996]). In this study, such a method was used to extract the channelized portion of the *Digital Elevation Model* (DEM), where debris flow initiation can appear, according to the following equation:

$$SPI = \sqrt{A}\cdot\theta, \tag{3}$$

where SPI denotes the *Stream Power Index* and A is the upstream area. The threshold value of SPI has been identified by trials, comparing the extracted network with the drainage network on the map, where also many ephemeral channel in the upper part of the basin were included in the network. That index can be used as an indicator of the local transport capacity of a single reach along the network and, therefore, to identify channel reaches were debris material preferentially accumulates (Dalla Fontana and Marchi [2003]). Clearly, the availability of an abundant debris thickness in the drainage network is a fundamental precondition for debris initiation and we therefore developed a conceptual framework for a qualitative evaluation of the debris availability in the river bed. We assume that the debris availability is a function of the convenience capacity of the

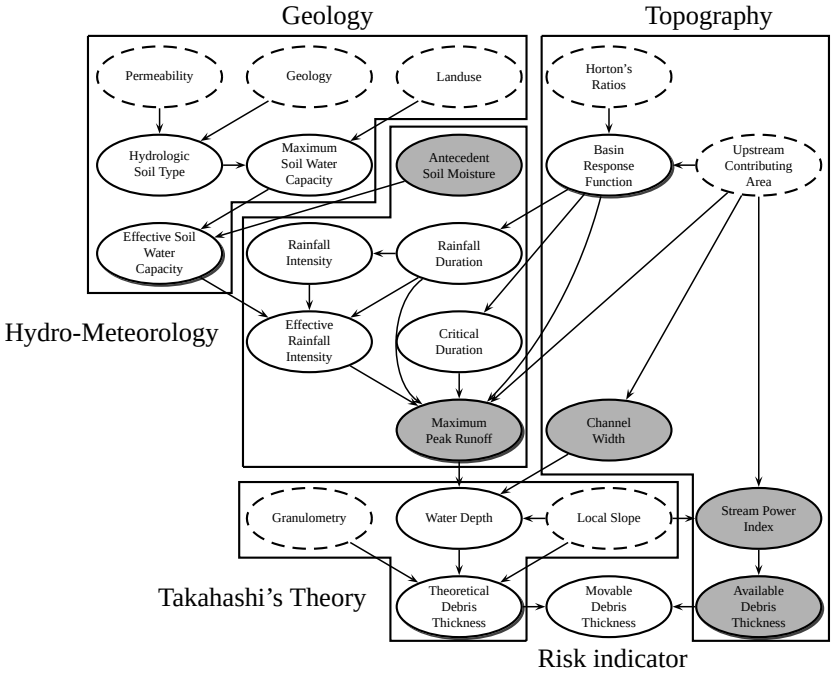


Figure 1: The directed acyclic graph of the credal network for hazard identification. The gray nodes denote the new factors or those for which a new quantification has been proposed with respect to the original network in Antonucci et al. [2004]. A dashed border denotes the nodes observed in the case study of Section 5.

river network associated to the SPI. Cells with SPI value exceeding the threshold for channel initiation correspond very often to areas where bedrocks emerge and local slope is quite high, and therefore the sediment deposition is zero or very low. On the contrary, in cells where SPI is much less than the selected threshold level, high deposition instead of erosion is expected and we therefore assume a high availability of debris material. These principles supported by expert knowledge have been used for an interval-valued probabilistic quantification of the node *Available Debris Thickness*.

Maximum Peak Runoff According to the model of the initiation mechanism considered in this study, the soil failure is induced by surface runoff and, consequently, the maximum discharge and the corresponding water depth must be estimated. Rodriguez-Iturbe and Rinaldo [1997] investigated how the variation of the characteristics of stream channel is expressed as a function of the discharge by a power law at a given cross section and also along the channel network. The parameters were estimated by using a few collected cross-section data, randomly distributed along the drainage network.

The maximum runoff along the drainage network was calculated according to the well-established theory of the *Instantaneous Unit Hydrograph* (IUH), expressed as the convolution integral of the effective rainfall input. The hydrograph shape strongly depends on the geomorphological features of the river basin, therefore, the *Geomorphologic Instantaneous Unit Hydrograph* (GIUH, Rodriguez-Iturbe and Valdes [1979]) presents obvious advantages in ungauged watersheds, since the GIUH only depends on the morphological characteristics of the watershed and the drainage network. According to this theory, the maximum peak runoff Q is obtained using the following:

$$Q = \begin{cases} I^{'} A \left[H\left(t^{*}\right) - H\left(t^{*} - t_{p}\right) \right] & 0 \leq t_{p} \leq \tau_{c} \\ \\ I^{'} A & t_{p} > \tau_{c}, \end{cases} \tag{4}$$

where I' is the effective rainfall intensity, $H(t)$ represents the integral of the GIUH from the beginning of the storm, t^* is the critical duration at the considered point, and it is a function of the

rainfall duration t_p, while τ_c is the concentration time. Effective rainfall intensity is determined using the well established *SCS Curve Number* infiltration method and the rainfall intensity modeled by multiscaling power law relationship. The critical duration t^* associated with the extreme peak runoff is independent of the return period and of the rainfall intensity; the corresponding rainfall volume is calculated for the rainfall duration t_p.

5 A CASE STUDY: DISTRIBUTED APPLICATION TO A RIVER BASIN

The case study we present in this section refers to the Acquarossa Creek in the Blenio Valley, an area located in the North-Eastern part of the Ticino Canton, Southern Switzerland. This area was selected because of the potential hazard caused by debris flows to communication lines and villages. That creek is a small tributary of the Brenno river, characterized by a high altitude range (from $530\mathrm{m}$ up to $2580\mathrm{m}$ a.s.l.) of the Simano Peak. Debris torrents are usually triggered by intense rainfall, following a period of abundant precipitation. Eight historical debris flow events were recorded in that area during the last 150 years. Most of them caused high damages to infrastructures on the alluvial fan, transporting several thousand cubic meters of material. For instance, during the last event in August 2003, a volume of about $15'000\mathrm{m}^3$ were estimated on the alluvial fan, and a similar pattern was observed in 1983 and 1987. That represents a relatively high frequency of debris flow events. Accordingly, the triggering factors appear to be already effective in many parts of the basin with storm events of low and medium return period.

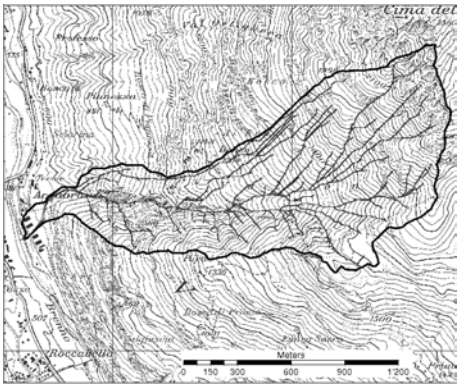


Figure 2: Acquarossa Creek Basin (area $1.6\mathrm{Km}^2$, length $3.1\mathrm{Km}$).

In order to gather evidential information about the geomorphological characteristics of the basin, a highly precise DEM based on airborne laser scanning produced by the Swiss Federal Office of Topography has been employed. That offers a spatial resolution of 4 meters, which is comparable with the typical channel width; that defines a drainage network of 6310 cells. Most of the morphological data used for our identification analysis (slope, flow-direction and flow-accumulation) were derived from this dataset, and the SPI was calculated as in Eq. (3).

Finally, regarding the observation of the granulometry, a field survey was conducted. The river bed and lateral debris levees were analyzed in order to determine the grain-size distribution of the debris material. A significant difference was observed for the grain-size distributions obtained from several samples. We have therefore decided to split the basin into two sub-regions of "uniform" granulometry, and describe the outcome of the sampling by a *soft evidence* modeled as a new unconditional credal sets for the corresponding node in the credal network, according to the procedure described in Section 3.1.

In order to avoid unnecessary computations, for each point of the basin, we have preliminarily checked whether or not the observed slope and the values of the friction angle compatible with

the soft observation of the granulometry were compatible with the constraint in Eq. (2). This deterministic pre-analysis detects 170 pixels where only ordinary sediment transport is possible and 135 pixels that are already unstable without complete soil saturation. For the remaining 6005 pixels, we have computed the posterior lower and upper probabilities for the movable debris thickness corresponding to observed geomorphological factors and rainfall intensity for a return period of 10, 30 and 100 years. These computations have been exactly performed by exhaustive approaches based on the iteration of standard algorithms for Bayesian networks as our credal network is equivalent to about 500 Bayesian networks. The network is thus expected to predict the probability of a debris flow event with the defined frequency level at each point of the drainage network. In this way, we aim at verifying whether the network would have been a valuable tool to predict considerable events of debris flows, which actually happened in the areas under consideration, and, more important, to identify the points where the debris flow is most likely to occur in the future. Figure 3 reports the results of the inference process for respectively 10 and 100 years return period rainfall event.

We observe that, according to the outputs of the credal net, debris flows are more likely to initiate on the main channel, even in the lower part of the basin. In fact, during a field survey conducted immediately after the debris flow event on August 2003, we observed typical evidences of bed erosion and channel-bed failure in the lower part of the main channel, up to an altitude of 700 m a.s.l; this observation is effectively confirmed by our results. Regarding the role of the return period in our tests, we observe an increase of the number of dangerous points, that spread upstream along the drainage network: for higher return periods even a small upstream area is sufficient to produce a peak runoff that can trigger a debris flow. The promising results of the credal net will be further compared and quantitatively evaluated by applying the model to other watersheds, where detailed geomorphological maps are obtainable and field observations of availability and characteristics of debris material along the drainage network have been recently collected.



Figure 3: Spatially distributed identifications for the basin in Figure 2 and rainfall return periods of 10 (left) and 100 (right) years. Points for which the credal net predicts the lower class of risk are depicted in gray, while black refers to points where higher levels of risk cannot be excluded.

6 CONCLUSIONS AND OUTLOOKS

We have presented a model for automatic identification of potential source areas of debris flows based on *credal networks*. Our network provides a refined description of the meteorological and hydrological processes contributing to the debris flow initiation, and allows to model also *vague* observations of the triggering factors. The identification procedure can be extensively applied to whole basins, and unnecessary computations are avoided for areas where the geomorphological conditions are not compatible with debris flow initiation. As a spatially distributed case study, we tested our model for a debris flow prone watershed in Southern Switzerland and the obtained results agree with field observations collected after the last debris flow event of 2003. The model is able to detect the areas inside the basin more prone to debris flow initiation and also shows that different rainfall return periods produce different hazard patterns. That makes it possible to

determine the return period of the critical rainfall that triggers debris flow as a result of channel-bed failure in a specific point along the drainage network. As a possible development of the present work, we intend to design a post-processing procedure for our simulations that produces integral risk indicators based on neighborhood relations among the detected dangerous points.

ACKNOWLEDGMENTS

This work was partially supported by the Swiss NSF grants 200021-113820/1 and 200020-116674/1, and by the Hasler Foundation grant 2233.

REFERENCES

Antonucci, A., A. Salvetti, and M. Zaffalon. Hazard assessment of debris flows by credal network. In *iEMSs 2004 International Congress: Complexity and Integrated Resources Management*, Osnabrück, Germany, 2004. International Environmental Modelling and Software Society.

Armanini, A. and C. Gregoretti. Incipient sediment motion at high slopes in uniform flow condition. *Water Resources Research*, pages W12431, doi:10.1029/2005WR004001, 2005.

Benda, L. and T. Dunne. Stochastic forcing of sediment routing and storage in channel networks. *Water Resources Research*, 33(12):2865–2880, 1997.

Berti, M. and A. Simoni. Experimental evidences and numerical modelling of debris flow initiated by channel runoff. *Landslide*, 2:171–182, 2005.

Costa, J. E. Physical geomorphology of debris flows. In *Costa, J. E. and Fleisher, P. J. Developments and Application of Geomorphology*, Berlin, Germany, 1984. Springer Verlag.

Cozman, F. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

Dalla Fontana, G. and L. Marchi. Slope-area relationships and sediment dynamics in two alpine streams. *Hydrological Processes*, 17:73–87, 2003.

Dietrich, W., D. Bellugi, and R. Asua. Validation of the shallow landslide model, shalstab, for forest management. In *Land use and watersheds: human influence on hydrology and geomorphology in urban and forest areas edited by M.S. Wigmosta and S.J. Burges*, Washington D.C., 2006. AGU.

Glade, T. Linking debris-flow hazard assessments with geomorphology. *Geomorphology*, 66: 189–213, 2005.

Griffiths, P., R. Webb, and T. Melis. Frequency and initiation of debris flows in grand canyon, arizona. *Journal of Geophysical Research*, 109:4002–4015, 2004.

Prosser, I. and B. Abernethy. Predicting the topographic limits of a gully network using a digital terrain model and process threshold. *Water Resources Research*, 32(12):2289–2298, 1996.

Reichert, P. On the necessity of using imprecise probabilities for modelling environmental systems. *Water Science and Technology*, 36(5):149–156, 1997.

Rodriguez-Iturbe, I. and A. Rinaldo. *Fractal river basins - Chance and Self Organization*. Cambridge University Press, Cambridge, UK, 1997.

Rodriguez-Iturbe, I. and J. Valdes. The geomorphological structure of hydrogeological response. *Water Resources Research*, 15:1409–1420, 1979.

Takahashi, T. *Debris Flow*. A.A. Balkema IAHR Monograph, Rotterdam, 1991.

Walley, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

Walley, P. Inferences from multinomial data: learning about a bag of marbles. *J. R. Statist. Soc. B*, 58(1):3–57, 1996.

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Learning conditional probabilities in event bushes with temporal labels

S. I. Nikolenko [a], **C. A. Pshenichny** [b], **R. Carniel**[c]

[a] *Steklov Mathematical Institute, St. Petersburg, Russia (sergey@logic.pdmi.ras.ru)*

[b] *VNII "Okeangeologia", St. Petersburg, Russia (pshenichny@yandex.ru)*

[c] *University of Udine, Udine, Italy (roberto.carniel@uniud.it)*
Departamento de Geofisica, UNAM / UNiversidad Autonoma de Mexico, D.F. Mexico

Abstract: The event bush is a recently developed formalism for knowledge representation optimized specifically for Earth science knowledge. It graphically represents logical and/or probabilistic dependencies between its nodes (events). Numerical data in geosciences often come in the form of time series: therefore, it is a very common situation to have tables with time-related numerical data for the variables which correspond to the events represented in an event bush. In this work, we present a way to use these numerical data to learn conditional probabilities that have to be specified in the intermediate event bush created on the basis of an event bush with temporal labels. The same idea works also if the bush also has spatial labels.

Keywords: event bush; time series; probabilistic inference; risk assessment

1 INTRODUCTION

The event bush is a newly developed formalism for knowledge representation optimized specifically for Earth science knowledge. It was first proposed by Pshenichny [2003] as a new approach to probabilistic knowledge representation. Event bushes underlie Bayesian Belief Networks (BBN) and perform probabilistic inference via reducing to a BBN, see Pshenichny et al. [2008] for more details.

Latest research has shown that the event bush formalism is extremely viable; we have developed ways to incorporate temporal and spatial data immediately into an event bush and have begun developing learning algorithms for event bushes based on time series data. Thus the event bush becomes an information technology linking time series and geodata, or, to put it more explicitly, bringing time series data into the geoinformation system format. The novelty of this approach is semantic and conceptual strictness, which is a virtue of event bushes. In the present paper we continue this work towards automatic learning of probabilities in event bushes.

For instance, a volcanic environment in mountainous area in a temperate climatic zone implies ice caps topping mountains including active volcanoes, deep valleys hosting rivers, and forested mountain slopes. Even slight warming under the volcano, or a regional earthquake may trigger ice avalanches turning into lahars downslope, which become yet more devastating as they capture tree trunks. Lahars may pond rivers and cause floods, which, in turn, evoke other dangerous consequences. Such processes are successfully modeled by event bush, and all of them can be attributed spatial or temporal values, or both.

Let us assume that the whole mapped area is divided into finite number of regions (e.g., by principles of geomorphology), so that each node of event bush is attributed a subset of regions on a map.

Inversely, for any region an event bush may be plotted that represents either a full, or, commonly, a reduced version of the general bush.

Importantly, (i) a process that originates in one region (e.g., heat flow in the summit area of the volcano) may need to last long enough to cause another process in the same place (ice melting). Otherwise, (ii) a lahar should travel long enough from its source area to capture tree trunks. Finally, (iii) a flood should spread well upstream the gullies and last for quite a time to cause slumping of gully sides. Obviously, we have "shifts" of cause-effect relationship (i) in time, (ii) in space, and (iii) both. These can be assessed from appropriate physical models or from observations, if only feasible. In the latter case we have time series or geostatistical datasets. Both can, and, ideally, should, be used for learning conditional probabilities. For the sake of simplicity, we consider learning based on one-dimensional data, i.e., the time series, but take into account that these data are regionalized.

Let us describe how we proceed from the initial structure of event bush through a time-and-space-labeled bush to a BBN. First, our algorithm creates copies of the event bush structure for each region. Then, we need to estimate whether an event X in region R influences event Y in region Q. By a simple learning algorithm, we can find the time delay (or a distribution over a set of possible delays) between X and Y in each region. Thus, we can complete the construction of an intermediate event bush (ready to convert into a BBN) inside each region.

In this paper, we concentrate on the issues of temporal modeling. In Section 2, we review the basics of event bush theory, namely the method to convert event bushes to BBNs and the basics of temporal data processing in event bushes. Section 2 should not be confused with a thorough introduction of the event bush formalism. Such an introduction may be found in Pshenichny et al. [2008] and even right here, under the same cover, in Pshenichny et al. [2007a]. We also do not deal with the question of how to build the structure of an event bush or a Bayesian Belief network. This can be done by either expert elicitation or machine learning, see e.g. Jensen [2002], Neapolitan [2004], Tulupyev et al. [2006].

Instead, in the present paper we dive into the details of a question on using time series data for the event bush formalism. Section 3 contains the new results introduced in this paper: how to use time series data (often available in geosciences) for learning probabilities. Section 4 describes a particular example of applying our methods based on the data recorded at the Stromboli volcano. Finally, Section 5 concludes the paper with a discussion of further work possibilities.

2 MATHEMATICAL ASPECTS OF EVENT BUSHES

2.1 Formal aspects of event bushes

Formally speaking, an event bush is a directed acyclic multigraph $B = \{V, E, S, M\}$, where

- V is the set of events (nodes);
- $E \subseteq 2^V$ is the set of edges;
- S is the set of operations (in the current case, logical connectives $\wedge$ and $\vee$);
- $M : E \to S$ is the set of labels on edges (in our case, they represent probabilities).

In the current event bush architecture there are two possible types of multiedges: $\vee$–edges with one source node and any number of target nodes and $\wedge$–edges with two source nodes and two target nodes. The semantics for an $\vee$–edge is "if source then one of the targets"; the semantics for an $\wedge$–edge is "if $\mathrm{src}_1 \wedge \mathrm{src}_2$ then tgt_1; if $\mathrm{src}_1 \wedge \neg\mathrm{src}_2$ then tgt_2". The semantics and other logical aspects of this construction, including its detailed motivation, can be found in Pshenichny et al. [2008].

As shown above, an event bush can have labels on edges. In this case, we consider probability labels that define conditional probabilities of the target nodes conditional on the source nodes. This gives rise to a general task of Bayesian inference on an event bush. This task (which is currently the primary numerical task solved by the event bush formalism) is solved by reducing an event bush to a BBN which is equivalent to the event bush, see Pshenichny et al. [2008] for more details.

The next developments which are absolutely crucial to make event bushes a success for geosciences would be adding time and space to the bushes. Once we have done that, it will be straightforward to incorporate the event bush technology into existing geoinformatic systems.

2.2 Temporal data in event bushes

It often happens that the logical structure of the event flow is not complete unless time is taken into account. Geological data are commonly related to time: to know *when* is no less important than to know *what*.

As shown above, the event bush is a qualitative tool of organization of a domain of knowledge. Together with the BBN formalism it forms a powerful tool for geohazard assessment. Thus, it it absolutely necessary to incorporate time into the structure of an event bush. BBNs have several known approaches to time handling. These can be divided into two main classes: discrete time (where it usually comes down to copying the nodes of the network several times corresponding to the time slices) and continuous time. For the time being, we do not touch continuous time handling in event bushes; but how do we account for discrete time?

This problem has been already addressed by Pshenichny and Nikolenko [2007]; Pshenichny et al. [2007b]. In this section we remind our progress and continue from there. We describe here the most basic approach to handling time-related information in an event bush and transforming such event bushes into BBNs and back. We take the most common (and sufficient for most practical purposes) approach of discrete time handling. That is, we divide the time scale into discrete periods with lengths known in advance. Each primary or secondary event in an event bush may have a time interval when it occurs (e.g. "magma ascent takes place from day 2 to day 5", where time is measured from some arbitrary point when the modeled processes begin, see also Jaquet and Carniel [2003]). If it is not specified, we take that the event occurs throughout the whole time scale (e.g. it is a landscape peculiarity).

Apart from the time interval, there are two other questions that should be answered about each of the nodes of the event bush in order to incorporate time properly. First, how fast does this event influence its successors? E.g. if magma ascent leads to lava doming, does it occur immediately, in an hour, or in two hours? There is room for intermediate options here: magma ascent may lead to lava dome in one hour with probability 0.4 and in two hours with probability 0.6. This can also be easily accounted for in the corresponding BBN (in the next section, we will describe this construction in detail and show how one can learn the delays from data given as time series).

Second, for some events it may be impossible to say when they end (we detect that magma ascent has begun, but it has not yet ended). In these cases it should be noted how an event occurring now influences the probability of the same event occurring in the next time interval. E.g. if magma is ascending, how probable is it that it will be ascending in the next hour? Usually answers to these questions depend also on the duration of an event; this can also be taken into consideration.

The simplest approach to both these problems would be to ask the user for missing data. If a user can input the duration of an event if it occurs or the time of its beginning and how probable it is for it to continue for different durations, this is all we need to completely specify the BBN structure. In the next section, we describe a more automated approach that can learn cause-effect delays from data.

The event bush "loaded" with time values is translated then into a BBN (where the actual propagation will take place). We are given the value of a discrete time step (e.g. an hour) and information described above about the nodes of the event bush. In order to build a time-specific BBN, we first build a regular BBN on this event bush, discarding for the moment all time-related information (note that we also discard cause-and-effect relationships that do not take place immediately). Afterwards, we copy this BBN n times, where n is the number of time periods under consideration (if it is undefined, we may take the maximum of all possible durations or just set a reasonable upper bound for our modeling).

In the resulting BBN, we keep the structure of the initial BBN but add edges corresponding to time-related relationships. Specifically, we add directed edges between time-related causes and effects. Of course, the edges should be directed from the earlier event to the later. The additional edges may connect similar events with different time ("if magma ascends at time x, it will with probability 80% ascend at time $x + 1$") or different time-separated events ("if magma ascends at time x, with probability 80% there will be a lava dome or a lava flow at time $x + 1$"). The BBN in which actual computation takes place is transformed back into an event bush to display results to the user.

Time-specific event bushes will allow to adequately capture monitoring data, fully incorporate abundant physical models of eruptive scenarios, seismic unrest, flow propagation, and others, and will be tested on a number of various hazardous objects (volcanoes first of all), for which good time-series exist.

3 Learning temporal information from time series data

3.1 Time series as input

In the previous sections we came to the conclusion that implementing temporal information in expert systems requires some additional data. In particular, apart from the logical structure of the event bush, one is required to specify how events represented in the bush relate to each other with respect to time (i.e. how long is the delay between cause and effect). In this section we offer a new technique that will allow to answer this question automatically.

It is common for geosciences to have input data represented as time series. Consider Fig. 1. On the left we have a series of observations of the value of X, while on the right observations of Y are depicted. Suppose that in our event bush logical structure there is an arc from X to Y (meaning that X affects Y). By looking at these graphs, it is fairly obvious that Y indeed happens after X, and a sudden increase in X leads to an increase in Y after about three units of time. Thus, we can enter the delay quantity as 3, and the intermediate event bush will have arcs from "X at time t" to "Y at time $t + 3$".

Naturally, if we have time series as data, we can try to extract the delays automatically. We distinguish two major approaches to this, both based on bayesian estimation, but leading to different structures of the intermediate bush graphs.

3.2 Detecting the maximum likelihood delay

One approach is to fix the delays for every cause-effect pair. We would like to estimate which delay is most probable (most likely) for a given pair of time series. The simplest algorithm would do the following.

1. Fix a reasonable upper bound on the delay.

2. For every possible value of the delay t find the correlation between the effect Y and the

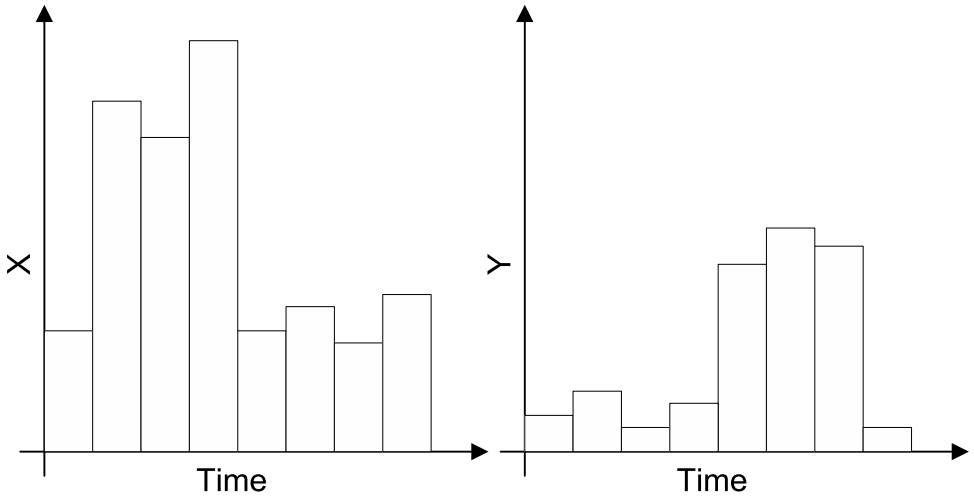


Figure 1: Uniquely determined delays.

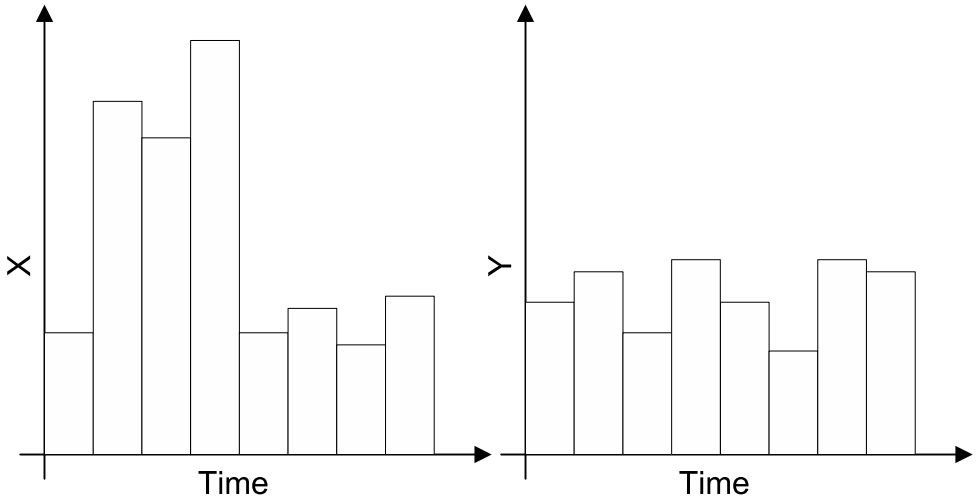


Figure 2: No reasonable correlation.

cause X shifted by t. We can use the Pearson correlation coefficient:

$$r_t = \frac{\sum (x_{i+t} - \bar{x})(y_i - \bar{y})}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}. \tag{1}$$

3. Choose the delay with maximal correlation:

$$d = \mathrm{argmax}_t r_t. \tag{2}$$

This simple algorithm allows to find the delay which maximizes correlation between the time series of the effect and the shifted cause. Therefore, the resulting delay value is most likely given these time series.

In the case when the delay value is clear and unique, like on Figure 1, this algorithm works fine. However, in other situations it has serious disadvantages.

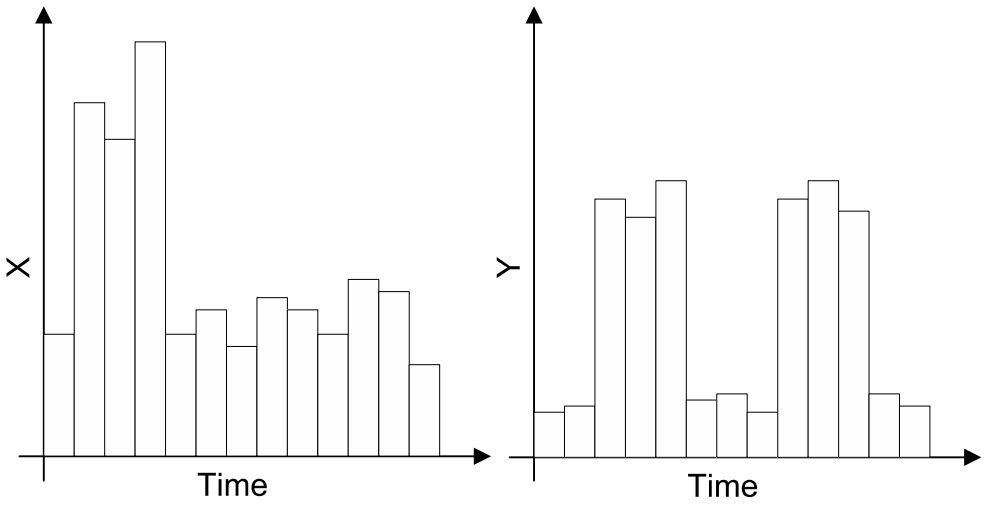


Figure 3: Two equally possible delays.

For example, on Fig. 2 there appears to be simply no reasonable correlation of the two time series (at least during the time span shown). In this case we would like our system to warn the user that these two nodes do not appear to be connected, while the above algorithm will just choose the most correlated delay (depending probably on the noise in data). However, this problem is easy to overcome: we can simply set a default threshold and on step 3 of the algorithm output d only if it exceeds this threshold.

But there are other problems. Look at Fig. 3. There are two plausible delays here: one time unit and seven time units. The algorithm above would choose one of them based on some miniscule correlation advantage which would most probably be accounted for by the noise in the data. In fact, we would like the system to accept the fact that it doesnt know what the delay really is. Of course, we could (and in some situations should) ask the user to choose one of the possibilities. But if the user does not have enough information to perform the choice, we should look for alternative ways.

3.3 Probability labels in the intermediate graph

Fortunately, our formalism is rich enough to handle this situation in the most natural way — consider all delays as possible and range the probabilities of these delays with respect to their correlations:

$$p_t = \frac{r_t}{\sum_t r_t}. \tag{3}$$

Let us make the assumption that there is only one value of the delay (that is, a single event cannot trigger another event twice). In this case, we can view the events "Y at time t" as mutually incompatible. On the level of the BBN this means that we can view Y as a node with T states corresponding to T statements: "Y at time 1", ..., "Y at time T". And we already know the probabilities we would like to assign:

$$p(Y \text{ at time } t_0 + t \mid X \text{ at time } t) = p_t = \frac{r_t}{\sum_t r_t}. \tag{4}$$

This BBN is shown on the left of Fig. 4. The right part of the same figure depicts the construction in the case when events are not instantaneous, but can last for several time intervals.

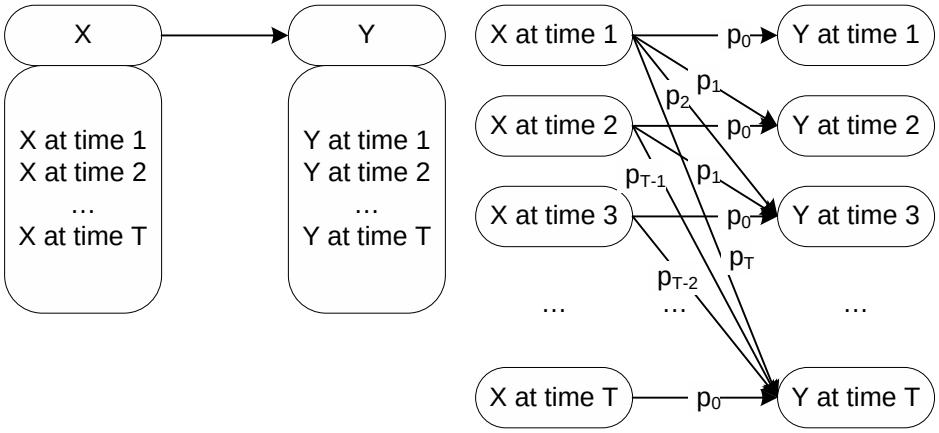


Figure 4: Intermediate event bush.

There is a natural modification of this algorithm if we do not want to have so many edges in the intermediate event bush. We can set a threshold of the correlation and assign nonzero probabilities only to the delays that exceed this threshold (assigning probabilities proportional to the correlations, as above). This way there will be much fewer edges in the intermediate graph with no substantial loss in representativeness.

3.4 Using variograms to estimate time intervals

Learning conditional probabilities is only a part of the job with specifying time labels on an event bush. Another important part would be to specify which time intervals to consider and how to break them up into smaller intervals. We propose a way to learn these attributes automatically from variograms of the input time series data.

Let us define for a stochastic process $V(t)$ its *variogram* as the following expression:

$$\gamma(\tau) = \frac{1}{2}\mathrm{E}\left[(V(t+\tau) - V(t))^2\right]. \tag{5}$$

Informally, a variogram, first applied to the estimation of volcanic hazard by Jaquet et al. [2000], captures the memory of the stochastic process V; it measures the correlation of V with itself shifted in time. If V has no memory of its past, $\gamma(\tau)$ will be simply equal to the variance of V for all $\tau > 0$. However, if V does depend on its own history, γ will gradually rise from 0 (note that $\gamma(0) = 0$) to the variance; as soon as $\gamma(\tau)$ stabilizes around the variance of V, this means that the process has effectively forgotten all about the zero point. For a much more detailed discussion of variograms see Carniel et al. [2008] and references therein.

For example, Carniel et al. [2008] depict the memory of the seismic noise recorded in March, 2005, on Tenerife, close to Las Cañadas caldera. The variogram reaches the variance after about 80 hours and remains at about the same level afterwards. This means that the process causing this seismic noise depended only on the history of its last 80 hours.

How does all this help us create event bushes? When we define the time intervals for nodes in event bushes, it would only make sense to define it as long as previous history may affect the current events. For example, we could go as far as 200 hours on the seismic noise, but since the autocorrelation of this noise reaches zero after 80 hours, there would be no meaningful links from the noise that far ago to the present time. Thus, for seismic noise it would suffice to consider a scale of 80 hours.

Note, however, that the overall time interval depends on all nodes in the bush. Perhaps there is

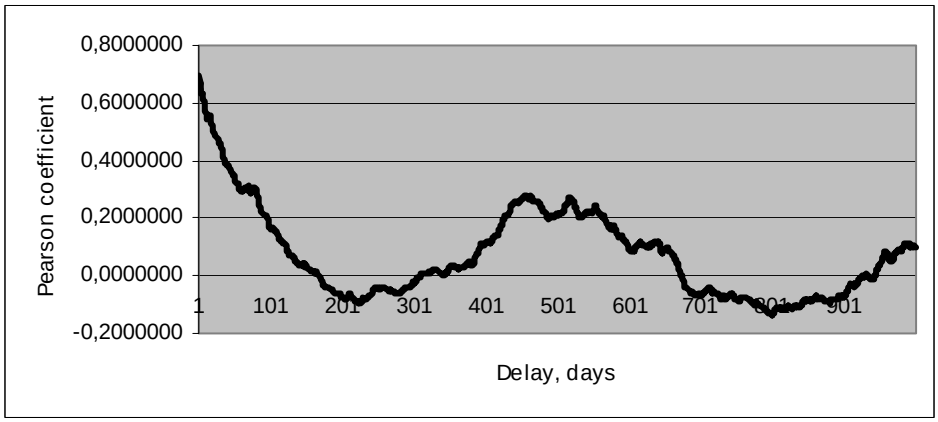


Figure 5: Correlation between events per day and tremor as a function of the delay in data.

some other event affected by seismic noise that has a longer memory (say, 200 hours). In this case, we would have to consider seismic noise from 200 hours ago because it might affect the current situation *via* this event. In general, however, calculating variograms provides us with an easy automatic way to estimate time scales of different events, and its use can be extended to inter-event dependence by the so called cross variogram (see e.g. Jaquet and Carniel [2003], Matheron [1962]).

4 AN EXAMPLE

As an example of our approach we took the data from Stromboli volcano (Italy) that has been thoroughly studied in Jaquet and Carniel [2001], Jaquet and Carniel [2003], Carniel et al. [2008]. These studies have shown that Stromboli exhibits an extraordinary long memory of its previous activity. Variogram analysis performed by Jaquet and Carniel [2001] showed that the memory could last longer than 400 days.

For our study we took the records collected from 1992 till 2001 of the number of volcanic event per day registered at Stromboli together with the average tremor intensity recorded on that day. The graph shown on Fig. 5 depicts the Pearson coefficient between these data sets as a function of the delay between the two measured variables.

The graph shows that the memory of these two mutually dependent datasets is probably limited in this case to about 150 days (this is when the correlation first reaches zero). However, one would lose much of the structure if he would have to limit the model to immediate consequences, discarding the delayed correlations at all. Thus, in this case it would be sensible to set a minimal threshold of the correlation (say, 0.2) and relate a node corresponding to tremor at time t to nodes corresponding to the number of events up to time $t + \Delta t$, where Δt is the period of time after which the Pearson coefficient reaches the threshold for the first time.

As for computational issues, if an event bush has a relatively simple structure, then multiplying it by 150 would not be unfeasible. However, if the bush has many nodes, the direct day–by–day copying method could fail. In this case, we suggest merging neighboring intermediate nodes (that is, enlarging the time unit). However, in the areas where the correlation is high the time unit may be split up further. For example, in this case we could use day–by–day nodes up until about day 50. After that, we could average the data over 5 or even 10 days without losing too much of the expressive power (and common sense).

5 CONCLUSION AND FURTHER WORK

We have presented some basic ideas on how to use time series data, very common in geosciences, for learning probabilities in event bushes with temporal and spatial labels. Note that this is the first implementation of an event bush mechanism where we actually learn probabilities in the Bayesian network from data, rather than assign them on the basis of logical considerations.

Obviously, open questions remain. The current implementation of the event bush technology leaves much for the user to specify concerning temporal and spatial labels. First, the spatial issues still remain largely uncovered by the current formalism. Space-bound data once again brings us to learning issues: an event bush should be able to automatically benefit from spatial data just like it does from time series. Hopefully, some of the learning could follow the lines of this paper, if an event bush is meant to be used for spatial and temporal modeling simultaneously, and we are provided with space-related time series. Other types of data (especially data of qualitative nature), however, may require additional insights.

In general, we plan to further automatize the creation of event bushes and, ultimately, embed a working prototype of the formalism into some existing GIS framework. This will bring us to real-life applications of the event bush technology.

ACKNOWLEDGEMENTS

The research has been funded by Project V4 of Istituto Nazionale di Geofisica e Vulcanologia, Italy, "Development of quantitative tools to evaluate volcanic hazard", a Human Capital Foundation grant, Information Technologies cluster project, 2006, and NATO-Russia collaborative linkage grant 982484.

The authors thank Oksana Kanzheleva for fruitful discussions and the anonymous referees for numerous useful suggestions.

REFERENCES

Carniel, R., O. Jaquet, and M. Tárraga. Perspectives on the application of the geostatistical approach to volcano forecasting at different time scales. In *Caldera volcanism: Analysis, modelling and response, Developments in Volcanology*. Elsevier, 2008.

Jaquet, O., S. Low, B. Martinelli, V. Dietrich, and D. Gilby. Estimation of volcanic hazards based on cox stochastic processes. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 25(6–7):571–579, 2000.

Jaquet, O. and R. Carniel. Stochastic modelling at stromboli: a volcano with remarkable memory. *Journal of Volcanology and Geothermal Research*, 105:249–262, 2001.

Jaquet, O. and R. Carniel. Multivariate stochastic modelling: towards forecasts of paroxysmal phases at stromboli. *Journal of Volcanology and Geothermal Research*, 128(1–3):261–271, 2003.

Jensen, F. V. *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag, 2002.

Matheron, G. *Traité de Géostatistique appliqueé. Tome* 1. Technip, Paris, 1962.

Neapolitan, R. E. *Learning Bayesian Networks*. Pearson Prentice Hall, 2004.

Pshenichny, C. A. A draft for complex formal approach in geoscience: modeling geohazards. In *Proceedings of IAMG '03*, 2003.

Pshenichny, C. A. and S. I. Nikolenko. Temporal assessment by means of an event bush. *Geophysical Research Abstracts*, 9(00497), 2007.

Pshenichny, C. A., S. I. Nikolenko, R. Carniel, A. L. Sobissevitch, P. A. Vaganov, Z. V. Khrabrykh, V. P. Moukhachov, V. L. Shterkhun, A. A. Rezyapkin, A. V. Yakovlev, R. A. Fedukov, and E. A. Gusev. The event bush as a potential complex methodology of conceptual modelling in the geosciences. In *Proceedings of the 4th Biennial Meeting of iEMSs*, 2007a.

Pshenichny, C. A., S. I. Nikolenko, R. Carniel, P. A. Vaganov, Z. V. Khrabrykh, V. P. Moukhachov, V. L. Akimova-Shterkhun, and A. A. Rezyapkin. The event bush as a semantic-based numerical approach to natural hazard assessment (exemplified by volcanology). *Computers and Geosciences*, to appear, 2008.

Pshenichny, C. A., S. I. Nikolenko, A. L. Sobissevitch, and A. V. Yakovlev. Spatial volcanic hazard assessment by the event bush method. In *Proceedings of the XXIV IUGG General Assembly*, 2007b.

Tulupyev, A. L., S. I. Nikolenko, and A. V. Sirotkin. *Bayesian Networks: A Probabilistic Logic Approach*. St. Petersburg, Nauka, 2006.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Towards a Framework for the Development of Intelligent Environmental Decision Support Systems

M. Sànchez-Marrè [a], J. Comas[b], I. Rodríguez-Roda[b], M. Poch[b] and U. Cortés [a]

[a] Knowledge Engineering and Machine Learning Group (KEMLG), Computer Software Dept., Technical University of Catalonia, Barcelona, Catalonia ({miquel, ia}@lsi.upc.edu)

[b] Chemical and Environmental Engineering Laboratory (LEQUIA), University of Girona, Girona, Catalonia ({quim, ignasi, manel}@lequia.udg.cat)

Abstract: Intelligent Environmental Decision Support Systems (IEDSS) are present in the environmental management process at different levels such as hazard identification, risk assessment, risk evaluation and intervention decision-making, but there is neither a well defined methodology nor a framework for the development of IEDSSs. The purpose of this paper is to provide a study on several approaches found in the literature to set-up a methodology and a plausible architecture for the development of IEDSSs. Also a general outlook into the classical decision making theories is done to provide a general and well based architecture. Taking into account the review of proposals, and some previous general architectures proposed by the authors, a new comprehensive architecture and associated methodology to develop an IEDSS for whatever domain is proposed. The methodology proposed and the architecture joins the main three cognitive tasks involved in a decision making process: analysis tasks, synthesis tasks, and prognosis tasks.

Keywords: Intelligent Environmental Decision Support Systems.

1. INTRODUCTION

Environmental Systems are usually very complex systems which are not easily managed. Environmental systems are stochastic and, very often, are multi-scale, spatial and temporal-dependent processes. They also tend to comprise complex interactions among physical, chemical and biological processes. These processes may not be known well and/or may be difficult to represent, causing considerable uncertainty. Some of the sources of this uncertainty can be tamed with additional data or further investigation, but this uncertainty becomes insurmountable especially when characterised by chaotic behaviour or self-organising processes.
Therefore, advocating a single perspective that encompasses everything in a system is becoming increasingly difficult and ineffective. The consensus is developing that environmental issues must be considered in terms of complex systems. But not all environmental systems present the same level of complexity in terms of both the degree of uncertainty and the risk associated with decisions. If the degree of complexity is represented as a function of uncertainty, on one hand, and the magnitude or importance of the decision, on the other hand, then we might distinguish three levels of complexity [Funtowicz and Ravetz 1993; 1999]:

- The *first level of complexity* would correspond to simple, low uncertainty systems where the issue at hand has limited scope. A single perspective and simple models would suffice to provide a satisfactory description of the system.

- The *second level* would correspond to systems with a higher uncertainty degree where simple models can no longer provide satisfactory descriptions. Acquired experience then becomes more and more important, and the need to involve experts in problem solving becomes advisable.
- Finally, the *third level* would correspond to truly complex systems, where much epistemological or ethical uncertainty exists, where uncertainty is not necessarily associated with a higher number of elements or relationships within the system, and where the issues at stake reflect conflicting goals. It is then crucial to consider the need to account for a plurality of views or perspectives.

In fact, these three levels of complexity are very similar to the original classification of different kind of problems proposed by H. A. Simon [1960]: structured, semi-structured and unstructured. These levels of complexity are depicted in figure 1.

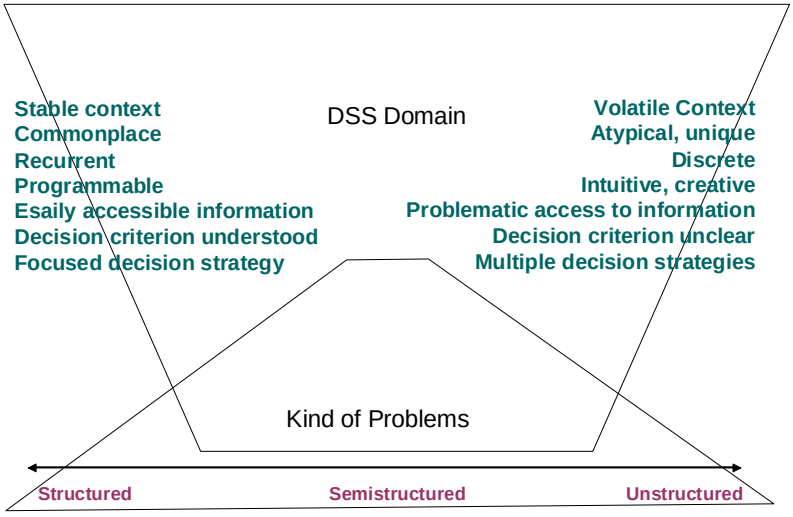


Figure 1. Kind of problems following Simon's proposal

In this sense, it is important to realise that environmental problems are characterized by dynamics and interactions that do not allow for an easy division between social and biogeophysical phenomena. Much ecological theory has been developed in systems where humans were absent or in systems where humans were considered an exogenous, simple and detrimental disturbance. The intricate ways in which humans interact with ecological systems have been rarely considered [Kinzig 2001]. Embracing a socioeconomic perspective implies accepting that all decisions related to environmental management are characterised by multiple, usually conflicting objectives, and by multiple criteria (Ostrom 1991). Thus, in addition to the role of experts, it becomes increasingly important to consider the role of wide public participation in the decision making processes. Experts are consulted by policy makers, the media, and the public at large to explain and advise on numerous issues. Nonetheless, many recent cases have shown, rather paradoxically, that while expertise is increasingly sought after, it is also increasingly contested [Ludwig 2001].

In the last decades, mathematical/statistical models, numerical algorithms and computer simulations have been used as the appropriate means to gain insight into environmental management problems and provide useful information to decision makers. To this end, a wide set of scientific techniques have been applied to environmental management problems for a long time and with good results.

But most of these efforts were focused on problems that could be assigned to the first level of complexity (i.e., simple, low uncertainty systems where the issue at hand has limited scope), following the classification of [Funtowicz and Ravetz, 1993; 1999]. Consequently, many complex environmental problems have not been effectively addressed by the scientific community. However, the effort to integrate new tools to deal with more

complex systems has led to the development of the so-called Environmental Decision Support Systems (EDSSs) [Guariso and Werthner, 1989],[Rizzoli and Young, 1997].

EDSSs have generated high expectations as a tool to tackle problems belonging to the second and third levels of complexity. Thus in a recent review of the relevant literature in the topic, more than 600 references were found (including journal articles, conference papers, and technical reports) during the 90s, with only 10 references in 1992 and more than 150 references per year towards the end of the decade. The range of environmental problems to which EDSSs have been applied is wide and varied, with water management at the top (25% of references), followed by aspects of risk assessment (11.5%) and forest management (11.0%). Equally varied are the tasks to which EDSSs have been applied, ranging from monitoring and data storage to prediction, decision analysis, control planning, remediation, management, and communication with society. Intelligent Environmental Decision Support Systems (IEDSS) are the envisioned new tools to cope with the inherent complexity of environmental problems [Sànchez-Marrè et al., 2008]. Other traditional computational approaches do not seem to be enough flexible and capable to safely and successfully handle complex real-world environmental problems.

2. IEDSSs: ORIGINS AND PREVIOUS APPROACHES

Environmental issues belong to a set of critical domains where wrong management decisions may have disastrous social, economic and ecological consequences. Decision-making performed by EDSSs should be collaborative, not adversarial, and decision makers must inform and involve those who must live with the decisions. EDSS should be not only an efficient mechanism to find an optimal or sub-optimal solution, given any set of whimsical preferences, but also a mechanism to make the entire process more open and transparent. In this context, Intelligent EDSSs or IEDSS can play a key role in the interaction of humans and ecosystems, as they are tools designed to cope with the multidisciplinary nature and high complexity of environmental problems. In the following we shall describe the nature of IEDSS.

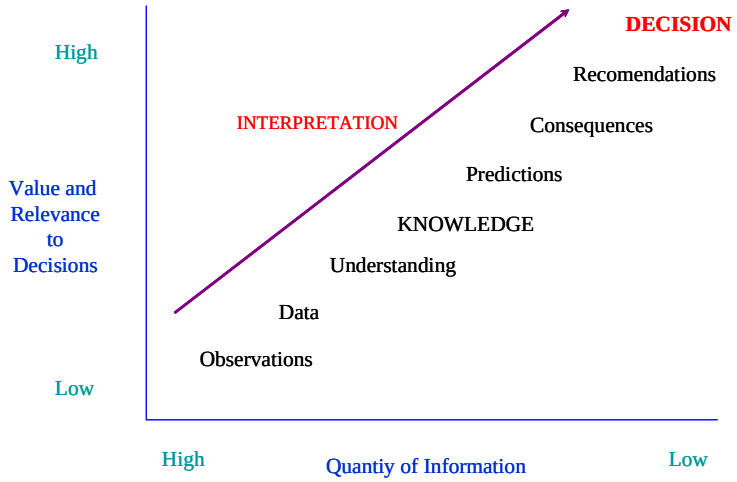


Figure 2. Interpretation process: from observations to decision

From a functional point of view, and taking into account the type of problem that the IEDSS solves, two kinds of IEDSS can be distinguished but of course most systems of interest fall between these two categories:

- The first category, which we could name as *dynamic IEDSS*, are those IEDSS which aim to control or supervise a process in real-time (or almost real-time), facing similar situations on a regular basis [Sànchez-Marrè et al. 1996]. They must guarantee robustness against noise, missing data, typos and any combination of input data. In general the end-user is responsible for accepting, refining or

rejecting system solutions. This responsibility can decrease, thereby increasing IEDSS confidence, over time in as far as the system is facing situations that were successfully solved in the past (real validation).

- In the second category, which can be named as *static IEDSS*, there are those IEDSS that give punctual support to decision-making, and are mainly used to justify multicriteria decisions of policy-makers more than to make real decisions on a day-to-day basis [Comas et al. 2003]. Here it is interesting for the end-user to play with what-if scenarios, to explore the response surface and the stability of the solution; for example how sensitive our decision is to small variations in the given weight and value of the relevant variables. The role of socio-cultural and economic issues limits the use of standard databases. Confidence can not be increased in the results when facing similar situations, because these IEDSS are very specific and sometimes are only built to make or justify one decision.

2. 1 Previous approaches

Decision theory is a field which has been studied long time ago from the perspective of business, economics and management sciences. One of the most well known decision making model was proposed by H. A. Simon [1960].

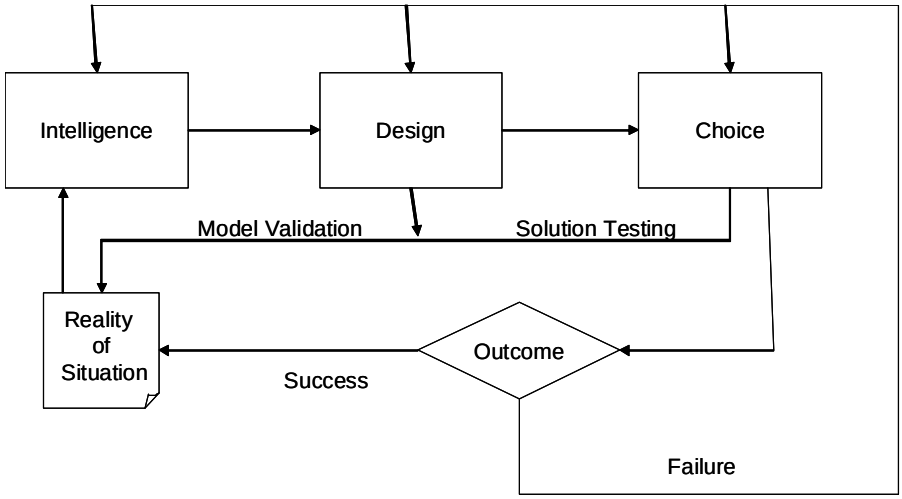


Figure 3. Decision making model by Simon

The three levels of decision process according to Simon are:

- Intelligence: at this level, there is the search for information and knowledge, in order to identify a problem requiring a decision-making task.
- Design: at this stage, the building and analysis of possible alternatives or strategies to solve the detected problem is done.
- Choice: the decisor selects one of the possible alternatives analyzed and generated in the previous step. The selected alternative satisfies the rational behaviour criterion. This means the selected alternative is the best one having in mind the fact that it provides the decisor with the maximum utility or benefit.

According to Fox and Das [2000], a decision support system is a computer system that assists decision makers in choosing between alternative beliefs or actions by applying knowledge about the decision domain to arrive at recommendations for the various options. It incorporates an explicit decision procedure based on a set of theoretical principles that justify the "rationality" of this procedure. Thus, an intelligent information system reduces the time in which decisions are made in a domain, and improves the consistency and quality of those decisions [Haagsma and Johanns 1994].

Thus IEDSSs could be defined [Sojda 2002] as systems using a combination of models, analytical techniques, and information retrieval to help develop and evaluate appropriate

alternatives [Adelman 1992; Sprague and Carlson 1982]; and such systems focus on strategic decisions and not operational ones. More specifically, decision support systems should contribute to reducing the uncertainty faced by managers when they need to make decisions regarding future options [Graham and Jones 1988]. Distributed decision making suits problems where the complexity prevents an individual decision maker from conceptualizing, or otherwise dealing with the entire problem [Boland et al. 1992; Brehmer 1991]. Other definitions could be found in D'Erchia et al. [2001].

Cortés *et al*. [2000] proposed an IEDSS architecture based on five steps (see Figure 4):

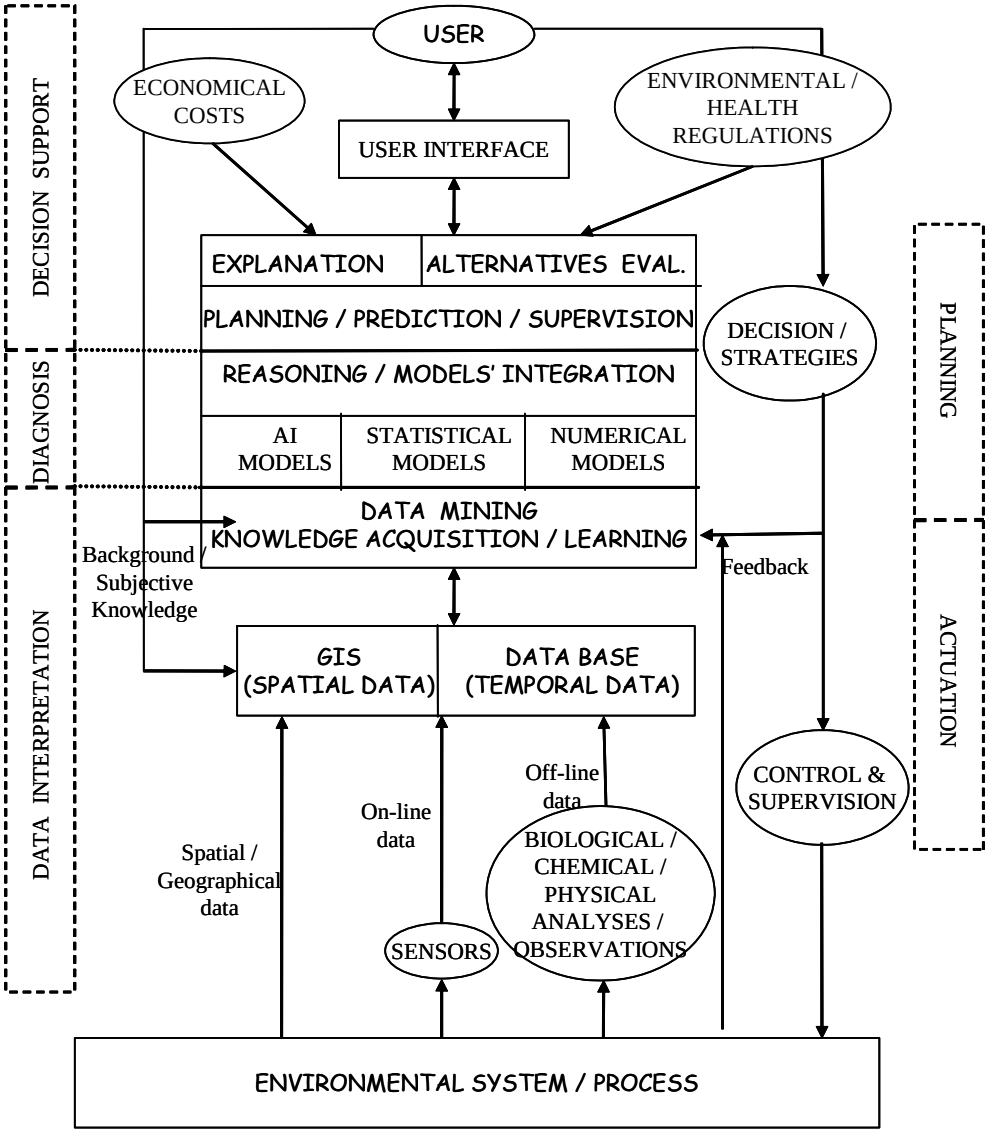


Figure 4. Dynamic IEDSS Architecture by Cortés et al. [2000]

- The first step of the IEDSS (*data interpretation*) encompasses the tasks involved in data gathering and registration into databases. Original raw data are often defective, requiring a number of pre-processing procedures before they can be registered in an understandable and interpretable way. Missing data and uncertainty must be also considered in this level. Also, the knowledge discovery step including data mining techniques are included here providing the IEDSS with the environmental process knowledge.
- The second step, *diagnosis level*, includes the reasoning models that are used to infer the state of the process so that a reasonable proposal of actuation can be

reached. This is accomplished with the help of statistical, numerical and artificial intelligence models, which will use the knowledge previously acquired.

- The third step, *decision support level*, establishes a supervisory task that entails gathering and merging the conclusions derived from AI knowledge models and numerical models. This level also raises the interaction of the users with the computer system through an interactive and graphical user-machine interface. When a clear and single conclusion can not be reached, a set of decisions ordered by their probability or certainty degree should be presented to the user.
- In the fourth level, plans are formulated and presented to managers as a list of general actions or strategies suggested to solve a specific problem.
- The set of actions to be performed to solve problems in the domain considered are the fifth and last step. The system recommends not only the action, or a sequence of actions (a plan), but a value that has to be accepted by the decision maker. This is the final step in the architecture that closes the loop.

Decisions are made when a deviation from an expected, desired state of a system is observed or predicted. This implies a problem awareness that in turn must be based on information, experience and knowledge about the process. Those systems are built by integrating several artificial intelligence methods, geographical information system components, mathematical or statistical techniques, and environmental/health ontologies, and some minor economic components. Examples are the works by Dorner et al. [2007], Reichert et al. [2007] and Cortés et al. [2001]. This progression in complexity of the methods, and in the intensive use of knowledge usually required to develop an IEDSS, corresponds to an increase in data required to support the models (see Fig. 2, adapted from Wittaker [1993]).

3. A COGNITIVE-ORIENTED APPROACH FOR DEVELOPMENT OF IDSSs

Taking into account the different approaches for the development of IEDSS proposed by several authors, there is the need for a common framework to develop an IEDSS either it is a static IEDSS or a dynamic IEDSS.

Studying the several proposals and taking into account the main cognitive tasks of human beings within the decision-making environment, the common underlying tasks in most of IEDSS can be extracted. The following general architecture is proposed for IEDSS. It is based on a three-layer architecture:

- **Analysis tasks**: In this layer is where most of interpretative processes are run. At this stage the data gathering processes as well as the knowledge discovery process by means of some data mining techniques are undertaken to get *diagnostic models*. These models will provide the IEDSS with hard analytical power to get an insight of the environmental system/ process being supervised in real-time or managed in an off-line basis. Most of these techniques come from the data mining field.

- **Synthesis tasks**: This layer wraps all the work necessary to synthesizes possible alternative solutions for the different diagnostics found in the previous step. This synthetisation can be done through several *solution generation methods* based on Statistical techniques, on Artificial techniques or on Mathematical/Numerical techniques. Of course the integration of different nature methods could enhance the problem solving ability of the IEDSS.

- **Prognosis tasks**: At this upper layer relay the inherent ability of IEDSS to decision support tasks. At this level, the several *predictive models*, which can be numerical (mostly simulations) or rather qualitative (qualitative reasoning or qualitative simulations), are used to estimate the consequences of several actions proposed in the previous step by the solution generation methods. These “what if” models let the final user/decisor to make a decision based on the evaluation of

several possible alternatives. At this stage, the temporal and spatial features could be very important for a good environmental modelling practise.

These three-layer architecture is depicted in figure 5.

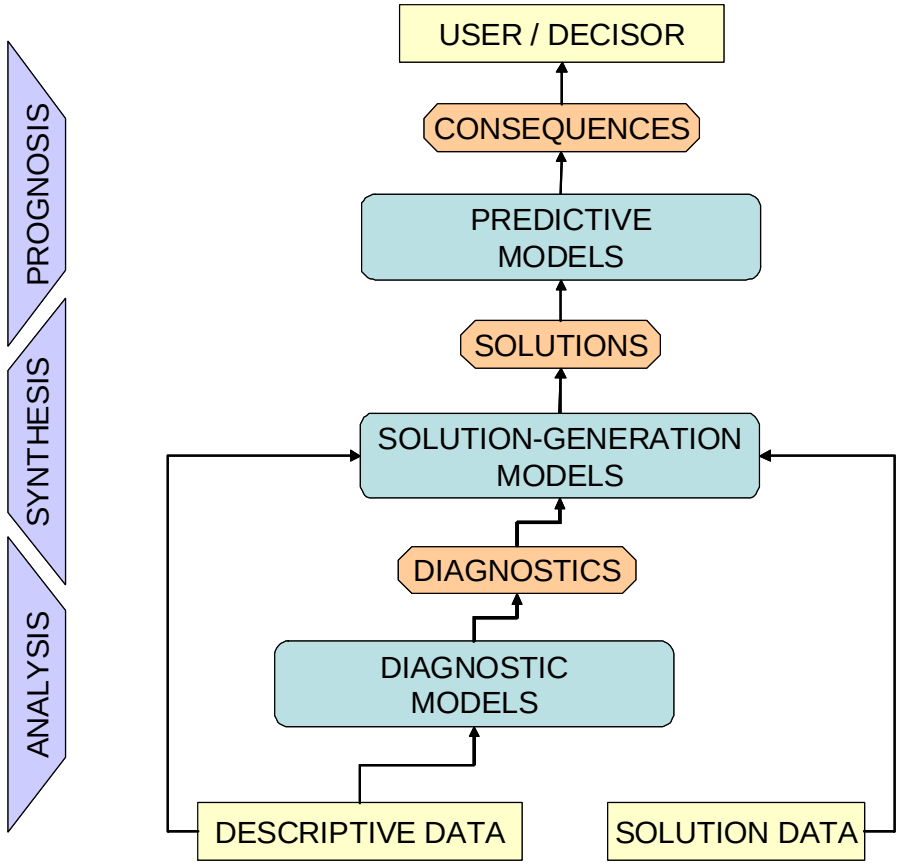


Figure 5. The cognitive-oriented approach for IEDSS development

4. CONCLUSIONS

IEDSS can play a key role in the interaction of human beings and ecosystem, as they are tools designed to cope with the multidisciplinary nature and high complexity of environmental problems. IEDSS help in the design of understandable, credible, and easily accessible information systems that are necessary to improve our still incomplete knowledge and understanding of the ecosystem.

Decisions are needed to manage complex environmental systems. This implies a problem awareness that in turn must be based on information, experience and knowledge about the process. IEDSS are built by integrating several artificial intelligence methods, geographical information system components, mathematical or statistical techniques, and environmental/health ontologies, and some minor economic components.

Although there are some architecture proposals found in the literature to combine all this techniques and knowledge, there is not a common framework to be taken into account as first guideline to deploy both static and dynamical IEDSS.

In this paper, a new cognitive-oriented architecture has been proposed, with three main cognitive tasks as a common framework for IEDSS. This architecture is wrapping both the the emphasis of static IEDSs, which relies in the second and third layer, and the dynamical IEDSS, which commonly focus on the first and second layers.

This approach has been used in the development of a wastewater treatment plant supervisory system as described in [Poch et al., 2004].

ACKNOWLEDGEMENTS

The authors wish to thank the partial support of Spanish project TIN2004-01368.

REFERENCES

Adelman, L., 1992. Evaluating Decision Support and Expert Systems. John Wiley and Sons, New York, NY.

Boland, R.J., Mahewshwari, A.K., Te'eni, D., Schwartz, D.G., Tenkasi, R.V., 1992. Sharing perspectives in distributed decision making. In: Proceedings of the Conference on Computer supported Cooperative Work. Association for Computing Machinery, New York, NY.

Brehmer, B., 1991. Distributed decision making: Some notes on the literature. In: Rasmussen, J., Brehmer, B., Leplat, J. (Eds.), Distributed Decision Making: Cognitive Models for Cooperative Work. John Wiley and Sons, Chichester, England.

Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M., Cortés, U., Freixó, A., Arráez, J., Poch, M., 2003. A knowldge-based approach to the deflocculation problem: integrating on-line, off-line, and heuristic nformation. Water Research 37 (10), 2377–2387.

Cortés, U., Sànchez-Marrè, M., Cecaronni, L., R.-Roda, I., Poch, M., 2000. Artificial intelligence and environmental decision support systems. Applied Intelligence 13 (1), 77–91.

Cortés, U., Sànchez-Marrè, M., Sangüesa, R., Comas, J., R-Roda, I., Poch, M., Riaño, D., 2001. Knowledge management in environmental decision support systems. AI Communications 14 (1), 3–12.

D'Erchia, F., C, Korschgen, M. Nyquist, R. Root, R. Sojda, and P. Stine. 2001. A framework for ecological decision support systems: building the right systems and building the systems right. U.S. Geological Survey, Biological Resources Division, Information and Technology Report USGS/BRD/ITR-2001-0002.

Dorner, S., Shi, J., Swayne, D., 2007. Multi-objective modelling and decision support using a Bayesian network approximation to a non-point source pollution model. Environmental Modelling and Software 22, 211–222.

Fox, J., Das, S., 2000. Safe and sound. In Artificial Intelligence in Hazardous Applications. AAAI Press/The MIT Press.

Funtowicz, S.O., Ravetz, J.R., 1993. Science for the post-normal age. Futures 25 (7), 739–755.

Funtowicz, S.O., Ravetz, J.R., 1999. Post-normal science—an insight nom maturing. Futures 31 (7), 641–646.

Graham, I., Jones, P.L., 1988. Expert Systems: Knowledge, Uncertainty, Decision. Chapman and Hall, New York, NY.

Guariso, G., Werthner, H., 1989. Environmental Decision Support Systems. Ellis Horwood-Wiley, New York.

Haagsma, I.G., Johanns, R.D., 1994. Decision support systems: an integrated approach. In: Zannetti, P. (Ed.), Environmental Systems, vol. II., pp. 205–212.

Kinzig, A., 2001. Bridging disciplinary divides to address environmental and intellectual challenges. Ecosystems 4, 709–715.

Ludwig, D., 2001. The era of management is over. Ecosystems 4, 758–764.

Olsson, G., Aspegren, H., Nielsen, M.K., 1998. Operation and control of wastewater treatment—a Scandinavian perspective over 20 years. Water Science and Technology 37 (12), 1–13.

Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. and Cortés, U. Designing and building real environmental decision support systems. *Environmental Modelling and Software*, 19(9), 857-873, 2004.

Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Sporri, C., Tockner, K., Truffer, B., 2007. Concepts of decision support for river rehabilitation. Environmental Modelling and Software 22, 188–201.

Rizzoli, A.E., Young, W.Y., 1997. Delivering environmental decision support systems: software tools and techniques. Environmental Modelling and Software 12 (2-3), 237–249.

Sànchez-Marrè, M., Gibert, K., Sojda, R. S., Steyer, J.P., Struss, P., Rodríguez-Roda, I., Comas, J., Brilhante, V. and Roehl, E.A. Chapter 8: Intelligent Environmental Decision Support Systems. *In: Jakeman, A.J., Rizzoli, A., Voinov, A. and Chen, S. (Eds.), State of the art and Futures in Environmental Modelling and Software*, Elsevier Science, Amsterdam, The Netherlands (in press), 2008.

Sànchez-Marrè, U., Cortés, J., Lafuente, I., R.-Roda, I., Poch, M., 1996. DAI-DEPUR: a distributed architecture for wastewater treatment plants supervision. Artificial Intelligence in Engineering 10 (3), 275–285.

Simon, H. A. *The New Science of Management Decision*. New York: Harper & Row, 1960.

Sojda, R.S., 2002. Artificial intelligence based decision support for trumpeter swan management. PhD Dissertation. Colorado State University. Fort Collins, Colorado.

Sprague Jr., R.H., Carlson, E.D., 1982. Building Effective Decision Support Systems. Prentice-Hall, Englewood Cliffs, New Jersey.

Wittaker A.D. Decision support systems and expert systems for range science. In *Decision Support Systems for the Management of Grazing Lands: Emerging Lands*, edited by J.W. Stuth and B.G. Lyons, pp. 69–81, 1993.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Soil Moisture Monitoring at Different Scales for Rainfall-Runoff Modelling

L. Brocca[a], F. Melone[a], and T. Moramarco[a]

[a] *National Research Council, Research Institute for Geo-Hydrological Protection, Via Madonna Alta 126, 06128 Perugia, Italy (brocca@irpi.cnr.it)*

Abstract: The soil moisture is a paramount quantity in the study of hydrologic phenomena and soil atmosphere interaction. Because of its high spatial and temporal variability, the soil moisture monitoring scheme was here investigated in view of its application in rainfall-runoff modelling. The results obtained through a "direct" analysis and an "indirect" analysis were found quite similar in terms of contents and useful to address the estimation of soil moisture at the catchment scale. Based on the direct analysis, the soil moisture samples number to get a reliable estimation of soil moisture at catchment scale was defined. The indirect analysis showed that the catchment wetness conditions at the beginning of a storm event can be accurately estimated through the soil moisture observations carried out in a small area.

Keywords: Soil moisture; rainfall-runoff modelling; experimental catchments; antecedent wetness conditions, Soil Conservation Service - Curve Number.

1. INTRODUCTION

Surface soil moisture status is widely recognized as a key parameter in environmental processes because of its control in partitioning the rainfall into runoff and infiltration. For storm rainfall-runoff modeling the knowledge of the wetness conditions at the onset of precipitation is one of the most important aspect mainly for basins in the Mediterranean area [Brocca et al., 2008a]. Because of its high spatio-temporal variability, soil moisture monitoring for large areas is by no means a simple task. In fact, conventional in situ measurements are costly and provide information at only a few selected points, independently by the method used (Time Domain Reflectometry, gravimetric, neutron probes, capacitance probe). Remote sensing techniques, as opposed to point measurements, show promise in assisting hydrologist in describing and measuring surface soil moisture. However, ground-based data are used to assess the accuracy of any remote sensing algorithm being a fundamental component of data analysis. Therefore, the key question for soil moisture monitoring on areas of interest for hydrological applications is how many points must be sampled to estimate a rational mean value. However, intensive field samplings carried out in different climatic regions have shown contrasting results with respect to the main statistical [Brocca et al., 2007a; Famiglietti et al., 2007] and geostatistical [Fitzjohn et al., 1998, Western et al., 2004] properties, to the factors influencing the variability [Grayson et al., 1997; Famiglietti et al., 1998] and, generally, to the entity of soil moisture variability.

In terms of simulation and flood forecasting the interest lies mainly in deriving Antecedent Wetness Conditions (AWC) for rainfall-runoff modelling. Several studies investigated the effect of soil moisture spatial variability on the hydrologic response of a catchment through simulation [Grayson et al., 1995; Merz and Plate, 1997; Bronstert and Bardossy, 1999] and experimental studies [Goodrich et al., 1994; Aubert et al., 2003]. Grayson and Western [1998] suggested that a network of a limited number of moisture sensors could provide reliable estimates of areal mean soil moisture thus addressing the AWC assessment.

Therefore, indicative wetness conditions derived by satellite and/or few point measurements might be sufficient to determine AWC, rather than the use of complex water balance models or detailed spatial measurements [Longobardi et al., 2003]. On this basis, new methodologies for the optimization of the monitoring strategy with indications of representative locations which guarantee information reliability are welcome [Vachaud et al., 1985; Brocca et al., 2008b].

Based on the above insights, this paper addresses two aspects: the optimal sampling and the AWC assessment. For the first one, the optimal soil moisture monitoring scheme, at different scales, has been investigated for rainfall-runoff modelling. In particular, several soil moisture spatial patterns collected in an inland region of Central Italy were investigated through a "direct" method based on statistical and temporal stability analysis. The second aspect has been addressed through an "indirect" analysis based on the estimation of the AWC for rainfall-runoff events occurred in five nested catchments. Because of the lack of significant flood events in the sampling period, the role of soil moisture data was investigated by using observations carried out continuously in an experimental plot included in the smallest catchment. The results of the different analysis were compared and some general findings for soil moisture monitoring at different scales have been derived.

2. METHODS

A monitoring scheme useful to address the soil moisture spatial mean at catchment scale relies on selecting representative observation sites where an optimal number of point measurements have to be carried out. It is well known that more measurement points are required to take account of soil moisture spatial variability at local scale. Therefore, the optimization of the monitoring scheme has been here addressed through a "direct" method based on the investigation of the soil moisture spatial variability at the plot and small catchment scale by using the classical statistical approaches. In particular, the optimal number of measurements to estimate the mean water content in a given area is determined through the relationship between the mean and the coefficient of variation [Famiglietti et al., 2007].

Moreover, the method also includes the temporal stability analysis, such as proposed by Vachaud et al. [1985]. This analysis is based on the parametric test of the relative differences, of which a brief description is reported below. Let θ_{ij} be the soil moisture at location i at the time j, the spatial mean for each sampling is:

$$\overline{\theta}_j = \frac{1}{N}\sum_{i=1}^{N}\theta_{ij} \tag{1}$$

where N is the number of measurement points. The relative difference, δ_{ij}, is given by:

$$\delta_{ij} = \frac{\theta_{ij} - \overline{\theta}_j}{\overline{\theta}_j} \tag{2}$$

The temporal mean of the relative differences for each location, i, is given by:

$$\overline{\delta}_i = \frac{1}{M}\sum_{j=1}^{M}\delta_{ij} \tag{3}$$

where M is the number of measurement campaigns. Obviously, a "representative" location of the mean is characterized by a low value of $\overline{\delta}_i$. Therefore, a regression analysis between the soil moisture values sampled in the most "representative" site having the lowest $\overline{\delta}_i$ value and the catchment-mean moisture would allow to assess the reliability of soil moisture prediction at large scale through local observations.

As far as the AWC assessment is concerned, the capability to use near-surface soil moisture observations carried out in a small experimental plot included in the catchment of interest can be analyzed through an "indirect" method [Brocca et al., 2007b]. For a specific rainfall-runoff event, the AWC determines the amount of rainfall infiltrating into the soil profile and

hence the surface runoff volume, R_d. Considering the Soil Conservation Service - Curve Number (SCS-CN) method for abstraction, R_d is given by:

$$R_d = \frac{(P - F_a)^2}{P - F_a + S} \qquad P \geq F_a \tag{4}$$

where F_a is the initial abstraction, S is the potential maximum retention, and P is the rainfall depth. The quantity F_a is assumed as a fraction of S, $F_a=\lambda S$, with λ=0.2 as the SCS-CN standard value [Ponce and Hawkins, 1996]. In particular, the parameter S is estimated considering the average dimensionless Curve Number (CN) assessed as a function of land use, hydrological soil group and the AWC. Assuming that the soil/land use characteristics are constant in time, the different values of S from storm to storm are only linked to variations in the AWC. When R_d is known, eq. (4) can be re-written in terms of $S=S_{obs}$ as:

$$S_{obs} = \frac{1}{2\lambda^2}\left(2\lambda P - \lambda R_d + R_d - \sqrt{\lambda^2 R_d^2 - 2\lambda R_d^2 + 4\lambda P R_d + R_d^2}\right) \tag{5}$$

where P and R_d were derived from the available rainfall-runoff events. Therefore, S_{obs} is the value to assign to the AWC to obtain the observed R_d. Considering the relationship between S_{obs} and the soil moisture observed in small area it is possible to asses indirectly the reliability of soil moisture prediction at catchment scale through local observations.

3. STUDY AREA

The Niccone stream is a right tributary of the Upper Tiber River and is located in an inland region of Central Italy (Figure 1). The catchment covers an area of 137 km^2 at Migianella River section (C1) and the elevation ranges between 887 and 249 m above sea level at the outlet, with a mean catchment slope of 25%.

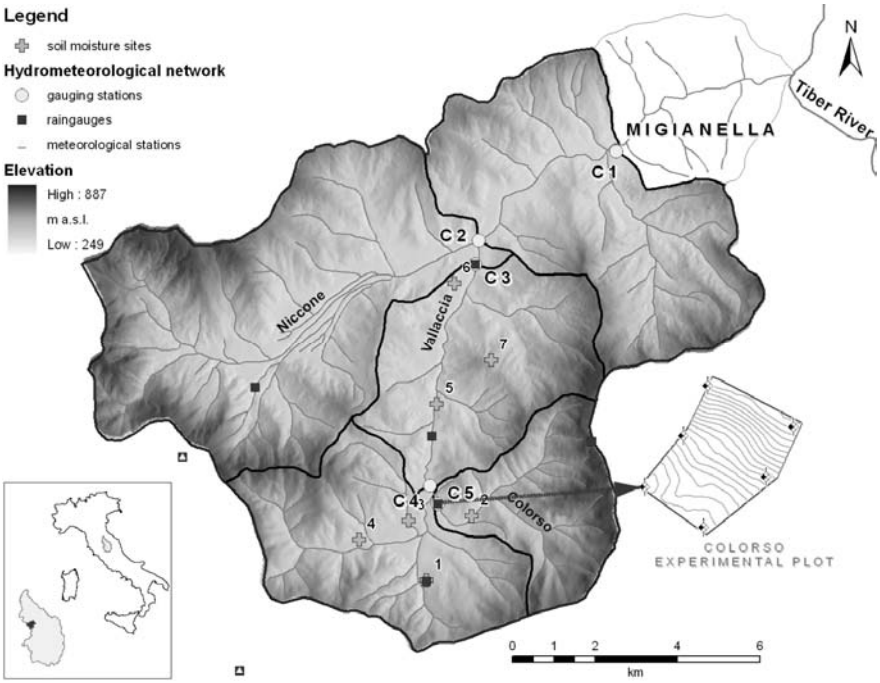


Figure 1. Hydrometeorological network operating in the Niccone catchment. The Colorso experimental plot for continuous soil moisture monitoring and the spot soil moisture measurements sites are also shown.

The catchment is characterized by a Mediterranean climate with average annual precipitation of about 930 mm. In the study area a dense hydrometeorological network has been operating since 1994 with 8 raingauges, 5 hydrometric gauges and 2 meteorological stations (see Figure 1). The data are recorded every 30 minutes and refer to instantaneous values except

for precipitation depth. As it can be seen by Figure 1, the hydrometric stations allow to analyse the hydrological behaviour of five nested catchments. Table 1 shows the main characteristic of each subcatchment including the soil/land use characteristics and the CN values for intermediate wetness conditions, $CN_{(II)}$. In particular, for the soil the percentage of clay and pelitic units is only reported because it affects significantly the watershed CN representing a relative measure of retention of storm water at catchment scale.

Table 1. Main characteristics of the five nested study catchments.

ID	Area (km^2)	Mean slope (%)	River length (km)	Clay and pelitic units (%)	Land Use (%)					$CN_{(II)}$
					wood land	crop land	range land	pasture	urban	
C1	137	25	18.1	15.0	55.0	32.0	10.0	2.7	0.3	77.9
C2	104	25	13.4	10.6	55.5	30.6	10.4	3.1	0.4	77.1
C3	57	24	12.8	6.2	57.7	29.6	7.4	5.1	0.3	74.8
C4	35	25	6.9	7.6	57.6	27.3	8.6	6.0	0.5	74.0
C5	13	28	5.4	1.3	71.0	17.9	0.0	10.9	0.1	72.9

4. DATASET

The near-surface volumetric soil moisture measurements (0-15 cm) were carried out on 7 sites, named $SM_{i=1,2,...,7}$, located inside the Vallaccia catchment (C3), having a drainage area of nearly 60 km^2. For each site thirty spot measurements were collected on a regular grid with a resolution of 10 m using a two-wire connector-type Time Domain Reflectometer (TDR). The measurements were repeated weekly for 35 times in a period of one year from November 2006 to November 2007 except for the summer due to soil hardness.

Inside Vallaccia catchment there is an experimental plot of nearly 1 ha where soil moisture measurements are carried out automatically with the same time step as of the hydrometeorological stations. The plot is covered by grass (permanent pasture) and the soil texture is sandy loam to a depth of 0.3 m and silty loam to a depth of 1.5 m. Six soil moisture probes, based on the Frequency Domain Reflectometry (FDR) technique, continuously measure soil moisture in the soil column at 10 cm, 20 cm and 40 cm, providing at each depth the volumetric soil moisture for a layer thickness of 10 cm. The soil moisture data set selected for this study covers the period from September 2002 to January 2006 except that from July to September 2003 and from September to November 2005, due to drawbacks in the acquisition. The average of the values measured by the six sensors buried at 10 cm depth was assumed as the 'observed' near-surface soil moisture.

Fifteen rainfall-runoff events were selected for this study and their main characteristics are summarized in Table 2 along with the 'observed' near-surface soil moisture at the beginning of each event.

5. RESULTS AND DISCUSSIONS

As mentioned before, the analysis on the soil moisture monitoring was addressed through two methodologies: the "direct" method applied to the dataset of soil moisture spot measurements, and the "indirect" one that makes use of rainfall-runoff data along with the continuous soil moisture data of the experimental plot.

5.1 "Direct" method

Observations in 1 year of soil moisture consist of 35 campaigns at 7 sites (area of 3000 m^2) where for each one 30 spot measurements have been collected for a total of 7350 measurements. Figure 2 shows the temporal pattern of the field-mean soil moisture for each site and clearly shows that the seven patterns are very similar and strongly linked to rainfall. Through this dataset, the spatial and temporal soil moisture variability was investigated considering the Coefficient of Variation (CV) computed in space and in time. For each site the spatial CV was found fairly low (~0.10). On the contrary, the temporal CV considerably

increased and it was found nearly three times the corresponding value obtained for the spatial one. As can be expected, the spatial CV increased up to 0.20 for the whole dataset, representing an area of ~60 km^2.

Table 2. Main characteristics of the selected rainfall-runoff events for the 5 study catchments. The rainfall depth for the largest catchment (C1) and the initial soil moisture 'observed' at the experimental plot, θ_i, are also reported.

Date	Rainfall depth (mm)	Runoff depth (mm)					θ_i (%)
		C1	C2	C3	C4	C5	
Dec 17, 2002	19.7	4.67	6.54	2.60	3.47	1.53	35.95
Dec 28, 2002	10.7	1.11	0.94	0.87	0.74	0.84	36.03
Dec 31, 2002	18.9	3.93	3.85	2.73	1.94	1.45	35.54
Nov 27, 2003	29.3	3.05	2.88	2.25	1.65	1.51	33.88
Feb 19, 2004	30.4	1.98	1.75	1.92	0.98	2.17	30.60
Feb 22, 2004	19.1	3.65	3.60	2.08	3.42	1.74	37.90
Feb 26, 2004	28.5	12.43	10.74	7.96	10.72	5.22	38.18
Mar 7, 2004	19.3	5.60	4.58	2.92	2.52	2.00	35.76
Apr 17, 2004	16.7	1.89	1.24	1.54	0.77	0.49	35.10
Apr 19, 2004	20.9	8.16	6.32	4.94	7.16	8.11	38.31
Apr 11, 2005	25.3	4.39	4.35	3.27	3.29	3.75	35.83
Apr 16, 2005	28.0	4.50	4.17	3.43	3.06	4.00	33.99
Dec 3, 2005	19.5	5.12	5.19	3.16	3.33	4.37	39.73
Dec 6, 2005	17.7	6.68	7.05	4.93	4.40	4.23	40.49
Dec 9, 2005	25.6	7.72	9.01	6.60	7.10	7.62	38.92

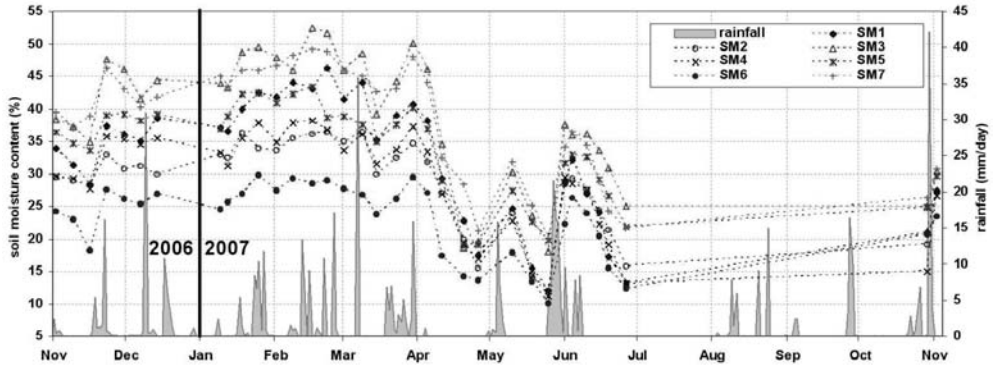


Figure 2. Temporal pattern of the field-mean soil moisture collected in the 7 sites.

Assuming the relationship between the field-mean soil moisture and the CV to be represented by an exponential law, it is possible to quantify the Number of Required Samples (NRS) in relation to the field-mean soil moisture and to the confidence level (for more details see Brocca et al., [2007a]). Figure 3a shows the NRS to obtain a value of the field-mean soil moisture within an error less than ±2%, at a 95% confidence level, for all investigated sites.

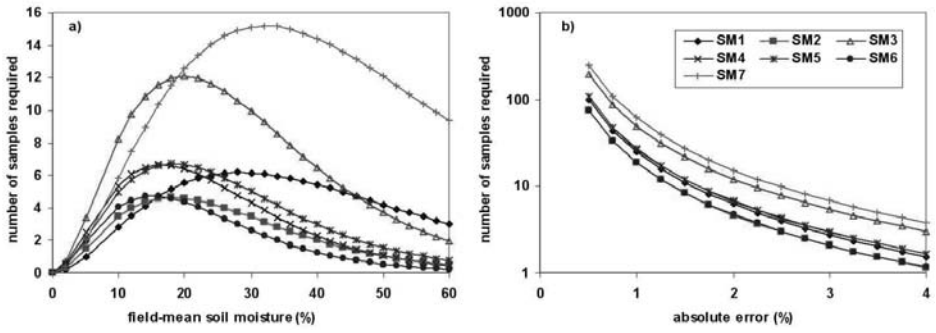


Figure 3. Number of soil moisture samples required to capture the field-mean soil moisture at 95% confidence level for: a) an absolute error of ±2 % and for different mean values; b) for the mean value requiring the maximum sampling and for different errors.

Similarly with other previous investigations [Famiglietti et al., 2007], the results show that for our study area the highest NRS is needed under almost dry conditions. In particular, to

assess accurately the field-mean soil moisture more resources are needed for SM7 and SM3 sites with a maximum NRS of 15 and 12, respectively. In Figure 3b, NRS is plotted against the absolute error for field-mean soil moisture value requiring the maximum sampling. For an accuracy ±4%, which is the value actually expected with remote sensing techniques, only from 1 to 4 measurements point are needed.

To investigate the NRS to obtain the catchment-mean soil moisture, the temporal stability analysis was conducted. In particular, the catchment-mean soil moisture is here assumed as the mean value of the whole dataset. Figure 4 shows the rank ordered mean relative difference, δ_i, with the standard deviation (vertical bar). As it can be seen, both the site SM1 and SM5 can be considered representative of the catchment-mean soil moisture as the corresponding δ_i was found nearly zero. Considering a single site, the determination coefficient (R^2) and the Root Mean Square Error (RMSE) were computed considering the linear regression between the catchment-mean soil moisture and the mean value obtained from different number of measurement points randomly selected in the SM1 site. As it can be seen in the inset of Figure 4 an accuracy of nearly 1% and a very high R^2 equal to 0.98 was obtained with 15 measurement points. Increasing the sample size the accuracy improves slightly.

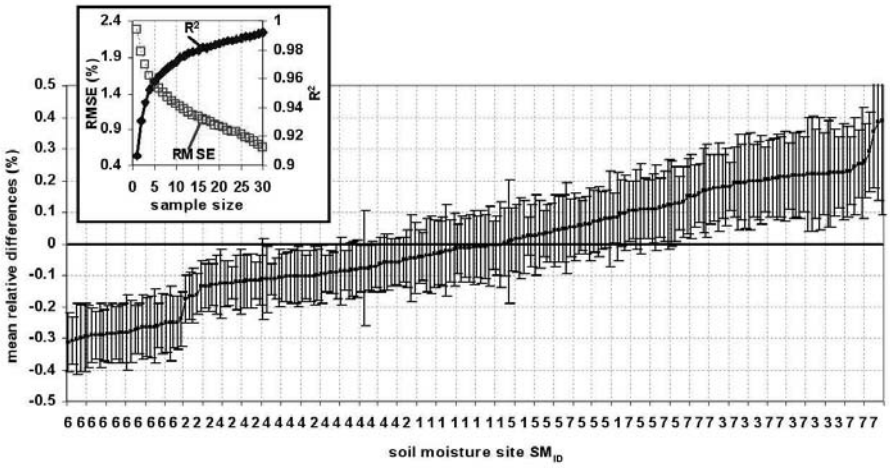


Figure 4. Rank ordered mean relative difference, δ_i, where error bars indicate ±1 standard deviation. The inset shows the Root Mean Square Error (RMSE) and the determination coefficient (R^2) between the catchment-mean soil moisture and the field-mean value obtained for site SM1 varying the number of point measurements.

5.2 "Indirect" method

The previous analysis has allowed to establish the number of measurements carried out on a representative site to get a reliable catchment-mean soil moisture. In this section, the relation between the soil moisture 'observed' at the experimental plot in the Colorso sub-basin and the AWC at the Niccone catchment scale is investigated. At the purpose, for each nested catchment the S_{obs} values were computed for the selected events and then normalized, in order to obtain a value between 0 and 1. In particular, the normalized S_{obs}, $\hat{S}_{obs}$, was given by:

$$\hat{S}_{obs} = 1 - \left(S_{obs} - 0.2S_{CN(I)}\right) / \left(1.2S_{CN(III)} - 0.2S_{CN(I)}\right) \quad (6)$$

where $S_{CN(I)}$ and $S_{CN(III)}$ are the S values corresponding to dry and wet Antecedent Moisture Conditions (AMC) as in the classical SCS-CN method. The $\hat{S}_{obs}$ was related to the 'observed' saturation degree, θ_e, given by the well-known relationship:

$$\theta_e = \left(\theta_i - \theta_r\right) / \left(\theta_s - \theta_r\right) \quad (7)$$

where θ_r and θ_s are the residual and saturated soil moisture, fixed as the minimum and maximum 'observed' values. Through eqs. (6) and (7) both quantities are varying between 0 and 1 and can be directly compared.

Figure 5a shows the $\hat{S}_{obs}$-θ_e relationship for all nested catchments. As it can be seen, the 'observed' saturation degree can be considered fairly accurate to address the catchment AWC. In particular, the regression performance decreases with increasing catchment area (see Figure 5b) but it can be still considered reliable for the largest catchment (137 km^2), with R^2 equal to 0.71. Considering a threshold value of 0.8 for the R^2 value, the 'observed' saturation degree in the plot can be considered appropriate to estimate AWC for catchments draining an area up to ~60 km^2. However, this result can be affected by the lumped approach used for relating rainfall and runoff depth. In fact, increasing the drainage area the spatial pattern of AWC, soil/land use characteristics, and rainfall becomes more and more significant in runoff generation.

It has to be noted that, for all catchments, the reliability of the 'observed' saturation degree for the AWC estimation was always better than other indicators commonly used for the AWC assessment, as the Antecedent Precipitation Index (API) or the BaseFlow Index (BFI) (not shown here for sake of brevity). Moreover, eq. (6) allows to take into account the variability of S from one catchment to another through the variability in the soil/land use characteristics as can be inferred from the $CN_{(II)}$ values (Table 1).

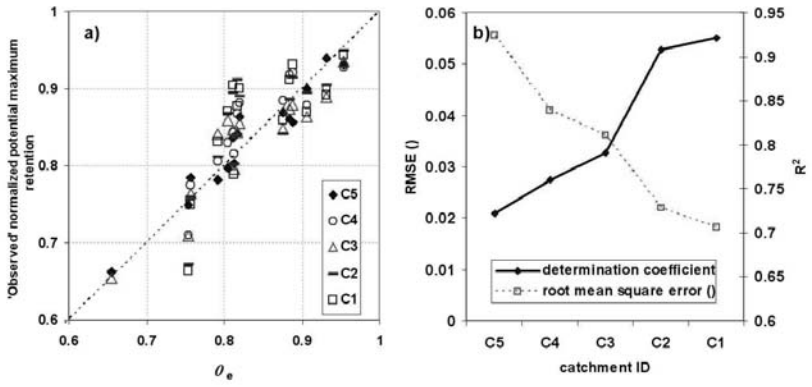


Figure 5. a) 'Observed' normalized potential maximum retention for the study catchments versus the 'observed' saturation degree, θ_e. b) Root Mean Square Error (RMSE) and determination coefficient (R^2) for the previous relationship and for each catchment.

6. CONCLUSIONS

Based on the analysis and the results reported in this study the soil moisture monitoring scheme for inland regions of Mediterranean area can be addressed. In particular: 1) the field-mean and catchment-mean soil moisture can be estimated with a good accuracy by a limited number of measurements, 2) soil moisture measurements carried out in a small area seem to be sufficient to obtain the wetness conditions of a catchment that can be effectively used for rainfall-runoff modelling. This aspect might have significant implications for flood prediction and forecasting and will be investigated in our future study. Finally, the widespread application of the SCS-CN method and the temporal stability of soil moisture observed in different studies make, in principle, the method proposed here reliable in other physiographic regions. However, a more detailed analysis is necessary to corroborate this insight.

ACKNOWLEDGEMENTS

The authors wish to thank B. Bani and R. Rosi for their assistance. Moreover, we would like to thank the students and colleagues who have worked for collecting soil moisture data. This work was funded by the National Research Council of Italy.

REFERENCES

Aubert, D., C. Loumagne, and L. Oudin, Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall runoff model. *Journal of Hydrology, 280*, 145–161, 2003.

Brocca, L., R. Morbidelli, F. Melone, and T. Moramarco, Soil moisture spatial variability in experimental areas of central Italy. *Journal of Hydrology, 333*, 356-373, 2007a.

Brocca, L., F. Melone, and T. Moramarco, Storm runoff estimation based on the soil conservation service – curve number method with soil moisture data assimilation., *Uncertainties in the 'monitoring-conceptualisation-modelling' sequence of catchment research,* L. Pfilster, L. Hoffmann Eds., IHP-VI, Technical Documents in Hydrology, 81 (UNESCO), 141-148, 2007b.

Brocca, L., F. Melone, and T. Moramarco, On the estimation of antecedent wetness condition in rainfall-runoff modelling. *Hydrological Processes*, 22, 629-642, 2008a.

Brocca, L., F. Melone, T. Moramarco, and R. Morbidelli, Soil moisture temporal stability over experimental areas of central Italy. *Geoderma, (submitted),* 2008b.

Bronstert, A., and A. Bardossy, The role of spatial variability of soil moisture for modelling surface runoff generation at the small catchment scale. *Hydrology and Earth System Sciences, 3*, 505-516, 1999

Famiglietti, J.S., J.W. Rudnicki, M. Rodell, Variability in surface soil moisture content along a hillslope transect: Rattlesnake Hill, Texas. *Journal of Hydrology, 210*, 259-281, 1998.

Famiglietti, J., D. Ryu, A.A. Berg, M. Rodell, and T.J. Jackson, Field observations of soil moisture variability across scales, *Water Resour. Res., 44,* W01423, doi:10.1029/2006WR005804, 2007.

Fitzjohn, C., J.L. Ternan, A.G. Williams, Soil moisture variability in a semi-arid gully catchment: implications for runoff and erosion control. *Catena, 32*, 55-70, 1998.

Goodrich, D.C., T.J. Schmugge, T.J. Jackson, C.L. Unkrich, T.O. Keefer, R. Parry, L.B. Bach, and S.A. Amer, Runoff simulation sensitivity to remotely sensed initial soil water content. *Water Resources Research*, *30(5)*, 1393-1406, doi: 10.1029/93WR03083, 1994.

Grayson, R.B., G. Bloschl, and I.D. Moore, Distributed parameter hydrologic modelling using vector elevation data: THALES and TAPES-C. *Computer Models of Watershed Hydrology*, Singh, V.P. Ed., Water Resources Publications, Highlands Ranch, Colorado, 669–696, 1985.

Grayson, R.B., A.W. Western, H.S. Chiew, and G. Blöschl, Preferred states in spatial soil moisture pattern: Local and nonlocal controls. *Water Resources Research, 33*, 2897-2908, 1997.

Grayson, R.B., and A.W. Western, Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *Journal of Hydrology, 207*, 68-82, 1998.

Longobardi, A., P. Villani, R.B. Grayson, and A.W. Western, On the relationship between runoff coefficient and catchment initial conditions. *Proc., MODSIM 2003 International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand Inc., Townsville, Australia, 2, 867-872, 2003.

Merz, R., and E.J. Plate, An analysis of the effects of spatial variability of soil and soil moisture on runoff. *Water Resources Research, 33(12)*, 2909-2922, doi: 10.1029/97WR02204, 1997.

Ponce, V.M. and R.H. Hawkins, Runoff curve number: Has it reached maturity? *Journal of Hydrologic Engineering, 1*, 11-19, 1996.

Vachaud, G., A. Passerat de Silans, P. Balabanis, and M. Vauclin, Temporal stability of spatially measured soil water probability density function. *Soil Science Society of America Journal, 49*, 822–828, 1985.

Western, A.W., S. Zhou, R.B. Grayson, T.A. McMahon, G. Blöschl, and D.J. Wilson, Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology, 286*, 113-134, 2004.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Radar-based surface soil moisture retrieval over agricultural used sites – A multi-sensor approach

M. Pause[a], **M. Volk**[a], **K. Schulz**[b]

[a] *Helmholtz Centre for Environmental Research – UFZ, , Department of Computational Landscape Ecology, Permoserstraße 15, D-04318 Leipzig, Germany, (marion.pause@ufz.de, martin.volk@ufz.de)*

[b] *Ludwig Maximilian University of Munich, Department of Geography, Chair for Geography and Geographical Remote Sensing, Luisenstraße 37, D-80333 München, (k.schulz@lmu.de)*

Abstract: Spatial distributed information of isochronal surface soil moisture is very important to compensate the inaccuracy of initial conditions and the uncertainty of parameters in hydrological models at the landscape scale. In this paper the conceptual procedure to derive spatial distributed surface soil moisture values from synthetic aperture radar data by using ancillary optical remote sensing data is presented. Different biophysical vegetation parameters like vegetation water content and leaf area index overlay the soil moisture information on the microwave signal and hamper the application of many existing models. Therefore the objective of the proposed study is to test the performance of artificial neural networks to extract soil moisture information from radar data. Multi- and hyperspectral remote sensing data provide spatial distributed information about above ground vegetation parameters and can thereby used as ancillary network input to support the soil moisture extraction. The results of the study are expected to provide an improved database (initial conditions, plant parameters, etc.) for hydrological models.

Keywords: Surface Soil Moisture; Remote Sensing; Artificial Neural Network

1. INTRODUCTION

Synthetic Aperture Radar (SAR) data from satellite may provide robust estimates of spatially distributed surface soil moisture (SSM) values for different applications of distributed hydrological modelling in catchment areas. It has been shown that SSM can be derived with a root mean square error of 3-7 vol.% [Loew et al., 2006; Wagner et al., 2007]. The spatial resolution of information derived from remote sensing data is often similar to or even finer than the scale of distributed model elements; and in many images it is easy to detect spatial patterns visually.

However, especially the vegetation cover and its spatial and temporal dynamics represent a problematic factor for the useful applicability of the most existing methods or models. The scattering behaviour of the radar signal from vegetated areas is thereby not only controlled by the soil dielectric properties but also by the dielectric properties of the vegetation and its geometric structure (leaves, stems, fruits). Over the last three decades fundamental research and many case studies have been performed to handle the active microwave signal in terms of SSM retrieval. Therefore, several different kinds of modelling techniques were performed. Theoretical models like the Kirchhoff Model as well as the prevalently used Integral Equation Model (IEM) [Fung et al., 1992] are electromagnetic backscattering

models and describe the radar backscattering as a function of sensor configurations and surface characteristics e.g. soil roughness or soil moisture. The parameterization of these models is very difficult and thus limited for operational applications on large areas. Simultaneously many empirical and semi-empirical models show in fact good results but must be explicitly designed for the research area. Furthermore, these methods require a large ground truth database to adopt simple linear or nonlinear regression analyses [Attema and Ulaby, 1978; Loew et al., 2006; Quesney et al., 2000; Zribi and Dechambre 2003] for a reliable model performance. An alternative solution that hasn't been applied that much for the retrieval of radar-based SSM is the application of an Artificial Neural Network (ANN). ANN techniques are readily available and have the ability to learn the relationships between input and output pairs given in a training phase. After the training, the network can be applied to new sets of input data. The demand on training data is in the most cases extensive and similar to empirical approaches. Hence, the presented study aims at the integration of appropriate other available remote sensing data to provide input information about the vegetation cover.

Based on the described background this paper outlines an ANN based SSM modelling on the basis of SAR data and ancillary optical remote sensing information. In this connection a moderate number of in-situ soil moisture values as well as in-situ vegetation parameters will be used to calibrate several remote sensing based vegetation indices and the network itself. Correlations of remote sensing data and many different bio- and geophysical parameters are in the most cases nonlinear. This is the reason why neural networks are inherently suitable for addressing remote sensing based vegetation and SSM information retrieval especially for input data collections with very different natural background and units.

The primary objectives of this presented study are focused to a) find the optimal set of remote sensing based indices to derive the leaf area index (LAI) and vegetation water content (VWC) information for different cereals; b) describe a moderate way for the ground truth sampling of SSM and vegetation parameters; and c) discuss the application constraints of radar based SSM retrieval.

2. MATERIAL AND METHODS

The data, used for the study, consist of space born multispectral and horizontal-horizontal as well as vertical-vertical polarized SAR data. For C-band SAR information ENVISAT-ASAR (Environmental Satellite-Advanced Synthetic Aperture Radar) Alternating Polarization Mode Products obtained by ESA EO Cat-1 Proposal 5086 are applied. SAR X-band data are provided by the new German TerraSAR-X satellite by Proposal HYD0315 through the German Aerospace Centre (DLR). The multispectral information will be obtained from the ASTER instrument on the NASA satellite Terra. Additionally it is planned to obtain further spectral data from the air born hyperspectral sensor AISA by an own flight campaign. The ground truth campaign will take place near the acquisition time of the ASAR data during spring and early summer 2008.

2.1 Remote Sensing based vegetation parameter estimation

Subsequently, the retrieval of vegetation parameters from SAR data and optical remote sensing data will be described. LAI and VWC are one of the most important parameters for vegetation canopy characterisation and can be related to different spectral vegetation indices (VI) or the radar backscatter coefficient. Normally multi- or even hyperspectral band information is reduced to a single numerical index on the basis of some spectral intervals. Generally, different VI must be analysed to find the most appropriate one, in terms of phenological stage and soil signal fraction, for simple linear or nonlinear regression modelling of LAI and VWC. Correlation analyses will be applied to quantify the relationship between SAR backscatter coefficient and vegetation parameters.

SAR based biomass information

The SAR C-band data can be used as SSM reference data set [Ulaby et al., 1996]. To provide radar based information about the above ground vegetation layer X- band SAR data can be applied. Therefore, the relation of co-polarisation ratios and the biomass respectively the vegetation structure (e.g. LAI) and VWC itself as well as their temporal change during the growing season can be analysed focused on biomass quantifications on the C-band signal.

LAI information

Spectral vegetation indices are designed to measure the greenness in which the amount of green leaves and the chlorophyll content plays a major role. Because of differences in canopy density and plant structure itself during the growing season it is obvious that different VI need to be taken into account.

For LAI estimations different VI derived from the multi- and hyperspectral data, e.g. like the Ratio Vegetation Index (RVI), the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI) or the Modified Soil Adjusted Vegetation Index (MSAVI) and Modified Soil Adjusted Vegetation Index 2 (MSAVI2), can be analysed focused on their applicability to establish simple linear and nonlinear empirical relationships to model spatial distributed LAI values.

Vegetation water content information

Vegetation water content has an important influence on the microwave signal especially at C- and X-band frequencies in terms of scattering and attenuation. By estimating the VWC the accuracy of the final SSM value is more accurate [Jackson et al. 2004, Anderson et al. 2004]. Therefore, different spectral remote sensing indices can be calculated from the ASTER and AISA data to obtain spatial distributed VWC estimations. A performance test to define the most proper indices will at least analyse the Normalized Difference Water Index (NDWI), the Moisture Stress Index (MSI) and the Equivalent Water Thickness of the canopy (EWT_{canopy}).

2.2 Ground Truth

The ground truth observations will be performed on three testsites in the south of the Parthe catchment in south-eastern Germany. Land management in the catchment is dominated by crop rotations such as winter barley, winter rye and winter rape. Presently five field campaigns in the period of March and June 2008 are planned. Table 1 gives an overview of different sampling variables and the time schedule. Due to the high diurnal dynamics of surface soil water content the corresponding SSM field measurements will take place in three-hours-periods around the ENVISAT ASAR overpass. Measurements will be carried out within the upper 6 cm of the soil surface by using mobile TDR probes. An average of five single measurements will be taken as representative value every 50 m along two or three profiles per testsite (see Figure 1). The profiles are parallel to the machine tracks on the field to ensure the accessibility during later campaigns in the growing cycle. To study the influence of SSM heterogeneity inside an ENVISAT ASAR pixel an extensive in-situ SSM sampling will take place in a defined grid (60m x 60m) on the winter rye field. Therefore 49 additional sampling points will be observed.

Table 1. Time table for in-field sampling

Variable	Time Window around ENVISAT overpass
Vegetation and Land Cover	1 day
- Phenological stage	
- Plant height	
- Stand density	
- Leaf Area Index	
- Biomass samples, to get fresh and dry weight of stems, leaves and ears	
- Spectral Emission	1 week
Soil Moisture	3 hours
Soil and Surface Temperature	3 hours
Surface Roughness	before campaign starts
- Row spacing, row direction	

Field LAI measurements will be performed by using a LI-COR, Inc. (Lincoln, Nebraska, USA) LAI-2000 Plant Canopy Analyzer. A destructive biomass sampling is scheduled to retrieve vegetation water content. Several single plants from different parts of each field will therefore be separated into stems, leaves and ears focused to a) analyse their single influence on the C- and X-band SAR signal and b) to obtain the EWT_{canopy} on the basis of the leave water content information. In the framework of this study the data acquisition from the hyperspectral sensor AISA (Airborne Imaging Spectrometer for Applications) [Mäkisara et al., 1993] shall be tested from a microlight aircraft. For that purpose field spectrometer measurements will be performed for the validation of several hyperspectral indices.

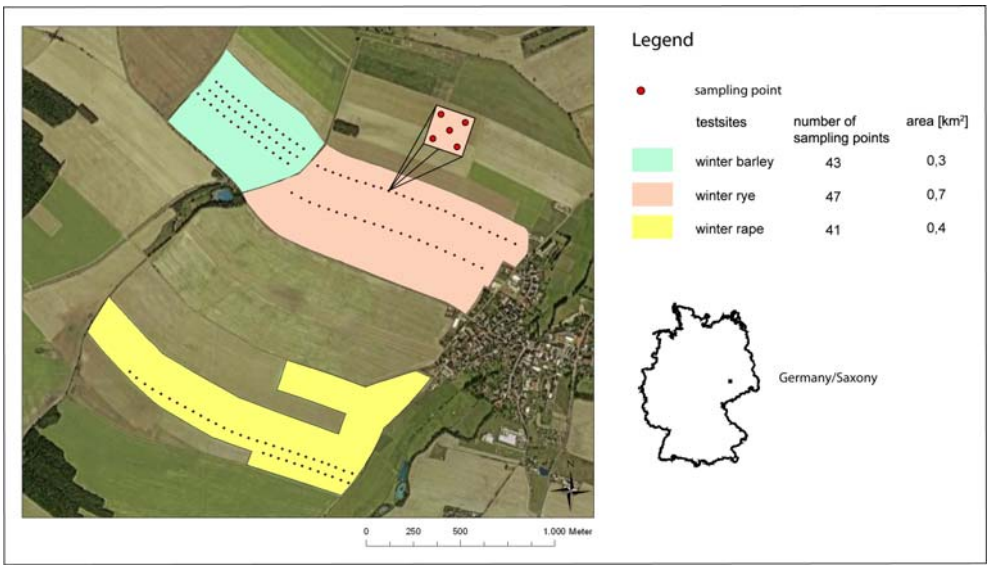


Figure 1. Testsites and schematic demonstration of the SSM sampling strategy (map source: Google Earth)

2.3 Methodology

All ASAR and TerraSAR-X data sets will pass an automated calibration and geocoding procedure. The rectification will include a digital elevation model to compensate terrain induced distortions. The temporal and spatial variations of the radar backscatter coefficient will be regarded as an influence of soil moisture and vegetation parameters in the case of identical sensor properties (incidence angle, acquisition track). The soil surface roughness will be assumed to be constant due to the fact that winter cereals are cultivated on it and the soil surface remains untilled from sowing during the experiment period until harvest.

Inferring SSM values from the C-band data among using other remotely sensed ancillary information implies that a functional relationship of the different vegetation parameters, the radar backscatter coefficient and the SSM itself must be made. The fact of the dynamics and complexity of many different environmental parameters (e.g. soil moisture, vegetation water content, plant geometry, plant interception) which are interacting to the signal formation limits a mathematical representation and its actual further application, respectively.

Artificial Neural Network training algorithms attempt to find the best nonlinear approximation based on a defined network complexity and structure without the constraint of linearity or pre-defined nonlinear relationships. A supervised network map inputs to the desired outputs by learning the mathematical function underlying the system. With this method it is possible to relate inputs and outputs without any assumptions about the mathematical relations. It is also a proven method to find a proper set of input parameters out of a large number to perform an essential modelling. Figure 2 shows the fundamental NN structure consisting of a M-dimensional input layer, a hidden layer and one output.

In the framework of this study a feed-forward multilayer perceptron (MLP) with some hidden layers of neurons between the input and output is designated. To train the weights of the network the backpropagation learning algorithm will be applied and a sigmoid activation function to introduce nonlinearity will be used. For example Brogioni et al. (2007) presented reasonable ANN based soil moisture results with a regression coefficient of R^2 = 0,82 by using only the C-band backscattering as input. An improvement to R^2 = 0,91 was obtained by using the backscattering and an ancillary soil surface roughness information.

To train the networks a set of ground truth vectors (GTV) will be created on the basis of the measured data for each testsite. The different inputs are indices adapted from direct ground truth measurements and the radar backscatter coefficients itself, the corresponding output is an aggregated in-situ SSM value. A second and only remotely sensed dataset for applying the learnt rules contains all input vectors where no direct ground truth information for the different indices exists.

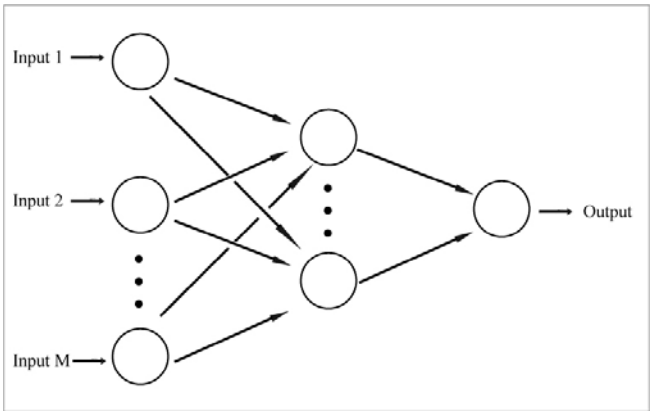


Figure 2. Principle Neural Network Structure

3. RESULTS

This paper introduces to a multi-sensor based method to derive spatial distributed SSM values through the application of an ANN and is currently under development. Results in form of modelled SSM or analysed empirical relationships of remote sensing data and different biophysical vegetation parameters can not be given yet, because the acquisition of necessary data is planned for the next months.

4. DISCUSSION

During the last three decades, many studies were performed to retrieve bio- and geophysical parameters from optical and active microwave remote sensing data. Significant correlations were carried out that provide important input for environmental modelling applications.

Surface soil moisture is highly variable in space and time and plays an important role for many hydrological processes (e.g. infiltration, surface runoff, evaporation) occurring at the land-atmosphere boundary. By proper selection of sensor parameters SAR C-band data mainly depend on SSM even on vegetation covered soils while the attenuation and vegetation contribution itself are larger observed in X- band SAR data [Fung und Ulaby, 1978; Ulaby and Wilson, 1985]. Multispectral and especially hyperspectral data may provide detailed information about vegetation phenological state and condition (e.g. VWC) which is important to quantify the effect of vegetation on the active microwave signal. Hence, the ancillary use of optical remote sensing data encourages radar- based SSM retrieval with a focus on moderate ground truth effort. Unfortunately, the availability of optical images is often limited through cloud cover. Therefore, the potential of X- band SAR data to quantify the effect of vegetation promises a further ancillary possibility to quantify the SSM information on the C-band signal. In general microwave systems, like radiometer and SAR systems, are quite reliable sources for remote sensing data, which is caused by their weather independency. Occurring problems are in the most cases limited to user conflicts by ordering the SAR data in a special sensor mode.

Regarding the complexity of radar signals the problem is that the final backscatter coefficient can be the result of a variety of corresponding conditions and the quantification of the several parameters is strongly hampered through its nonlinear nature. However, neural networks provide the potential to handle complex nonlinear systems and need to get studied and applied in future remote sensing applications especially with focus in minimizing the ground truth effort.

5. CONCLUSION

SSM and their spatial variability play a key role for many hydrological processes and are a major factor in watershed science and modelling. Satellite SSM products are an important component for long term observations of the earth surface and provide high potential to perform that task. However, there is a need to improve SSM retrieval from spaceborn SAR in respect of reliability and operational applications. The aim of the outlined project is to prepare an optimal processing chain for multi-sensor (optical and SAR) data and indices to quantify the vegetation content on the C-band SAR signal and to built up a NN based SSM retrieval procedure.

6. REFERENCES

Anderson, M.C., C.M.U. Neale, F. Li, J.M. Norman, W.P. Kustas, H. Jayanthi and J. Chavez, Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery, *Remote Sensing of Environment, 92*(4), 447−464, 2004.

Attema, E., and F. Ulaby, Vegetation modeled as a water cloud, *Radio Science, 13*, 357-364, 1978.

Brogioni, M., S. Paloscia, P. Pampaloni, S. Pettinato and E. Santi, Soil moisture maps of agricultural fields in Northern Italy from ENVISAT/ASAR images, 5th International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, Italy, 2007.

Fung, A.K., Z. Li, and K.S. Chen, Backscattering from a randomly rough dielectric surface, *IEEE Transactions on Geoscience and Remote Sensing, 30*(2), 356-369, 1992.

Fung, A.K. and F. Ulaby, A scatter model for leafy vegetation, *IEEE Transactions on Geoscience and Remote Sensing, GE-16*(4), 281-286, 1978.

Jackson, T.J., D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy and R. Hunt, Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans, *Remote Sensing of Environment, 92*(4), 475−482, 2004.

Loew, A., R. Ludwig and W. Mauser, Derivation of Surface Soil Moisture From ENVISAT ASAR Wide Swath and Image Mode Data in Agricultural Areas, *IEEE Transactions on Geoscience and Remote Sensing, 44*(4), 889-899, 2006.

Mäkisara, K., M. Meinander, M. Rantasuo, J. Okkonen, M Aikio, K. Sipola, P. Pylkkö and B. Braam, Airborne Imaging Spectrometer for Applications (AISA), Digest of IGARSS'93, 2, 479– 481 (Tokyo, Japan), 1993.

Quesney, A., S. Le Hegarat-Mascle, O. Taconet, D. Vidal-Madjar, J.P. Wigneron, C. Loumagne and M. Normand, Estimation of Watershed Soil Moisture Index from ERS/SAR Data, *Remote Sensing of Environment, 72*(3), 290-303, 2000.

Ulaby, F., P.C. Dubois and J. Van Zyl, Radar mapping of surface soil moisture, *Journal of Hydrology, 184*, 57-84, 1996.

Ulaby, F., and E. Wilson, Microwave attenuation properties of vegetation canopies, *IEEE Transactions on Geoscience and Remote Sensing, GE-23*(5), 746-753, 1985.

Wagner, W., G. Blöschl, P. Pampaloni, J.C. Calvet, B. Bizzarri, J.P. Wigneron and Y. Kerr, Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, *Nordic Hydrology, 38*(1), 1-20, 2007.

Zribi, M., and M. Dechambre, A new empirical model to retrieve soil moisture and roughness from C-band radar data, *Remote Sensing of Environment, 84*(1), 42-52, 2003.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Overland flow and soil erosion in an undisturbed Brazilian Northeastern Semiarid Experimental Plot

L. F. F. Moreira[a], F. de O. Silva[a], A. M. Righetto[a] and V. M. de A. Medeiros[a]

Sanitation Engineering Post-graduation Program, Federal University of Rio Grande do Norte, Campus Universitário, SN, 59082-970, Natal-RN, Brazil. (lucio@ct.ufrn.br; flaviane@semurb.natal.gov.br; righetto@ct.ufrn.br; victormoises@bol.com.br)

Abstract: In the Brazilian semi arid region, a combination between human activities and climatic factors produces an increase in the ecosystem susceptibility to soil erosion. In this study, runoff and soil erosion processes due to natural rainstorm events on a 0.025 ha undisturbed experimental plot is analyzed. The site was instrumented to monitor rainfall, overland flow, runoff and erosion. Automatic devices were installed to monitor rainfall and the level of a 5 m^3 tank at a 1-minute time step. The sediment transported by the flow was collected after each event. Plot average slope is 2.5%, vegetated of both permanent and annual plants, composed by a combination of xerophytes shrubs, brushes and grass species, whose patterns change with growing season as a result of its natural development dynamics. It was observed that runoff and soil erosion was controlled by rainfall pattern, antecedent soil moisture and vegetation. Also, vegetation height and density is influenced by soil structure and faunal activity, increasing both soil storage capacity and surface roughness. Wetting up period data (February 13 to April 25/2007) showed that runoff generation is the result of a non-linear interdependency of rainfall pattern, soil wetness condition and vegetation characteristics. Runoff coefficients less than 0.1 occurred at approximately 82 per cent of the observed events. Moreover, a positive feedback was observed involving vegetative species dynamics, soil properties and faunal activity producing annual plants regeneration. Also, annual plants distribution within the plot seemed to be controlled by permanent trees canopy, where density was higher. Finally, it was obtained an empirical relationship between short-term sediment production and rainstorm amount valid for 2006-2007 wetting up period individual rainstorm events. Hence, this approach has provided a continuous record of rainfall-runoff-sediment production field monitoring data which can be used to analyze hydrologic and vegetative dynamics at plot scale. Also, it highlights the role of vegetative cover as a controlling factor on erosion protection management.

Keywords: Semiarid, Undisturbed plot, Erosion, Runoff.

1. INTRODUCTION

Native vegetation in the Brazilian semi arid region has been reduced due to inadequate grazing activities and woodland removal for domestic use and ceramic industries. In upland areas, natural vegetation can be considered as the main controlling factor on runoff generation and soil erosion, which confirms previous studies (Kosmas *et al.*, 1997; Nicolau *et al.*, 1996; Bergkamp, 1998). Conversely, extreme rainstorm events in this region may increase the susceptibility of the ecosystem to soil erosion induced by human activities (Wainwright, 1996; Reid *et al.*, 1998).

Historically, the lack of annual rainfall distribution uniformity has motivated the construction of tenths of small to medium sized reservoirs used in supplying human and

grazing demands (Cadier, 1996). Currently, there has been a need of assessing the impact of human activities on soil degradation and the infilling of these reservoirs.

Runoff generation and soil erosion processes in semi arid environments are complex and strongly non-linear. Field monitoring using an experimental plot allows a detailed analysis of the mechanisms in a small area and helps to analyze small scale processes (Bracken and Kirkby, 2005; Parsons *et al.*, 1997). Since 1981, hydrologic field studies have been conducted with the aim of analyze the influence of vegetation and crop management on runoff and soil erosion processes in Brazilian semi arid region (Cadier et al., 1983; Molinier et al., 1989; Srinivasan et al, 2003).

In general, surface runoff generation on tropical semi arid environments is the result of a combination between rainfall-excess and saturation-excess mechanisms. Natural vegetation dynamics during the wetting up season have an important effect on infiltration and interception; annual species density increase enhances both soil infiltration and soil water storage capacity. Also, litter and organic matter accumulated on soil surface both protect against raindrop impact and increase storage water due to its strong moisture holding capacity (Boer and Puigdefábregas, 2005).

This paper was based on fieldwork data. Obtained results were provided by continuous record of rainfall-runoff-sediment production field monitoring data used to analyze hydrologic and vegetative dynamics at plot scale. Also, in a context of natural resources mismanagement it highlights the role of vegetative cover as a controlling factor on erosion protection management. Hence, this approach analyzed runoff and soil erosion processes by using 2006-2007 wetting up period monitoring data on an upland undisturbed representative experimental plot established in Brazilian semi arid region.

2. STUDY AREA

The study area is located 300 km far western from the city of Natal, state of Rio Grande do Norte-Brazil. Plot area was established in Serra Negra do Norte experimental watershed, a 2.108 km^2 monitored catchment (coordinates 6°34'42" S; 37°15'56" W). Further details on the experimental watershed are shown in Moreira et al. (2006). The mean annual temperature is approximately 28˚C. Topographic contour lines and vegetation species at the end of the rainy season on the plot area are shown in Fig. 1. The drainage area of the SNN experimental plot is 0.025 Ha (5 m x 50 m), and its elevation is 205 m above MSL. Although the plot micro relief varies with the dominant flow axis, its average slope is 2.5%. Natural vegetation features such as height and density increased dynamically as a function of time approximately ten days after the beginning of the wetting up period.

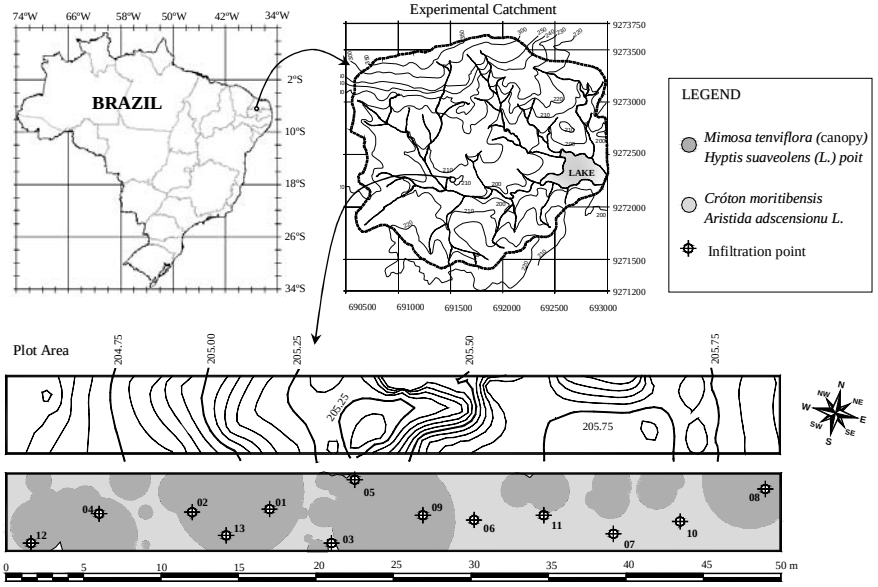


Figure 1. Study area and plot configuration, topographic contour lines, vegetative species.

In contrast with the beginning of the wetting up period, the plot was largely vegetated with annual plants, indicating the influence of rainfall regime, soil moisture and the presence of a large soil seed bank available to promote germination. At the end of the rainy period, approximately 96% of the canopy cover was composed by a combination of xerophytes shrubs, brushes and grass species, the most representative being *Mimosa tenviflora, Cróton moritibensis, Aristida adscensionu L.* and *Hyptis suaveolens.* Vegetation patterns changed with growing season as a result of its natural development dynamics. During the drought period, vegetation is prone to local collapses due to a combination of both long dryness spells and high evapotranspiration. The end of the drought season coincides with a high water stress level, and permanent species, *Mimosa tenviflora,* cover approximately 20% of the soil surface, the remainder covered with accumulated litter (seeds, leaves and wood). The soils in the SNN plot site are well drained, compacted and shallow, with depth varying from 0 to 140 cm, poorly sorted bimodal chromic luvisoils (fine and coarse modes present in the bulk sample), with approximately 20% rock and gravels at surface cover. Eleven soil samples were collected from the top 10 cm and taken to the laboratory for size distribution analysis. Soil texture is gravelly sandy loam with approximately 15.6% gravel, 52.2% sand, 26% silt and 7.2% clay. Organic matter and litter accumulates on soil surface as a result of an intense vegetation biological dynamics. The average annual rainfall over the past 11 years (1995-2005) is 689 mm. Precipitation is highly variable from year to year in the area. Approximately 52 per cent of the precipitation is high intensity thunderstorms of limited area extent. A previous study has shown the influence of high spatiotemporal rainfall heterogeneity on runoff in the watershed area (Moreira et al., 2006). Monthly rainfall shows considerable variation during the rainy season, especially in the period January-May. Daily rainfall data statistical analysis revealed that approximately 25% of the annual depth occurs during the maximum daily precipitation, which may give rise to erosion processes of high magnitude.

3. FIELD RUNOFF AND EROSION MONITORING

The experimental plot was defined considering representative mosaic areas within the catchment. It was established by a 0.2 m high surrounding cement wall. A fence was installed surrounding the plot for protection from grazing activity. At the plot downstream end, a 5 m^3 volumetric tank was built to monitor water and sediment discharge and organic matter for each storm event. A sediment trap structure, dimensions 0.15 m x 0.15 m x 2 m (wide-depth-length) was built at the tank bottom, where an inclined surface promotes sedimentation during the storm event. Tank water level during the storm was monitored by using a water level logger (OTT *Thalimedes*) adapted into a 150 cm high PVC tube of 6" diameter, which recorded water level data at 1-minute interval. A rainfall automatic recorder (*Scientific Campbell*) was installed adjacent to the plot. Rainfall data were recorded at 5 and 1-minute intervals on 2006 and 2007 respectively. During the rainy period, the tank was emptied manually after each storm by using a portable pump, when all the sediment and particulate organic matter was collected and taken to the laboratory, where sediment was dried, weighted and sampled for grain size analysis. Also, organic matter was dried, weighted, identified and biologically classified. For each storm event, runoff was obtained by applying hydrologic budget including runoff, rainfall and the tank water level variation during the storm. Topography surface survey was performed on regular grid of 200 points by using an electronic GPS device.

4. EMPIRICAL OBSERVATIONS

Tropical semiarid regions are strongly influenced by rainfall distribution and soil water availability. Rainfall controls biological activity, such as seed germination and vegetative regeneration. On the beginning of the wetting up period (approximately 15 days), sheet erosion prevails and controls topsoil particles detachment caused by rainstorm. Located at the bottom of a hillslope, plot area is characterized by a high proportion of stoniness and bare soil areas. Overland flow down the hillslope washes the topsoil. A mixture of water and solid particles, including insects, seeds and foliage is produced. Although a

considerable seed bank proportion has been washed by the sheet flow, native annual plants germination after a sequence of rainfall events was widespread. Field observations revealed an accelerated increase of annual plants density as a function of time. This pattern was probably caused by differences in seed germination and topsoil moisture maintenance. Furthermore, it was observed the establishment of annual plants communities, as can be seen in Figure 1. Indeed, it helps to explain the patterns of vegetation spatial distribution. Also, vegetation and biological activity enhanced topsoil water storage capacity through a positive feedback. Therefore, increase on vegetation density caused a considerable decrease of overland flow and sediment production, which confirms previous studies (Ludwig et al., 2005). Fig. 2 shows observed hydrographs of four events on plot area. Generally, they have steep rising and recession limbs, with more than one peak values. Hydrograph behavior is the result of a combination of factors such as storm characteristics, previous rainfall, soil wetness and vegetation cover (Kidron and Yair, 1997). It can be seen on 14 February 2007 hydrograph a pulsating behavior with a quick response. Although rainfall was moderate, soil runoff reflected the previous rainfall height of 101 mm. On 19 February 2007 hydrograph values reflect both previous 3 days rainfall height of 102 mm and rainfall pattern.

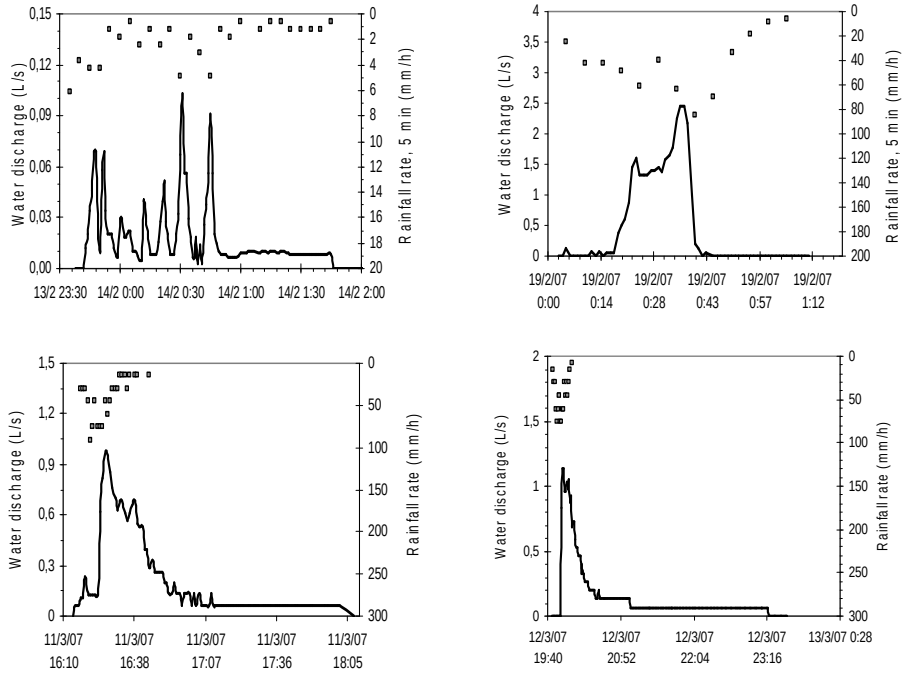


Figure 2. Observed hydrographs of four different rainstorm events on plot area.

Figure 2(b) shows that although rainfall event maintained after peak discharge, recession limb decreased rapidly. On Figures 2(c) and 2(d), short duration and high magnitude events produced hydrographs with steep rising limbs and recession limbs of longer duration. A combination of rainfall patterns, antecedent soil moisture and vegetation control runoff and soil erosion (Kidron and Yair, 1997). The occurrence of a considerable increase on soil moisture between high magnitude events on 11-12 March 2007 was reflected on hydrograph behavior. Rainstorm intensity exceeded soil infiltration capacity in some less permeable areas, resulting in depressions filling and progressive overflowing to adjacent depressions. For these events, a fully connected overland flow over a saturated surface probably minimized surface micro-relief effect (Darboux et al., 2001). Figure 3 shows two photographs that reveal the difference on plot characteristics during the drought and the rainy seasons. The first one was taken on 03 February and the other one on 28 June 2007.



Figure 3. Plot characteristics during the drought (a) and the rainy (b) seasons.

5. RUNOFF GENERATION MECHANISMS

Runoff generation in semiarid areas is a process influenced by a non-linear and complex interdependency between rainfall patterns, soil wetness condition and vegetation characteristics (Beven, 2002). Runoff coefficient, rainfall height and maximum rainstorm intensity behavior as a function of time is shown on the graph of Fig. 3 for the wetting up period between 13 February and 25 April 2007. It can be observed that approximately 82 per cent of the rainfall events produced flow rate values less than 0.1. The remaining flow rate values, which varied between 0.16-0.28, were influenced by a combination of a sequence of antecedent rainfall events and high intensities. The highest flow rate value 0.49 was observed on 12 March and is the result of a 3-days sequence of high intensity rainfall events, and was probably due to a considerable increase on soil wetness. It was observed that although only 26 per cent of the rainfall events were higher than 20 mm, maximum rainfall intensities varied within a broader range. Moreover, most of the rainfall events occurred during the first 10-days period, with heights varying between 29-101 mm. It was observed a sequence of flow rate values less than 0.05 after 12 March. From this time on, vegetation and biological activity were the main controlling factors on runoff generation and sediment production. Flow rate values reflected a positive feedback of vegetation and infiltration on soil storage capacity. Vegetation cover protect soil surface from splash, increase surface roughness and enhances soil structure and macroporosity. Generally, the graph shows that runoff generation is mainly controlled by soil antecedent wetness and vegetation characteristics.

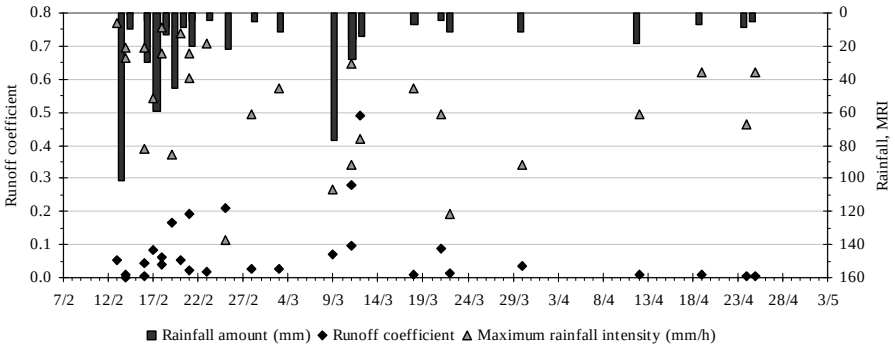


Figure 3. Runoff coefficient, rainstorm amount and maximum rainstorm intensity behavior for the 2006-2007 wetting up period.

6. INFILTRATION FIELD EXPERIMENTS

Previous studies outlined soil infiltration capacity as a factor of crucial importance on runoff and soil erosion mechanisms. Since water stress controls seed germination and plants regeneration, infiltration increases soil water availability for the plants. Field

infiltration experiments were performed at 13 points randomly distributed within the study area by using a constant head permeameter (Moreira et al., 2002). The infiltration experiments were conducted between 3 February and 5 May 2007. Measured infiltration rates were used to fit the Horton (1933) empirical equation, described as follows,

$$f(t) = (f_o - f_c)\exp\{-kt\} + f_c \qquad (1)$$

where f(t) is infiltration capacity at time t, f_o is an initial infiltration rate, f_c is a final infiltration capacity and k is an empirical constant. Horton parameters were fitted for each point within the study area and are shown on Table 1. Field infiltration results show that final infiltration capacity f_c had a great local dispersion, with a coefficient of variation of 1.2. This demonstrates the effect of spatial heterogeneity on soil hydraulic characteristics and confirms previous studies on infiltration rates in semiarid areas. Moreover, highest infiltration parameters were observed in areas located under the *Mimosa tenviflora* specie permanent tree canopy cover. Particularly in these sites, annual plants density was high, indicating the influence of both tree protection from radiation and soil moisture. Nevertheless, a combination of soil properties, vegetation dynamics and faunal activity controls runoff generation within the study area. According to Dune and Dietrich (1980), this mechanism is applicable for semiarid conditions, where plants intercept rainstorm through its branches and stems directly to the soil surface and canalize it to the roots.

Table 1. Horton parameters obtained from field infiltration experiments within plot area.

Infiltration Point	f_c (mm h^{-1})	f_o (mm h^{-1})
1	57	100
2	4	50
3	3	30
4	50	100
5	6	60
6	10	30
7	13	50
8	5	25
9	195	300
10	85	200
11	120	200
12	10	50
13	170	300

7. SOIL EROSION MECHANISMS

Sediment production in the study area showed considerable variation as a function of time. During the drought period, radiation and high temperatures produce soil surface degradation, which results in accumulated sediment available to transport. In fact, the first high magnitude rainfall events 13-16 February 2007 showed that sediment production occurred mainly as sheet erosion, which explains the existence of many large pebbles at the soil surface. Also, high amounts of seeds and organic matter were transported by overland flow, although it didn't seem to affect vegetative species germination and replenishment during the wetting up period (Gallart et al., 2002). The increase in soil water moisture controls both annual plants germination and growth. Field observations pointed that the increasing density of annual plants as a function of time increased infiltration and reduced both runoff and sediment yield. Runoff and sediment are intercepted by these patchy plants as it comes from upland areas. Organic matter was made up of washed seeds, leaves, faunal microbiotic and small roots. Once the first high magnitude rainfall events have occurred

and annual plants seeds germinated, sediment production and organic matter decreased considerably. An empirical relationship between sediment production and rainstorm amount was statistically adjusted for the 2007 wetting up period is shown on the graph of Fig. 4. Two points were rejected (triangles). The reason was that they seem to represent limited source conditions in the sediment mobilization.

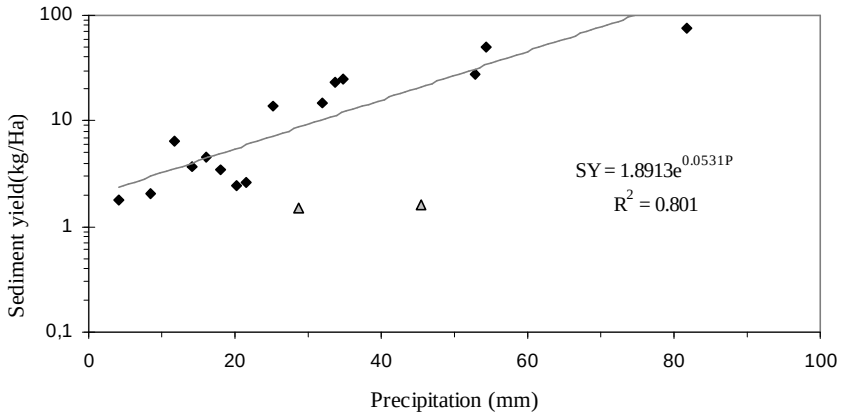


Figure 4. Relationship between sediment yield and rainstorm amount measured data for individual events on the plot.

8. CONCLUSIONS

Rainfall patterns on semiarid Brazilian region are dominated by high interannual variability and convective thunderstorms of high intensity. As other semiarid regions, a strong vulnerability to erosion processes may result in land degradation, desertification and reservoir siltation, causing social and economic consequences. These processes are strongly induced by human activities, such as inadequate grazing and deforestation. This study aimed to analyze runoff and soil erosion processes by using monitored data obtained during the 2006-2007 wetting up period on an upland undisturbed plot established in Brazilian semi arid region. Both continuous field survey and monitored data confirmed that vegetation cover is the major control on erosion process due to the influence on soil properties and soil infiltration process. It was observed that seed germination, annual vegetation replenishment and faunal activity influenced runoff and sediment production in a positive feedback. On the beginning of the rainy season, sediment and organic matter were transported by a sequence of rainfall events on degraded soil surface. Although the importance of such processes, they didn't seem to affect vegetative regeneration on the plot. Indeed, vegetative patterns of annual specie communities seemed to be controlled by a feedback relationship. Observed hydrographs demonstrated interdependency between rainfall sequences, antecedent soil wetness, vegetation patterns and runoff-erosion mechanisms. Runoff coefficients less than 0.1 occurred at approximately 82 per cent of the observed events. Obtained results highlight the role of vegetative species dynamics in a positive feedback with soil properties and faunal activity in Brazilian semiarid region.
Finally, an empirical relationship between short-term sediment production and rainstorm amount measured data for individual events on the plot showed an increasing trend.

9. ACKNOWLEDGEMENTS

The research of this study was supported by CNPq–Technologic and Scientific National Development Council/ Science and Technology Ministry/Brazilian Government. This support is gratefully acknowledged.

10. REFERENCES

Bergkamp, G. (1998) A hierarchical view of the interactions of runoff and infiltration with vegetation and microtopography in semiarid shrublands. *Catena*, **33**, 201-220.

Beven, K. (2002) Runoff generation in semi-arid areas. In: *Dryland Rivers: Hydrology and geomorphology of semi-arid channels.* (ed. By L.J. Bull & M.J. Kirkby), 57-105, John Wiley & Sons Ltd, England.

Boer, M. & Puigdefábregas, J. (2005) Effects of spatially structured vegetation patterns on Hill-slope erosion in a semiarid Mediterranean environment: a simulation study. *Earth Surface Processes and Landforms*, **30**, 149-167.

Bracken, L.J. & Kirkby, M.J. (2005) Differences in hillslope runoff and sediment transport rates within two semi-arid catchments in southeast Spain. *Geomorphology*, **68**,183-200.

Cadier, E. (1996) Small watershed hydrology in semi-arid in north-eastern Brazil: basin topology and transposition of annual runoff data. *Journal of Hydrology*, **182**, 117-141.

Corradini, C., Morbidelli, R. & Melone F. (1998) On the interaction between infiltration and Hortonian runoff. *Journal of Hydrology*, **204**, 52-67.

Darboux, F., Davy, Ph., Gascuel-Odoux C. & Huang C. (2001) Evolution of soil surface roughness and flow path connectivity in overland flow experiments. *Catena*, **46**, 125-139.

Dune, T. & Dietrich, W.E. (1980) Experimental studies of Horton overland flow on tropical hillslopes. 1. Soil conditions, infiltration and frequency of runoff. *Zeitschrift fur Geomorphologie Supplement Band*, **35**, 40-59.

Gallart, F, Solé, A, Puigdefàbregas, J. & Lázaro, R. (2002) Badland systems in the Mediterranean. In: *Dryland Rivers: Hydrology and geomorphology of semi-arid channels.* (ed. By L.J. Bull & M.J. Kirkby), 299-326, John Wiley & Sons Ltd, England.

Horton R. E. (1933) The role of infiltration in the hydrological cycle. Transactions of the American Geophysical Union, **14**, 446-460.

Kosmas, C., Danalatos, N., Cammeraat, L.H., Chabart, M., Diamantopoulos, J., Farand, R., Gutiérrez, L. Jacob, A., Marques, H., Matínez-Fernandez, J., Miozara, A., Moustakas, N., Nicolau, J.M., Oliveros, C., Pinna, G., Puddu, R., Puigdefràbegas, J., Roxo, M., Simao, A., Stamou,G., Tomais, N., Usai, D. & Vaca, A. (1997) The effect of land use on runoff and soil erosion rates under Mediterranean conditions. *Catena*, **29**, 45-59.

Ludwig, B., Boiffin, J., Chadoeuf, J. & Auzet, A.V. (1995) Hydraulic structure and erosion damage caused by concentrated flow in cultivated catchments. *Catena*, **25**, 227-252.

Molinier, M. & Cadier. E. (1985) Les sécheresses du Nordeste brésilien. Cah. ORSTOM, Série Hydrology, **4**, 23-49.

Moreira, L.F.F., Righetto, A.M. & Medeiros, V.M. de A. (2002) Infiltrômetro de Carga Constante. Available on http://www.hidro.ufcg.edu.br/twiki/bin/view/Rehisa/DocPublicacoes

Moreira, L.F.F., Righetto, A.M. & Medeiros, V.M. de A. (2006) Uncertainty analysis associated with Rainfall Spatial Distribution in an Experimental Semiarid Watershed, Northeastern Brazil. Summit on Environmental Modeling and Software, Burlington, VT, USA.

Nicolau, J.M., Solé-Benet, A., Puigdefràbegas, J. & Gutiérrez, L. (1996) Effects of soil and vegetation on runoff along a catena in semi-arid Spain. *Geomorphology*, **14**, 297-309.

Parsons, A.J., Wainwright, J., Abrahams, A.D. & Simanton, J.R. (1997) Distributed dynamic modeling of interril overland flow. *Hydrological Processes*, **11**, 1833-1859.

Srinivasan, V.S. & Galvão, C.O. (2003) Bacia Experimental de Sumé: descrição e dados coletados. (ed. by UFCG/CNPq), 129 p., Campina Grande, Paraíba, Brasil.

Wainwright, J. (1996) Infiltration, runoff and erosion characteristics of agricultural land in extreme storm events, SE France. *Catena*, **26**, 27-47.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Entropy Theory Application for Flow Monitoring in Natural Channels

T. Moramarco[a], A. Ammari[b], A. Burnelli[a], D. Mirauda[c] and V. Pascale[c]

[a] *National Research Council, IRPI, via M. Alta 126, 06128 Perugia, Italy (t.moramarco@irpi.cnr.it)*
[b] *Ecole Nationale Superieure de l'Hydraulique,09000 Blida, Algerie (ammariabdelhadi@yahoo.fr)*
[c] *Department of Environmental Engineering and Physics (DIFA), University of Basilicata, viale Ateneo Lucano 10, Potenza, Italy (mirauda@unibas.it)*

Abstract: A quick and accurate determination of flow passing through a river section is fundamental for a large number of engineering applications, such as flood forecasting models and the real time water resources management. Velocity is one of the basic variables of open channel flow and its variation across a section is described by the velocity profile distribution. This aspect has also been investigated through a probabilistic approach based on the entropy concept, which also expresses the mean flow velocity as a linear function of the maximum velocity through a dimensionless entropy parameter M. Therefore, the entropic relationship between the mean flow velocity and the maximum velocity has been investigated here for nine gauged river sections. Five of them are located in Italy and the other four are in Algeria. Based on the flow velocity measurements sample of each site, the robustness of the linear entropic relationship has been proved. The entropic parameter, M, which characterizes the relationship, has been found uniform for seven river sites with a value in the range of 2.1 – 2.5. For two river sites the value has been lower and equal to -0.22 and 0.33, respectively. As this parameter is fundamental for addressing the flow velocity measurements at equipped river sites, the morphological characteristics of river reaches where the gauged sections are located have been investigated and identified in terms of Rosgen classification. The seven gauged sections with uniform M value have been found belonging to the same stream classification, i.e. type C, except one located in Algeria which is type D. The other two with different M values have been ascribed as different types. The analysis, although preliminary, allows foreseeing new developments in the application of the entropy approach addressed to quantify high flow discharges, especially in new gauged river sites where M is unknown.

Keywords: Entropy theory; flow velocity; maximum flow velocity, rivers; Rosgen classification; hydraulic parameters.

1. INTRODUCTION

The accurate assessment of discharge at river sites is fundamental to address the water resource management in terms of both quality and quantity. Unfortunately, the discharge estimation is often affected by the lack of velocity measurements, often restricted by the excessive cost involved, which does not permit to establish a reliable stage-discharge relationship. Besides, even if velocity measurements are available, they are restricted in most cases to low flows, thus enabling a good discharge assessment only for lower water levels. Therefore, many studies attempted to overcome this shortcoming by addressing the analysis in terms of velocity profiles distribution. Although there are a large number of studies, few studies have been addressed for estimating the spatial velocity distribution during high flood conditions, wherein it is not possible to sample the whole velocity field, particularly in the lower portion of the flow area. In this case, the sampling procedure of velocity measurements at a river cross section could be difficult and particularly dangerous

for cableway operators. On the contrary, the value of maximum flow velocity could be more easily obtained since its position is located in the upper portion of the flow area where velocity measurements can be carried out also during high flow conditions (Moramarco et al., 2004). Therefore, a model able to assess the velocity profiles, also when velocity data are not available in any portion of the flow area should be welcome. For this purpose, many researchers have resorted to the use of the analysis tool based on the entropy theory (Shannon, 1948) and the pioneer among them is Chiu (1987, 1988, 1989). Based on this theory, it was found that the mean flow velocity can be also be estimated from the value of maximum velocity through a linear relationship identified by the entropic parameter M (Xia, 1997). Moreover, it was found that the M value is constant not only at each gauged site but also the same along the main channel reach, where these sites are located (Moramarco et al., 2004).

This insight might be very interesting for new equipped river sites where an accurate linear relationship can not be fitted. Therefore, sampling of the maximum flow velocity in the upper portion of the flow area of these new sites would be sufficient for estimating the mean flow velocity.

Based on the above insights, the main purpose of the paper is to more deeply investigate the entropic linear relationship and to attempt in justifying the M value also in terms of morphological stream characteristics, such as proposed by Rosgen (1994). To this aim, stream-gauges located in some natural channels with different hydraulic and geometric characteristics are analyzed. Velocity measurements data carried out in the investigated gauged sites during different periods were used for the analysis.

2. THE ENTROPIC LINEAR RELATIONSHIP

The mean velocity, $\bar{u}$, and the maximum velocity, u_{max}, sampled at river cross section can be expressed by Chiu's relationship (Chiu and Said, 1995):

$$\bar{u} = \Phi(M)\, u_{max} \tag{1}$$

where:

$$\Phi(M) = \left(\frac{e^M}{e^M - 1} - \frac{1}{M} \right) \tag{2}$$

and M is the entropy parameter. Eq. (1) shows that $\bar{u}$ and u_{max} together can determine $\Phi(M)$ and then the entropy parameter M. Plotting the pairs ($\bar{u}$, u_{max}) sampled at gauged stations, the best fit line can be estimated thus providing the values of $\Phi(M)$ and M for the gauged section. Xia (1997) by investigating some equipped sites along the Mississippi river found that the M value was quite similar for sections located along straight branches and equal to 2.1; whereas for sites along bends the M value was different and equal to 4.8. At similar findings Moramarco et al. (2002, 2003 and 2004) arrived by investigating some gauged sections located along natural channels of the Upper Tiber basin. They found M values not different from the ones furnished by Xia (1997) and, further, the M value didn't change at sections along the same river or stream. By way of example, Figure 1 shows the entropic linear relationships for three gauged sections (see Table 1) located along the Upper Tiber River, in Central Italy: Santa Lucia (933 km^2), Ponte Felcino (2040 km^2) and Ponte Nuovo (4135 km^2). As it can be seen the same value of $\Phi(M)$=0.667 can be assumed for the three sites, obtaining M=2.1. Therefore, it would seem that along the investigated Tiber River reach between Santa Lucia and Ponte Nuovo, the $\Phi(M)$ value can be surmised as constant. The inference was verified by Moramarco and Saltalippi (2002) at a gauged river section, named Pierantonio, located between Santa Lucia and Ponte Felcino sites. It was found that the best fit line with $\Phi(M)$ = 0.667 well reproduced the pairs of ($\bar{u}$, u_{max}) belonging to the sample of velocity measurements carried out at Pierantonio. This preliminary results establishes a possible linkage between the entropic parameter M and the hydraulic and geomorphological characteristics of the river.

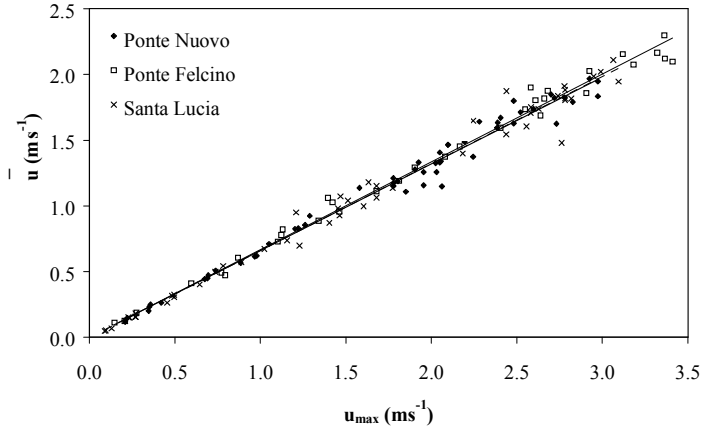


Figure 1. Relations between mean and maximum velocity in three gauged sections along the Tiber River. $\Phi(M)$ = 0.663.

To test the entropic relationship accuracy at gauged sections different from those already investigated along the Tiber River, velocity data collected from six sites of different rivers were analyzed. Four of these gauged sites belong to rivers of North Algeria; the two remaining are located on the Basento River, in the South of Italy. The selected sections are equipped with remote ultrasonic water level gauge while the measurements of velocity were made by current meter from cableway or bridge cranes. By comparing the point velocities sampled along different verticals and applying the well-known velocity-area method, the maximum and mean velocities were estimated (Herschy, 1985). Topographical surveys of the equipped sections were also available in different years of the sampling. The values of $\Phi(M)$ and the entropy parameter M are shown in Table 2, for Tiber River sites.

Table 1. Flow characteristics, discharge, Q, and water depth, D, of the available velocity measurements, N, for the three gauged sections located along the Tiber River. The period of sampling is also shown.

Site	N	Q (m^3/s)	D (m)	Period
Santa Lucia	51	1-215	0,9-5,2	1986-2005
Ponte Felcino	34	2-412	0,8-6,2	1986-2005
Ponte Nuovo	54	3-537	1,1-6,7	1986-2005

Table 2. Value of the entropic parameter, M, and of the best fit slope, $\Phi(M)$, for the gauged sites along the Tiber River. The determination coefficient, R^2, is also shown.

Site	Φ(M)	M	R^2
Santa Lucia	0.662	2.08	0.98
Ponte Felcino	0.668	2.17	0.99
Ponte Nuovo	0.660	2.05	0.98

The four Algerian streams are located in the coastal Algiers hydrographic basin, which is characterized by a Mediterranean climate with mean annual precipitation of 800 mm. All these sites are subjected to high erosion processes during flood events because of the presence of sand sediments at the channel bed and banks. Table 3 shows the main flow characteristics of the investigated gauged sections.

Table 3. As in Table 1, but for the Algerian streams.

River	Site	N	Q (m^3/s)	D (m)	Period
Chiffa	Amont des Gorges	166	0,02-28	0,1-1,1	1990-2004
Sebaou	Baghlia	121	0,01-173	0,1-6,65	1990-2001
Mazafran	Fer à Cheval	317	0,01-350	0,1-7,8	1985-2005
Belah	Pont RN11	53	0,08-66,3	0,13-2,04	1993-2006

The $\Phi(M)$ value for each gauged station has been assessed based on the sampled pairs of ($\bar{u}$, u_{max}). Figure 2 shows the relationship between mean and maximum velocities for the Algerian gauged sections, assuming the real maximum velocity equal to the maximum value sampled at each site (Chiu, 1988).

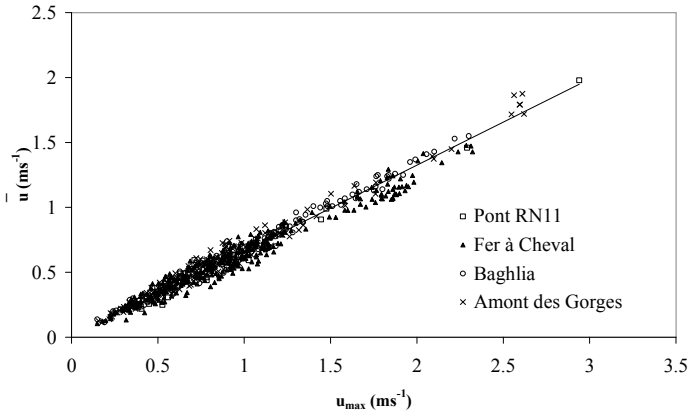


Figure 2. Relations between mean and maximum velocity in the gauged sections located in the Algiers basin. *Φ(M)* = 0.667.

For each investigated river site, the values of $\Phi(M)$ and of the entropy parameter M are shown in Table 4, along with the determination coefficient of the best fit line.

Table 4. As in Table 2, but for the Algerian sites.

River	Site	Φ(M)	M	R^2
Chiffa	Amont des Gorges	0,689	2,49	0,98
Sebaou	Baghlia	0,681	2,37	0,99
Mazafran	Fer à Cheval	0,641	2,02	0,96
Belah	Pont RN11	0,663	2,10	0,97

A good linear relationship was found for a large range of discharges, water levels and sediment concentrations. Therefore, based on the estimated $\Phi(M)$ value, the assessment of $\bar{u}$ through the u_{max}, can be easily done without losing in accuracy. Considering that the maximum velocity is located at the deepest vertical near the water surface, the velocity measurements can be addressed at the purpose. We need to point out that also in this case the $\Phi(M)$ value can be considered uniform and quite similar to that estimated for the sites located in the Upper Tiber basin.

The analysis has been also extended for two more gauged sites on the Basento River, in Southern Italy. In this case, the sections are stable and no erosion processes have been taken place during the period of study. The first site, Campomaggiore, is located in the upper portion of the basin (840 km^2), characterized by a steep slope value and by the presence of boulders and cobbles in the cross section. The second one, Torre Accio, shows features

typical of a valley reach (1380 km^2), with increasing sinuosity, extremely reduction of slope and the presence of fine sediments and silts. Both sites are characterized by the presence of shrubby vegetation and trees. Table 5 shows the main flow characteristics of the two river sites.

Table 5. As in Table 1, but for the Basento River sites, in Southern Italy.

Site	N	Q (m^3/s)	D (m)	Period
Campomaggiore	6	1-62	0,7–2,0	2002 - 2005
Torre Accio	7	1-180	0,5-3,3	2002 - 2005

As can be seen, the number of sample measurements in terms of Φ assessment, might be not properly representative for investigated sites. However, as some pairs of measured ($\bar{u}$, u_{max}) pertain to high flows, the analysis can be still valid, despite the limited sample size. The values of $\Phi(M)$ and of the entropy parameter M are shown in Table 6, for both Basento River sites.

Table 6. As in Table 2, but for the Basento River sites, in Southern Italy.

Site	Φ(M)	M	R²
Campomaggiore	0,48	-0,22	0,98
Torre Accio	0,53	0,33	0,94

In this case the entropic parameter was found different both among the two sites and from the one obtained for the Tiber River and the four Algerian rivers. Figure 3 shows the relationship between mean and maximum velocities for these gauged sections. Based on the $\Phi(M)$ values, it would seem that for both river sites the flow is slower than the one along the Tiber River and Algerian rivers.

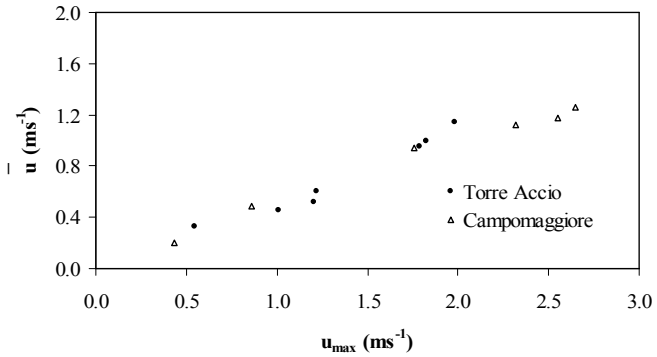


Figure 3. Relations between mean and maximum velocity in the gauged sections located along the Basento River (South Italy).

3. ENTROPIC PARAMETER AND RIVERS MORPHOLOGICAL CHARACTERISTICS

Considering the uniformity of Φ values at gauged sections, it has to be investigated if the Φ value can be explained in terms of stream types such as defined through the Rosgen classification (1994). Therefore, the morphological characteristics of rivers where the investigated gauge sections operate have been analyzed. In particular, the main stream types can be identified into seven major groups, i.e., A, B, C, D, E, F and G. Summarizing, rivers quite straight can be identified as type A; a river pattern with low sinuosity, B and G, and meandering type C and F. Complex patterns with high tortuosity and braided are defined as E and D, respectively. All these groups can be defined through the following morphological parameters (Rosgen, 1994): a) the entrenchment ratio; b) the W / D ratio,

where W is the bankfull channel width and D is the bankfull mean depth; c) the sinuosity defined as the ratio of stream length to valley length and d) the slope evaluated in a reach of about 20 channel widths. In this case, the term "bankfull" is referred to water level corresponding to the discharge with a return period of 1.5 years. The entrenchment ratio is estimated as the ratio between the channel width, measured twice at the maximum bankfull depth and W, as previously defined. For more details the reader may refer to Rosgen's work (1994). For each river site, the above mentioned morphological parameters have been assessed by using the available data from topographical surveys, cartographic maps as well as hydrological outcomes. Table 7 shows the above parameters for all investigated gauged sections with the corresponding stream type.

Table 7. Results of the Rosgen classification for the nine investigated sites.

Site	River	Entrenchment	W/D	Sinuosity	Slope %	Rosgen Class
Santa Lucia	Tiber	> 3.3	8	1.7	0.2	**C**
Ponte Felcino	Tiber	3	10	1.4	0.15	**C**
Ponte Nuovo	Tiber	> 11	12	1.8	0.14	**C**
Campomaggiore	Basento	1.2	8	1.5	1.62	**G**
Torre Accio	Basento	40	12	3.2	0.16	**E-C**
Amont des Gorges	Chiffa	2.4	29	1.4	0.5	**C**
Baghlia	Sebaou	2.3	51	1.5	0.2	**D**
Fer à Cheval	Mazafran	2.4	31	1.5	0.12	**C**
Pont RN11	Belah	1.2	24	1.3	1.3	**C-F**

It can be inferred that the gauged sections having the M values greater than 2 belong to stream type C, except the Sebaou River at Baghlia which is of type D, characterizing a braided shape (see Figure 4).

With regard to Pont RN11 on Belah River, the type can not be fairly defined because of lack of further information useful to accurately assess the entrenchment ratio. In this case, however, the river type might be referred to type C or F. On the contrary, for the two river branches located along the Basento River in South Italy, the upper reach has been identified as type G; whereas uncertainty occurs for the Basento at Torre Accio where the branch type can not be properly defined. Also in this case, further accurate information are required for assessing the entrenchment and W/D morphological parameters. However, the type E or C might represent the range more appropriate for the branch. Therefore, through this preliminary analysis, it would seem that a link between the entropy parameter and the morphological parameters might be found, at least in terms of overall insight. However, Figure 5 reveals that difference can be found between the morphology of the Tiber River at Ponte Nuovo (type C) from that of Sebaou River (type D) (shown in Figure 4) which has evident braiding characteristics.

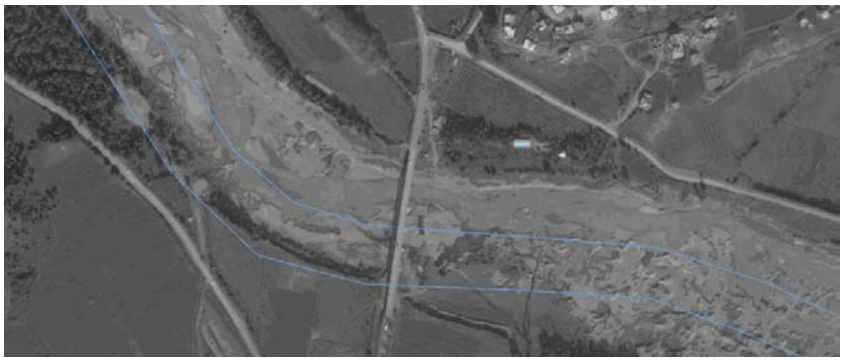


Figure 4. Plan view of the Sebaou River branch at Baghlia (mean width 200 m).

Therefore, at this level of analysis, it is not possible to clearly explain the differences between these two river types, although they have the same Φ value. In this context and also based on the findings of Xia (1997), who estimated Φ values around 0.64 and 0.79 respectively along straight Mississippi River reaches and on bends, one would be prompted to reconsider the Rosgen classification, which could thus address the analysis on the relationship between the entropic parameter, M, and morphological characteristics of river branches.



Figure 5. Plan view of the Tiber River branch at Ponte Nuovo (mean width 60 m).

5. CONCLUSIONS

The entropic linear relationship has been investigated by using the velocity measurements data sampled at nine gauged river sites, operating in six rivers, of which two are in Italy: Tiber and Basento rivers, and four are in Algeria: Mazafran, Chiffa, Belah and Sebaou rivers. The value of the entropic parameter M was found quite similar (M≅2.1) for the Tiber River and the four Algerian rivers. A different value was obtained for the two sites located along the Basento River. Through a preliminary analysis on morphological characteristics of rivers and based on Rosgen classification, it would seem that gauged stations belonging to the same river type might have the same value of M. However, further data referring to other rivers are required to confirm this inference.

REFERENCES

Chiu, C.L., and Chiou, J.D., Structure of 3-D flow in rectangular open channels, *J. Hydr. Engrg*., *ASCE, 112(11)*, 1050-1068, 1986.

Chiu, C.L., Entropy and probability concepts in hydraulics, *J. Hydr. Engrg*., *ASCE, 113(5),* 583-600, 1987.

Chiu, C.L., Entropy and 2-D velocity distribution in open channels, *J. Hydr. Engrg*., *ASCE, 114(7)*, 738-756, 1988.

Chiu, C.L., Velocity distribution in open channel flow, *J. Hydr. Engrg*., *ASCE, 115(5)*, 576-594, 1989.

Chiu, C.L., and Abidin Said, C.A., Maximum and mean velocities and entropy in open-channel flow, *J. Hydr. Engrg*., *ASCE, 121(1)*, 26-35, 1995.

Herschy, R.W., Streamflow measurement, *Elsevier*, London, 1985.

Moramarco, T., and Saltalippi, C., Stima della velocità media in sezioni di un corso d'acqua naturale, *Proceedings of the 28th National Hydraulics and Hydraulic Engineering Conference*, *September 16-19, Vol. 3*, Potenza (Italy), 263-274, 2002.

Moramarco, T., and Saltalippi, C., A practical method for mean flow velocity estimation during high flood, *Proceedings of the IASTED International Conference APPLIED SIMULATION AND MODELLING, September 3-5,* Marbella (Spain), 249-254, 2003.

Moramarco, T., Saltalippi, C., and Singh, V.P., Estimation of mean velocity in natural channel based on Chiu's velocity distribution equation, *J. Hydrol. Engrg*., *ASCE, 9(1)*, 42-50, 2004.

Rosgen, D.L., A classification of natural rivers, *Catena*, *(22)*, 169-199, 1994.

Shannon, C.E., A mathematical theory of communication, *Bell System Technical Journal*, *(27)*, 379-423, 623-656, 1948.

Xia, R., Relation between mean and maximum velocities in a natural river, *J. Hydr. Engrg*., *ASCE, 123(8)*, 720-723, 1997.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

On the Lateral Inflows Assessment Within a Real-Time Stage Monitoring Addressed to Flood Forecasting

S. Barbetta[a], L. Brocca[a], F. Melone[a] and T. Moramarco[a]

[a] National Research Council, Research Institute for Geo-Hydrological Protection, Via Madonna Alta 126, 06128 Perugia, Italy [s.barbetta@irpi.cnr.it]

Abstract: The role of the lateral inflows component for an on-line stage forecasting model Muskingum based is investigated. In particular, the original empirical formulation is replaced by a physically based approach incorporating the error on lateral contribution that is updated using the stage observations available in real-time. The updating procedure for lateral inflows assessment is different depending on the weight of the lateral contribution in the downstream flood evolution. The model requires the estimation of four parameters if the downstream rating curve is unknown, otherwise only two parameters have to be determined. The accuracy of the new model formulation is tested by its application to different flood events observed along an equipped river reach of the Upper Tiber River basin, in Central Italy, characterized by a significant intermediate drainage area. Assuming the rating curve known at the downstream site, the forecast stage hydrographs are found in good agreement with the observed ones for all the investigated events and more accurate than those provided by the original model. A preliminary analysis has highlighted that the forecast reliability is not significantly affected by the rainfall spatial variability in the intermediate basin, whereas it is strictly linked to the ratio between the rainfall depth in the upstream and intermediate drainage area (RRI). Moreover, a threshold of RRI has been defined in order to identify the forecast stages which could be affected by a not negligible degree of uncertainty. On this basis, the proposed model can be conveniently employed as part of a real-time flood forecasting system involving the RRI continuous monitoring during critical storm events.

Keywords: real-time flood forecasting; lateral inflows; Muskingum method; rainfall spatial variability.

1. INTRODUCTION

Short-term (or real-time) forecasting is fundamental for both minimizing the effects of forthcoming flood events in flood-prone regions and optimizing water resources management where dam safety is also of concern. Models within a flood forecasting system aim at producing future estimates of hydrological variables, based on the present state and the past behavior of the investigated catchment or river reach, thus allowing to reduce the uncertainty in decision making for the evolution of the flood event. The mathematical models applied at the purpose generally belong to either conceptual or black-box approaches. For medium-sized catchments (drainage area of the order of 10^3 km^2) conceptual rainfall-runoff models of the semi-distributed type seem to be most reliable for operational activities provided that an updating procedure is adopted [Michaud & Sorooshian, 1994; Boyle et al., 2001].

To reduce the complexity of the flood forecasting system, simplified models based only on flood routing have been also used [O'Connell & Clarke, 1981; Lamberti & Pilati, 1996] for flood-prone sites located downstream of a gauged station and at a distance allowing an appropriate forecasting lead-time. Moreover, this type of models reduces the level of

forecast uncertainty since the water level measurements are definitely more accurate than the rainfall ones. In this context, the Muskingum model can be a useful tool. However, critical points in hydrological routing are the representation of lateral inflows contribution and the knowledge of stage-discharge relationships. As regards the former, O'Donnell [1985] proposed a three-parameter Muskingum procedure assuming the lateral inflows proportional to the contribution entering upstream. Using this approach, Franchini and Lamberti [1994] presented a simple model Muskingum type to provide forecast water levels at the downstream end by selecting a routing time interval and, hence, a forecasting lead-time which allows to express the forecast stage as a function of only observed quantities. Moramarco et al. [2006] enhanced the modeling scheme by Franchini and Lamberti [1994] incorporating a procedure for adapting the parameter linked to lateral inflows. This last model, called STAFOM (STage FOrecasting Model), was successively extended by Moramarco et al. [2006] to a two connected river branches schematization in order to improve significantly the forecasting lead-time. The STAFOM model provided satisfactory results for most of the analysed flood events in each investigated river reach in the Upper-Middle Tiber River basin in Central Italy. However, the updating procedure for the parameter linked to the lateral contribution was based on empirical assumptions.
This might be questionable for floods with characteristics different from those there employed for model calibration and evaluation phase. On this basis, the STAFOM model has been modified by incorporating a physically based simplified approach for lateral inflows representation.
The objective of the paper is, then, to describe the general features of the modified stage forecasting model and to test its performance along a river reach selected in the Upper Tiber River basin, in Central Italy. A preliminary analysis has been also carried out to assess the effect of the rainfall field on the intermediate basin on the accuracy of the forecast stage hydrographs. For this aim, two indexes have been used. The first addresses the rainfall field variability on the basin and the second one represents the significance of the intermediate drainage area contribution.

2. THE STAGE FORECASTING MODEL STAFOM

2.1 Modelling Scheme

For a river reach bounded by two hydrometric sites with not negligible lateral inflows the Muskingum approach is expressed as:

$$Q_d(t+\Delta t) = C_0 I(t+\Delta t) + C_1 I(t) + C_2 Q_d(t) \quad (1)$$

where Q_d is the discharge at the downstream end. C_0, C_1 and C_2 are the Muskingum parameters depending only on K and x coefficients value and on the time interval, Δt. The term I in equation (1) is the total discharge entering into the selected river reach defined by the upstream contribution, Q_u, and the lateral inflows, Q_l:

$$I(t) = Q_u(t) + Q_l(t) \quad (2)$$

In particular, the lateral inflows are mainly due to the tributaries contribution which is determined by the overland flow formation and propagation process during the flood event.
In order to get the formulation for a real-time forecasting model, the term $C_0 I(t+\Delta t)$ in equation (1) has to be eliminated and this can be achieved by assuming a time interval $\Delta t = \Delta t^* = 2Kx$ for which $C_0 = 0$ [Franchini & Lamberti, 1994]. For each time of forecast, t_f, the estimation of the downstream discharge carried out Δt^* hours in advance is then given by:

$$Q_d(t_f + \Delta t^*) = C_1^* \left[Q_u(t_f) + Q_l(t_f) \right] + C_2^* Q_d(t_f) \quad (3)$$

where C_1^* and $C_2^* = 1 - C_1^*$ are now referred to the Muskingum parameters K and x with the constrain $\Delta t^* = 2Kx$ [Moramarco et al., 2006]. The term in equation (3) linked to the lateral inflows, $Q_l(t_f)$, can be assessed by considering the continuity equation written in the characteristic form. Following Moramarco et al. [2005] the lateral contribution per unit

channel length, q, is expressed as a function of the upstream and downstream flow area, A, and of the wave travel time, T_L, which is the time shift to match the rising limb and the peak region of the upstream and downstream dimensionless stage hydrographs:

$$q(t) = \left(A_d(t) - A_u(t - T_L)\right) / T_L \tag{4}$$

Assuming $T_L=\Delta t^*$, $Q_l(t_f)$ is given by:

$$Q_l(t_f) = q(t_f)L = \left[\left(A_d(t_f) - A_u(t_f - \Delta t^*)\right) / \Delta t^*\right] L \tag{5}$$

where L is the channel length.

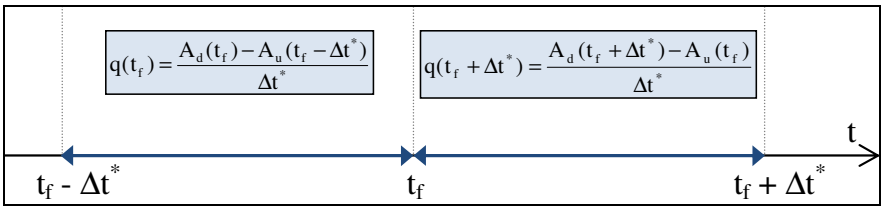


Figure 1. Lateral contribution per unit channel length, q, referred to different time intervals (for symbols see text).

In order to determine the downstream discharge at time $(t_f+\Delta t^*)$, in the model formulation the q value referred to the time interval $[t_f; t_f+\Delta t^*]$ should be employed (see Figure 1). However, it is an unknown quantity and, hence, the lateral contribution has to be assumed invariant over the forecasting lead-time, Δt^*, and equal to $q(t_f)$. However, for floods characterized by a short duration, the lateral inflows can be expected to be highly varying in time. Moreover, considering the simplified approach adopted for lateral inflow representation an additive error, ε, for q is introduced, so that equation (3) is rewritten as:

$$Q_d(t_f + \Delta t^*) = C_1^* Q_u(t_f) + C_1^* \left[q(t_f) + \varepsilon\right] L + C_2^* Q_d(t_f) \tag{6}$$

Assuming a kinematic relationship, $Q=\lambda h^{\delta}$, for the rating curve at the downstream section, the general formulation for the STAFOM model can be obtained from equation (6):

$$h_d(t_f + \Delta t^*) = \left\{(1/\lambda)\left[C_1^*\left(Q_u(t_f) + q(t_f)L + \varepsilon L\right) + C_2^* \lambda h_d^{\delta}(t_f)\right]\right\}^{1/\delta} \tag{7}$$

where h_d is the water level at the channel outlet and λ and δ are parameters to be estimated.

2.2 Updating Procedure

If the rating curves at the ends are accurate, the main uncertainty can be assumed to be linked to the lateral inflows assessment. Therefore, the error on the lateral contribution estimate, that is the ε term, is updated through an automatic procedure which, obviously, is based on the last measurements carried out by the monitoring network operating on-line. In particular, for each time of forecast, t_f, ε is estimated assuming that the model furnishes the optimal forecast at t_f when the lateral inflows in the forecasting period is considered. This means that the error on lateral contribution is given by the difference between the discharge observed at the downstream section, $Q_{d_obs}(t_f)$, and the discharge forecasted by the model Δt^* hours before using the lateral contribution per unit channel length, q, at time t_f or $(t_f-\Delta t^*)$ for events characterized by predominant upstream or lateral inflows, respectively:

$$\varepsilon(t_f) = \left(Q_{d_obs}(t_f) - C_1^* Q_u(t_f - \Delta t^*) - C_2^* Q_d(t_f - \Delta t^*)\right) / \left(C_1^* L\right) - q \tag{8}$$

The different approaches for the q estimate are necessary for taking the lateral inflows underestimation into account occurring during the rising limb when the intermediate contribution is predominant in the downstream flood wave formation. The choice of the approach to be applied is based on the real-time stage observations at the two channel ends which allow to assess the significance of the lateral inflows.

2.3 Model Parameters Evaluation

In order to employ the stage forecasting model expressed by equation (7), all the involved parameters have to be defined. In particular, beside the adaptive parameter, ε, the values of λ, δ and C_1^* have to be assessed. As far as the rating curve coefficients, λ and δ, are concerned they are easily inferred when an accurate stage-discharge relationship is available at the downstream site. Otherwise, they can be derived through the Rating Curve Model (RCM) proposed by Moramarco et al. [2005] which also takes the lateral inflows contribution into account along the selected river reach. The C_1^* value depends only on the Muskingum coefficients, K and x. In particular, K is evaluated considering that in the classical Muskingum formulation it represents the mean wave travel time along the investigated river branch. Therefore, it can be estimated through few observed flood events. For each event the wave travel time, T_L, is computed as the time shift necessary to overlap the rising limb and the peak region of the normalized stage hydrographs [Moramarco et al., 2005]. Finally, K is obtained as the average of the T_L values. It represents a characteristic of the selected reach and the value of the forecasting lead-time, Δt^*. The assumption $\Delta t^*=K$ implies that x has to be set equal to 0.5 which, however, is in the typical range of the parameter values [Chow et al., 1988] and, hence, C_1^* and C_2^* are set equal to 1 and 0, respectively, in the model formulation. This means that the Muskingum approach follows the kinematic wave dynamic [Kundzewicz & Strupczewski, 1982].

3. MODEL APPLICATION AND RESULTS

The accuracy of the modified forecasting model was verified considering some flood events occurred along one equipped river reach in the Upper Tiber River basin bounded by the hydrometric sites of Pierantonio and Ponte Nuovo, whose location is shown in Figure 2 along with the main sub-basins and the existing raingauge network operating in real-time.

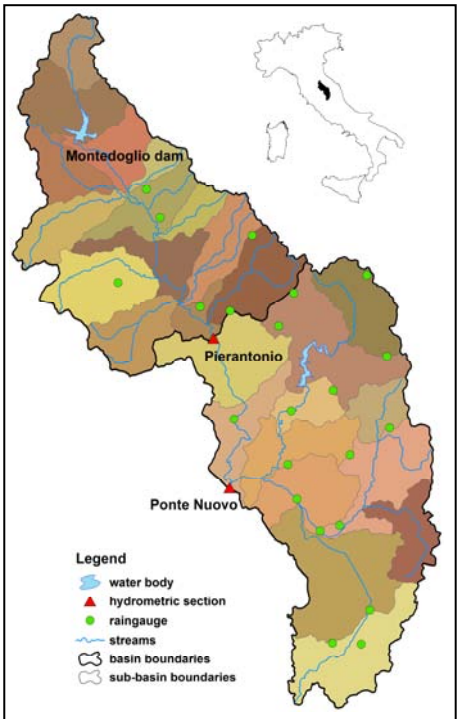


Figure 2. Location of the selected sections in the Upper Tiber River basin. The existing raingauge network and the main sub-basins are also shown.

The main geomorphological properties of the reach are summarized in Table 1. As can be seen, the reach is characterized by a significant intermediate drainage area, equal to 2340 km^2, which is about 56% of the total basin and, hence, the lateral inflows contribution can be fundamental in flood wave formation process. Moreover, an artificial reservoir (Montedoglio dam) is located in the upper part of the basin subtending a drainage area of about 275 km^2 (see Figure 2). An accurate stage-discharge relationship was available for both the upstream section and the downstream one. Significant flood events observed during the last decade were selected for the analysis and their main characteristics are summarized in Table 2. As can be seen, they are characterized by different values of the percentage lateral inflow contribution, LC. In particular, the events with LC<50% are considered 'Prevalent Upstream Contribution' (PUC) events, while the others 'Prevalent Lateral Contribution' (PLC) ones. For each water level the flow areas in equation (5) were estimated by the sections geometry defined through accurate topographical surveys. The model parameters λ and δ were derived from the knowledge of Ponte Nuovo rating curve, while K was assessed considering the mean

wave travel time, T_L, of the selected river reach. T_L values are shown in Table 2 and an average can be surmised equal to 4.5 hours. Therefore, K can be assumed equal to 5 hours considering also the need to adopt a lead-time as high as possible.

Table 1. Main geomorphological properties of the selected reach: area subtended by each section, A_b, channel length, L, mean channel slope, S_0, mean section width, B.

Sections	A_b [km^2]	L [km]	S_0	B [m]
Pierantonio	1805	40.23	0.0013	46
Ponte Nuovo	4145			

Table 2. Main floods characteristics (peak stage, h_p, peak flow, Q_p, wave travel time, T_L) and model results (percentage error on peak stage, e_{hp}, and time to peak, e_{tp}; coefficient of persistence for the event, E_v, and the rising limb, E_r; Nash-Sutcliffe coefficient, NS). The Spatial Deviation Index, SDI, and the Rainfall Ratio Index, RRI, are also shown along with the percentage lateral inflows contribution, LC.

Date	Pierantonio		P.Nuovo		T_L	LC	e_{hp}	e_{tp}	E_v	E_r	NS	SDI	RRI
	h_p [m]	Q_p [m^3/s]	h_p [m]	Q_p [m^3/s]	[h]	[%]	[%]	[h]	[%]	[%]	[%]		
Dec. 96,13	4.32	378	6.43	728	4.5	47	-1.3	2.0	77.9	78.0	91.5	0.200	0.99
Apr. 97, 20	4.67	429	5.16	502	5.5	18	-2.2	-3.0	95.3	96.5	98.9	0.154	0.80
Nov. 97, 22	3.82	308	4.56	409	3.5	36	6.6	4.0	83.4	83.3	95.4	0.131	0.85
Dec. 97, 26	2.17	121	3.0	189	3.0	44	5.6	1.0	80.4	79.0	93.8	0.112	1.04
Jan. 98, 19	3.12	221	3.79	302	4.5	33	1.7	1.0	96.4	96.2	98.1	0.215	1.00
Feb. 98, 3	1.56	70	2.91	177	2.5	63	12.7	3.0	43.0	75.7	86.2	0.116	1.44
Apr. 98, 15	3.12	221	4.77	441	7.0	51	10.7	4.0	33.8	79.6	90.0	0.176	1.62
Oct. 98, 5	1.84	92	4.85	453	0.0	78	16.6	2.0	39.7	58.6	86.8	0.161	1.65
Dec. 98, 3	3.24	235	6.37	717	3.5	62	22.6	3.0	23.4	59.1	56.8	0.127	1.87
Feb. 99, 9	4.66	428	6.61	764	6.0	45	-5.4	2.0	85.6	89.2	94.5	0.093	0.85
Dec. 99, 14	2.31	134	6.62	766	2.0	80	8.4	-2.0	60.4	62.5	93.3	0.169	2.03
Dec. 00, 25	5.52	552	7.17	879	8.0	36	8.7	1.0	66.9	70.4	95.3	0.125	0.65
Mar. 01, 1	2.83	188	4.26	367	4.0	50	2.4	3.0	73.0	76.8	91.5	0.348	0.80
Apr. 01, 7	3.28	240	3.78	300	5.5	30	3.3	-1.0	96.9	97.1	97.0	0.118	0.96
Nov. 05, 25	6.70	779	8.19	1110	6.0	39	-2.6	5.0	79.2	92.5	96.4	0.085	0.99
Dec. 05, 2	4.70	434	5.91	631	5.5	35	2.9	2.0	86.2	86.2	92.6	0.233	0.77
Dec. 05, 5	5.09	495	7.13	871	4.0	47	-2.4	2.0	81.3	80.5	93.1	0.104	1.26
Dec. 05, 30	4.59	417	6.04	655	7.0	41	-2.2	1.0	86.8	85.1	98.2	0.139	1.06
Feb. 06, 20	1.88	96	4.19	357	2.0	74	8.2	3.0	34.5	37.8	85.3	0.107	1.74

The results provided by the model are summarized in Table 2 in terms of error on peak stage and time to peak along with the coefficient of persistence (E) [Kitanidis & Bras, 1981] and the Nash-Sutcliffe efficiency coefficient (NS) [Nash & Sutcliffe, 1970]. In particular, the coefficient E compares the prediction of the model against the obtained one by the no-model, which assumes the steady state over the forecasting lead-time. It has been computed for the rising limb (E_r) and up to the time in the recession limb when the dimensionless stage becomes equal to 0.5 (E_v). Considering the PUC events (LC<50%) the model was found very accurate in terms of error on peak stage which was less than 10% with a mean absolute error equal to 3.6%. On the contrary, this last error goes up to 13.2% for the PLC floods. As regards the error on time to peak, it has to be underlined that for events with a wide peak region the difference between the observed and the forecast peak time is not adequate to assess the model reliability in capturing flood-peak timing. In general, the model provides forecast peak stage with a mean absolute time error equal to 2.1 and 2.8 hours for the PUC and the PLC floods set, respectively. The coefficients of persistence, E_v and E_r, show that the proposed model is very useful, particularly during the rising limb period, for most of the selected flood events. In particular, for the PUC events set the values of E_v and E_r are greater than 70% with a mean value of about 85%. Moreover, the NS coefficient shows that the stage hydrograph shape is well simulated with a value always greater than 90% and equal on average to 95.1%. As regards the other flood events, the model performance can be still considered satisfactory even if the E_v and E_r assume lower values on average equal to 39.2% and 62.2%, respectively. Figure 3 shows the comparison between the observed and the forecast stage hydrograph for the floods occurred on April 97

and November 05, belonging to the PUC floods set, and those observed on December 99 and February 06 characterized by prevalent intermediate basin contribution.

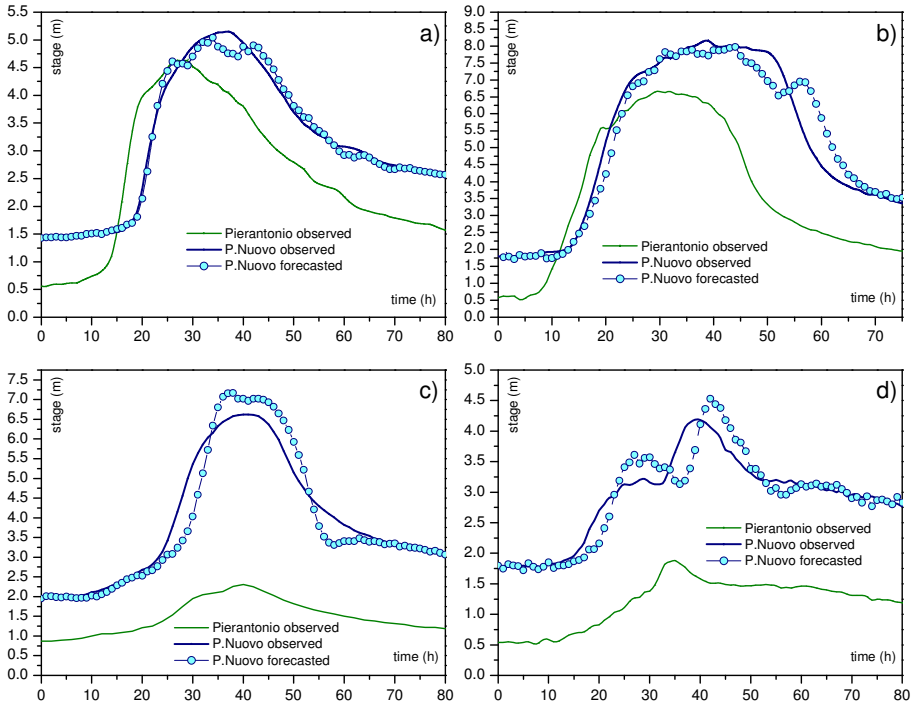


Figure 3. Comparison between observed and forecast stage hydrograph for the floods occurred on: a) April 97; b) November 05; c) December 99; d) February 06.

Considering that for the present case study the upstream drainage area and the intermediate one are equal to 44% and 56% of the total area, respectively, we can suppose that the model accuracy might be significantly influenced by the ratio between the contribution entering upstream and the lateral inflows as well as by the spatial rainfall distribution in the intermediate basin. Therefore, a preliminary analysis was carried out quantifying for each storm the rainfall field variability over the intermediate drainage area through the Spatial Deviation Index proposed by Segond et al. [2007]:

$$SDI = (1/N)\sum_{i=1}^{N}\left(\left|P_{ri} - P_I\right|/P_I\right) \tag{9}$$

where N is the number of sub-basins in which the intermediate basin is divided. P_{ri} and P_I is the Thiessen averaged areal rainfall for the i-th sub-basin and for the intermediate basin, respectively. In particular, the intermediate basin between Pierantonio and Ponte Nuovo sections was subdivided in 12 sub-basins shown in Figure 2 along with the raingauge network used for the analysis. Figure 4a shows the plot of E_v and E_r against the SDI index, (see Table 2) suggesting that a correlation does not exist. In principle, this result is affected by a sort of uncertainty linked to the representation of rainfall spatial variability through the existing raingauge network and the adopted spatial discretization. However, the different model accuracy for events characterized by an almost uniform rainfall (e.g. February 99 and February 06) suggests that this uncertainty can be considered negligible. In order to investigate this, the Rainfall Ratio Index (RRI) was computed for each storm as the ratio between P_I and the Thiessen averaged areal rainfall for the basin between Montedoglio dam and Pierantonio site, P_U, ($RRI=P_I/P_U$). Figure 4b shows that there is a linear relationship between E_r, more significant for flood forecasting purpose, and the RRI index. The floods occurred on April 98, December 00 and February 06, highlighted by red squares in Figure 4b, were not considered in the relationship assessment since they are characterized by a complex shape and multiple peaks. On the basis of the obtained results, it can be inferred that for the investigated river reach the accuracy of the proposed model is mainly influenced by the significance of the intermediate basin contribution in the flood wave formation process. In particular, for RRI less than 1.0÷1.2 the model results can be considered

accurate being the coefficient of persistence greater than 80%. On the contrary, when RRI becomes greater than 1.4, which can be assumed as a preliminary threshold value, the provided forecast stages should be considered affected by a not negligible degree of uncertainty. However, it has to be emphasized that the NS coefficient was found greater than 85% for all the investigated events except for the flood occurred on December 98.

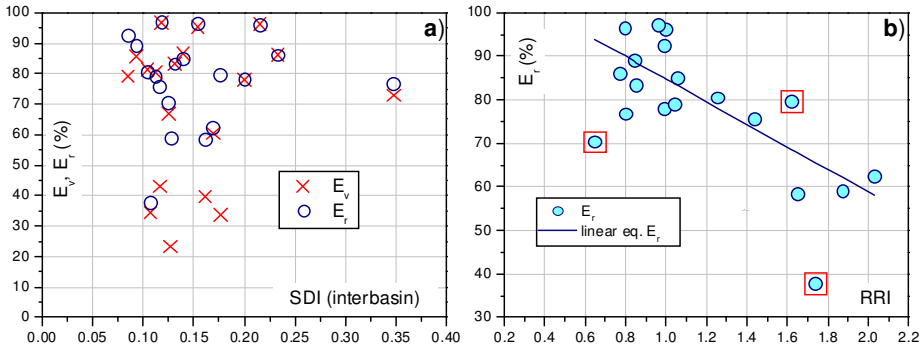


Figure 4. Relationships between coefficients of persistence E_v and E_r and a) SDI index; b) RRI index. Red square indicates floods characterized by a complex shape and multiple peaks (for symbols see text)**.**

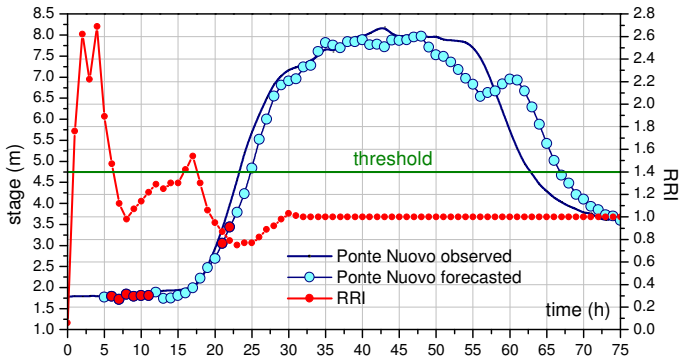


Figure 5. November 05: comparison between observed and forecast stage hydrograph along with the rainfall index, RRI. The red points indicate the forecast stages referring to RRI values greater than the fixed threshold.

However, these results refer to the total rainfall depth. In order to check the usefulness of the RRI index for having indications on the forecast stage reliability in real-time, it was computed at hourly time step considering the rainfall occurred up to the time of forecast. Specifically, the real-time RRI value monitoring at hourly time step has been simulated for the floods occurred on December 98, poorly forecast, and November 05, accurately forecast. As regards the former flood, the ratio P_I/P_U was found significantly greater than 1.4 for all the times of forecast, whereas for the second flood RRI overcomes slightly the fixed threshold just at eight forecast times whose corresponding forecast stages are shown as red points in Figure 5. As can be seen, they refer to periods before and at the beginning of the rising limb and, hence, are not fundamental for flood forecasting purposes.

4. CONCLUSIONS

The stage forecasting model proposed by Barbetta et al. [2003] for real-time applications was modified in the lateral inflows representation by introducing a physically based approach. In particular, two approaches were proposed for floods with prevalent upstream or lateral contribution, respectively. The accuracy of the modified model has been tested considering several floods occurred along the Pierantonio-Ponte Nuovo river reach, in the Upper Tiber River basin in Central Italy, characterized by a significant intermediate drainage area. Assuming the rating curve known at the downstream site, the forecast stage hydrographs were found in good agreement with those observed for all the investigated case

studies even if a slightly lower model reliability characterized the floods affected by significant intermediate basin contribution. Moreover, the performance of the new formulation was found more accurate than that of the original one for all the investigated events. In particular, the new approach for lateral inflows assessment allowed to significantly reduce the error on peak stage, whose mean absolute value decreased from 13.5 to 6.6%, and to increase the coefficients of persistence, E_v and E_r, as well as the Nash-Sutcliffe. As far as the error on time to peak is concerned, it was found out that the two model formulations are characterized by similar accuracy. The analysis on spatial storm distribution has shown that the forecast stage reliability is mainly affected by the ratio between the intermediate basin contribution and the upstream one, defined by the RRI index. In particular, when this index is lower than 1.0÷1.2 the model results are expected to be accurate, while when it becomes greater than 1.4 the provided forecast stage might be characterized by a not negligible degree of uncertainty. On these bases, the proposed model can be a useful tool within a real-time flood forecasting system for the Pierantonio-Ponte Nuovo reach provided that RRI is monitored in continuous during critical storm events. Further investigations on other river reaches with different hydraulic and geometric properties should be carried out in order to verify the new model formulation reliability. Moreover, it could be interesting to analyze the effect of different representation of rainfall spatial variability on the model performance.

ACKNOWLEDGEMENTS

The authors wish to thank B. Bani and R. Rosi for their technical assistance. They are also grateful to Umbria Region, Dept. of Soil Conservation, for providing most of the analysed hydrometric and pluviometric data.

REFERENCES

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z. & Smith, M., Towards improved streamflow forecasts: the value of semi-distributed modelling, *Water Resour. Res*., 37[11], 2749–2759, 2001.

Chow, V. T., Maidment, D. R. & Mays, L. W., Applied hydrology, McGraw Hill International Editions, 1988.

Franchini, M. & Lamberti, P., A flood routing Muskingum type simulation and forecasting model based on level data alone, *Water Resour. Res*., 30[7], 2183–2196, 1994.

Lamberti, P. & Pilati, S., Flood propagation models for real time forecasting, *J. Hydrol.*, 175, 239–266, 1996.

Michaud, J. D. & Sorooshian, S., Comparison of simple versus complex distributed runoff models on a midsize semiarid watershed, *Water Resour. Res.,* 30[3], 593–605, 1994.

Moramarco, T., Barbetta, S., F. Melone, F. & Singh, V.P., Relating local stage and remote discharge with significant lateral inflow, *J. Hydrol. Engng* ASCE, 10[1], 58–69, 2005.

Moramarco, T., Barbetta, S., Melone, F. & Singh, V.P., A real-time stage Muskingum forecasting model for a site without rating curve, *Hydrol. Sci. J.*, 51[1], 66–82, 2006.

Nash, J. E., & Sutcliffe, J. V., River flow forecasting through conceptual models, Part-1: A discussion of principles, *J. Hydrol*., 10, 282-290, 1970.

Kitanidis, P.K. & Bras, R., Real-time forecasting with a conceptual hydrologic model. 2. Applications and results, *Water Resour. Res.*, 16[6], 1034-1044, 1981.

O'Connell, P. E. & Clarke, R. T., Adaptive hydrological forecasting – a review, *Hydrol. Sci. Bull*., 26[2], 179–205, 1981.

O'Donnell, T., A direct three-parameter Muskingum procedure incorporating lateral inflow, *Hydrol. Sci. J*., 30[4/12], 479–496, 1985.

Segond, M.L., Wheater, H.S., Onof, C., The significance of spatial rainfall representation for flood runoff estimation: a numerical evaluation based on the Lee catchment, UK, *J. Hydrol*., 347, 116-131, 2007.

Kundzewicz, Z.W., Strupczewski, W.G., Approximate translation in the Muskingum model, *Hydrol. Sci. J*., 27[1], 19–27, 1982.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Field Measurement Network in the Mesoscale Catchment of the Gera River

M. Fink, M. Wetzel, D. Varga, S. Kralisch and W.-A. Flügel

Institute for Geography, Department of Geoinformatics, Friedrich-Schiller-University, 07737 Jena, Germany

Abstract: In many cases, applications of water-quality models have a poor basis of calibration and validation due to limited or lack of input data. Water-quality monitoring often is limited to a few sites within the stream network at which samples are collected and analysed at some intermittent frequency (e. g. monthly). In particular, this limited temporal sample coverage causes problems in modelling water-quality variables such as phosphorous that are mainly transported during short-time events. The resultant loads calculated based upon these data are often subject to large errors, due to this sparse time resolution.

In order to address this problem, a meso-scale catchment was equipped with a measurement network for obtaining spatially and temporally higher-resolution data. As test site, the upper catchment of the Gera River (approx. 850 km²) was selected for this study. The catchment features different landscape forms that are typical for middle mountainous regions in middle Europe.

The measurement network consisted of six online water-quality stations and 23 additional water-quality sample points distributed in the river network. At these monitoring sites, streamflow, temperature, pH-Value, electrical conductivity, turbidity, suspended solids, dissolved oxygen, nitrate, ammonium, total phosphorous, total organic carbon and other variables were measured or analyzed in the laboratory. In addition, eight weather stations and 24 soil-moisture sites were installed on farm land distributed throughout the catchment.

The goal of this study is to improve the existing components describing the hydrological, nitrogen loadings, and overall land-use management using the J-2000-S model with the measured data. These data will also help to implement an erosion-phosphorus component and a stream water-quality component for this particular model.

Keywords: Water Quality, Measurement Network, Modelling, Validation.

1. INTRODUCTION

Water-quality models for meso-scale catchments often are confronted with a poor data base for calibration and validation purposes. Water-quality monitoring is normally limited to a few points within the stream network at which samples are collected and analysed at some intermittent frequency (e. g. monthly). The resultant calculated loads are often subject to large errors, due to this sparse temporal resolution of the available data. This in particular is true in cases of erosion and sediment-linked contaminant transport, such as phosphorous loadings generated from diffuse sources (Dijkhuis et al. 1992). Erosion normally takes place during short-term events (thunderstorms or spring snowmelt runoff). Data with better spatial and temporal resolution are more commonly available for plots or for small catchments.

The results taken from such micro-scale measurements are not adequate for management decision-making, because most of these decision-making processes involve management

issues on a meso- or even macro-scale basis. The water-quality models used to aid in the decision-making process are often based on limited available physical data.

In order to attain higher spatial and temporal resolution data for model development, calibration and validation, a meso-scale test site was equipped within the scope of the Friedrich-Schiller-University-Jena's ILMS-Project (Integrated Land Management System). The design of this measurement network focuses on the diffuse nutrient loadings and nutrient transport in stream, due to the erosion processes occurring from agricultural lands.

2. UPPER GERA TEST SITE

As a test site for a so called field laboratory, we choose the upper catchment of the Gera River (approx. 850 km²). It is a typical meso-scale catchment located on the northern bound of the middle mountain range of the Thuringian forest and south to the city of Erfurt (Figure 1). The main streams in the catchment are the mainstem Gera River draining the central part from south to north, the Apfelstädt River draining the western part and the Wipfra River draining the eastern part of the catchment. The upper Gera River catchment was selected for this study, due to its heterogeneity of natural conditions and the land use. These conditions can be considered as representative for the State of Thuringia in Germany as well as large regions of middle Europe. The catchment itself features the following four specific landscape forms:

1. The "Thüringer Wald" (Thuringian Forest) (400 to 980 m a.s.l.) is underlain by rhyolite, conglomerate and sandstones. Most of the bedrock is covered with periglacial slope deposits which are important pathways for subsurface runoff. The soils developed from these deposits and the underlying bedrock range from leptosols on the steep slopes over dystric cambisols and podzols to fluvisols in the valleys. The Thuringian Forest area has distinct border to the other landscape forms in the catchment (Seidel, 2003). The land cover consists of coniferous forest with small areas of deciduous forest, pasture and meadows in the valleys. The average temperature in January is about -3 °C; during July, the value increases to 14 °C. The precipitation ranges up to 1200 mm a^{-1} (Hiekel et al., 2004).

2. The "Ilm-Saale-Ohrdrufer-Muschelkalkplatte" (Ilm-Saale-Ohrdrufer-Limestone-Plate) (300–600 m a.s.l.) primarily consists of underlying limestone, dolomite and marlstones, which locally exhibit karst features. The rendzic leptosols and mostly shallow phaeozem soils have been developed from limestone weathering and aeolic sediments. Most of the soils are shallow and exhibit limited field capacity. Due to this, there is little resistance to nutrient leaching. The texture of the soils is clayey to loamy. According to this, the erodibility is medium, while small spots show also a high erodibility depending on the amount of aeolic sediments in the soil (Ad-hoc AG Boden, 2005). The land use is a mixture of forest, pasture, meadows and crops. The average temperature is higher than in the "Thüringer Wald". In January, it reaches an average of about -1.5 °C and during July around 17°C. The precipitation is between 550 mm a^{-1} in the northern part and 750 mm a^{-1} at the southern border with the "Thüringer Wald" in the southeast (Hiekel et al., 2004).

3. The "Innerthüringer Ackerhügelland" (Central Thuringian Hilly Field Country) (200–400 m a.s.l.): The geology exhibits an alternating stratification of marlstones, limestones, claystones and sandstones. It is overlain by partly deep loess and fluviatil deposits. The loess deposits are the dominating source for soil development. The typical soil types are pheozens, luvisols cambisols, and rendzic leptosols. In the floodplain, gleyic pheozems, mollic gleysols and fluvisols are formed. Most of the soils are characterized by high field capacity and base saturation, resulting in relatively high soil fertility. Due to this high soil fertililty, the most of the land use consists of crops and only small areas of forests and pasture. The texture of the soils is silty to clayey where the loess deposits are dominating and sandier where the fluvial deposits have a higher influence on the soil formation. According to this, the erodibility of the soils are very high to medium (Ad-hoc AG Boden, 2005). The average temperature in January is approximately -0.5 °C and in

July around 17 °C. The mean annual precipitation is relatively low for middle European conditions and ranges between 500 mm a^{-1} and 600 mm a^{-1} (Hiekel et al., 2004).

4. The "Paulinzellaer Buntsandstein Waldland" (Paulinzellaer Sandstone Woodland) (400–600 m a.s.l.): The geology is dominated by sandstones with minor inclusions of marlstones, claystones, and glacial and fluvial deposits. Due to the type of bedrock, the soil texture is dominated by the sand fraction. The main soil types are cambisols, podsols, and small areas of luvisols. In the valleys are formed sandy fluvisols and gleysols. Due to the sandy soil texture, the soils have a limited field capacity, a low-to-medium soil fertility and a medium erodibility (Ad-hoc AG Boden, 2005). The land use is a mixture of coniferous forests, crops, pasture and meadows. The average temperature for January is approximately -2 °C and for July approximately 15 °C. The mean annual precipitation ranges from 550 mm a^{-1} to 800 mm a^{-1} at the border to the "Thüringer Wald" (Hiekel et al., 2004).

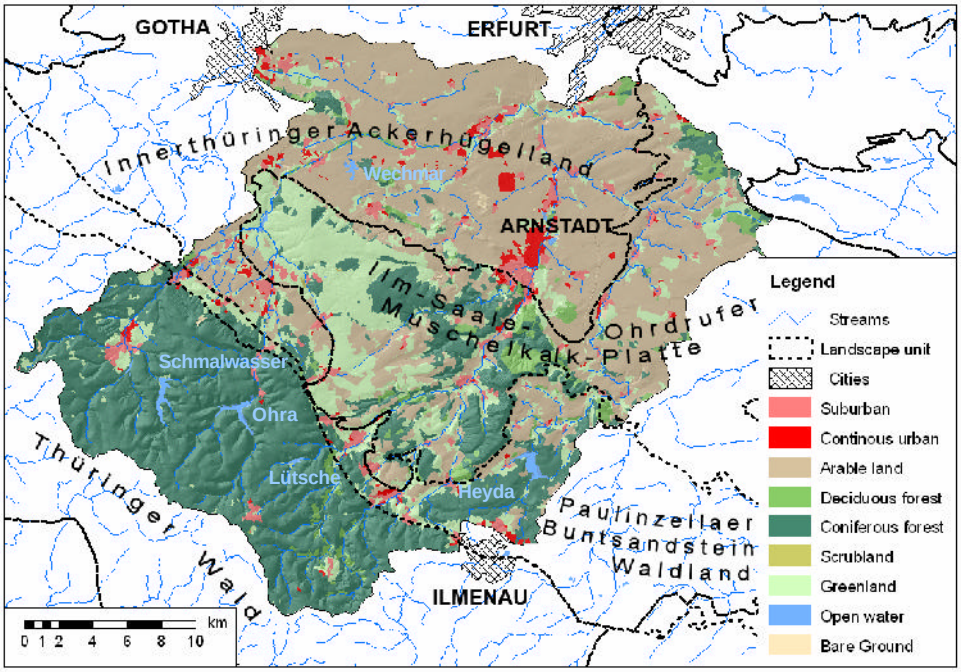


Figure 1. Landscape forms and land-use types in the upper Gera catchment.

In the entire catchment the pattern of erosible precipitation storm events during the year has only a slight seasonality with a maximum in July an August.

3. MEASUREMENT NETWORK

The configuration of the observation network, established in this field-laboratory case study, was designed to delineate the characteristics from each of the landscape forms. In addition, a nested catchment approach was implemented (Figure 2).

3.1 Water-quality stations

A key component of the field-laboratory design consists of six online water-quality stations distributed in the river network of the test site. The measurement devices were manufactured by WTW (Wissenschaftlich Technische Werkstätten GmbH). All of the water-quality stations measure the following parameters: streamflow, water temperature, pH, electrical conductivity, turbidity, suspended solids, dissolved oxygen, nitrate, ammonium, total phosphorous, redox-potential and spectral absorption coefficient (to correlate with carbon

parameters such as total organic carbon and chemical oxygen demand). The pH, electrical conductivity, and redox-potential are measured with electro-chemical sensors. Ammonium is measured with an ion selective sensor, compensated against potassium. Within the same device also a nitrate sensor is integrated without a specific compensation. Turbidity and suspended solids are measured with infrared optical sensors. The spectral absorption coefficient is measured with an ultraviolet optical sensor. This sensor also measures nitrate. For this optical sensor a filter system was installed to provide a solid-free and bacteria-free sample. The dissolved oxygen is measured with an optical sensor on the basis of the photoluminescence method. Total phosphorous is measured with a photometric sensor using the molybdenum blue method. In advance of the photometric analysis, a digestion that converts phosphorous compounds to orthophosphate takes place.

The temporal resolution was set at one recorded value per hour in the case of the phosphorous analyzer; whereas, the other variables are recorded every 15 minutes. The intended start of the regularly data recording is in June 2008.

Within the designated subcatchments, the locations of the water-quality stations are indicated in Figure 2 (black solid lines). The locations of the water-quality stations were selected to monitor the influences of the different landscape units on the water-quality parameters. The upper northern station ("Möbisburg") monitors conditions at the outlet for the entire Gera catchment. Located nearby to the west of "Möbisburg", the station "Ingersleben" collects the values of the largest tributary of the catchment, the Apfelstädt River, comprising three landscape forms 1, 2 and 3. West of the station "Ingersleben", the station "Wandersleben" was installed. This station monitors the values of the Rot stream, which has in its source area only the landscape form 3. For the mainstem Gera River, the station "Arnstadt" has been equipped (Figure 2). This station provides data from landscape forms 1 and 2. South-east of the station "Möbisburg" is located the station "Eischleben", measuring the values from the second largest tributary, the Wipfra River. This river catchment comprises landscape forms 4, 2 and a minor part of 3. The station "Niederwillingen" at the Wipfra River was installed, because it only comprises landscape form 4. Using the combined data for stations "Niederwillingen" and "Eischleben" the characteristics for landscape form 2 can be delineated. The spatial distribution of the water-quality stations in general strives to obtain distinctive characteristics for the four different landscape forms. However, the values of the landscape form 1 are not measured separately, because agricultural land use is very limited in landscape form 1. Therefore, only low concentrations of nutrients are found in the sampled water.

The stations "Möbisburg", "Ingersleben", "Eischleben" and "Arnstadt" were installed at already existing gauging stations operated by the Thuringian environmental agency. The stations "Wandersleben" and "Niederwillingen" are installed in field containers.

3.2 Water-quality sampling sites

Additional to the water-quality stations, 23 sample points throughout the catchment's stream network are selected. These points are sampled monthly to increase the spatial resolution of the water quality measurements. The locations of this sample points are chosen, so that the influence of single branches within the river network as well as the influence of landscape unit 1 can be estimate. This provides a total of 29 *in-situ* sampling locations, including the six water-quality locations were also water samples were taken. This program has been conducted to achieve a better linkage between the data of the sample points and the data of the water-quality stations and to test or confirm the online measurements independently. The subcatchment of the water quality station Wandersleben (Rot Creek) in the northern part of the test site has a higher measurement point density due to its high percentage of arable land. The goal is to investigate this part of the catchment more intensively.

At these sample points, similar parameters are measured compared to those at the water-quality stations. Water temperature, pH, electrical conductivity, turbidity, dissolved oxygen and redox-potential are measured with handheld instruments directly at the sample points. Chemical oxygen demand, nitrate, nitrite, ammonium, total phosphorous, organic carbon, potassium, magnesium, and calcium concentrations are photometrically measured with cell

tests in the laboratory. The samples are taken on one day, on the next day the laboratory analysis is conducted. The sensors and cell test are manufactured by Hach-Lange GmbH.

3.3 Meteorological stations

A primary focus of this study is on processes associated with storm-event based characteristics. Therefore, climate data with a high temporal resolution are necessary and nine meteorological stations (Feingerätebau K.Fischer GmbH) have been installed for this study. The locations of those stations were selected to obtain data for the various landscape forms. The landscape form 1 ("Thüringer Wald") is only equipped with one station, because of the relatively greater focus on processes in agricultural land in this study.

For this type of station, locations were sought to the extent possible in enclosures to protect the stations from vandalism. The station towards the northeast just outside of the catchment area is also used for another part of the ILMS-project which is not described in this article, but it gives additional information about the test-site area.

The measured parameters are air temperature, precipitation, wind speed, wind direction, air pressure, humidity and radiation balance. These data are being recorded every 5 minutes. A GSM-modem is integrated to provide remote access to these data.

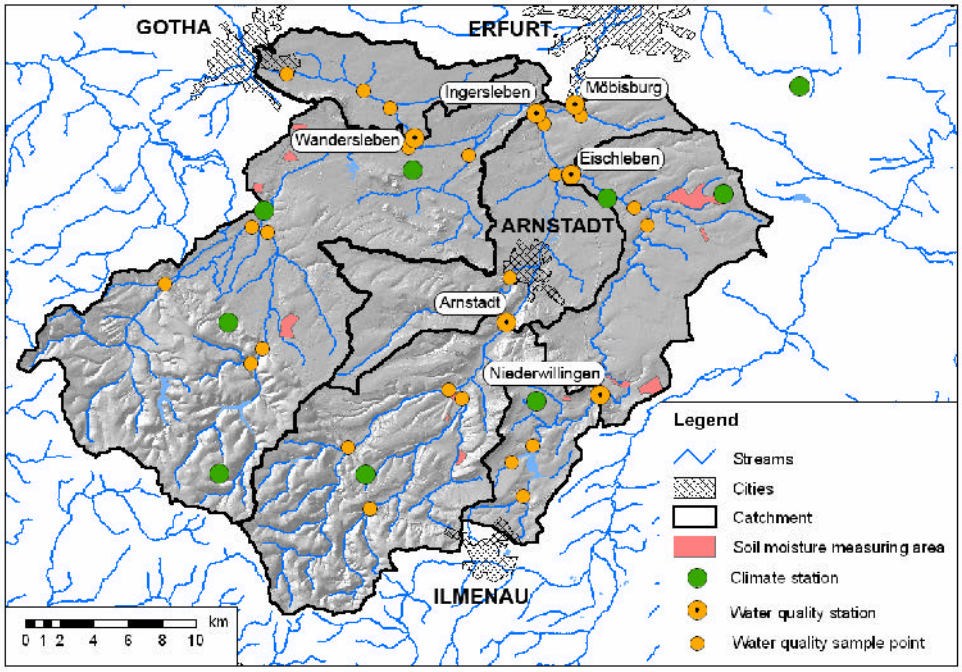


Figure 2. Measurement network of the upper Gera catchment.

3.4 Soil-moisture stations

For the description of the hydrological balance, soil-water storage is an important parameter. It is of vital importance for estimating evapotranspiration. The condition of this storage term has also a significant influence on the distribution of water in the different runoff components – surface runoff, interflow and percolation.

We measure soil moisture with frequency-domain probes (Sentek EnviroSMART) at five different depths at a total of 24 locations on farmlands distributed throughout the different landscape forms. The goal of those measurements is to measure water infiltration into the soil and to derive information about surface-runoff events. The values are stored every 5 minutes. The data will be transferred via a remote-telemetry network to a central station which then can be transferred through modem connection to the central database.

4. FIRST MESUREMENT RESULTS

The first field-measurement survey to collect water samples at the various water-quality sample points began in January 2008. Selected results of this initial survey are given in Figures 3 and 4 for calcium and nitrate, respectively. In Figure 3, the areal distribution of calcium concentrations are indicated. These concentrations reflect to some degree the natural background of the different properties of the soils as well as of the underlying bedrock (except from sources of calcium fertilisers which have been applied over large areas of the catchment). The low calcium values in landscape form 1 (sample points a5, o2, o3, g9 and g10) are a result of the siliceous lithology and the acidic soils. Downstream of the Apfelstädt River (a4, a3 and a1), the concentrations rise, due to the limestone and loess deposits in the downstream parts of the catchment. This effect could be clearly seen in the tributaries of the Apfelstädt River (o1, r1 – 4 and a2) which have higher calcium concentration then the mainstem stream. The Gera River (g7, g8, g4, g3 and g1) shows basically a similar situation to that for the Apfelstädt River. The values are higher, because of a lower part of the landscape form Thüringer Wald which contributes water with low calcium concentrations. The Wipfra River shows slightly higher values in the headwater in comparison to the other catchment streams (w5, w6 and w7). This might be caused by the different properties of the landscape form 4 in the Wipfra River and form 1 in the other stream areas. Sample point w4 indicates a higher value than the points in the headwater, despite the fact that this point is still in the same landscape form. One reason could be that the higher percentage of agricultural land in this area which receives more fertilisers with calcium than the forested areas. The increase of the calcium concentration downwards the river (w1, w2 and w3) is caused by the limestone-dominated landscape form 2 similar to the increase in the other catchment streams.

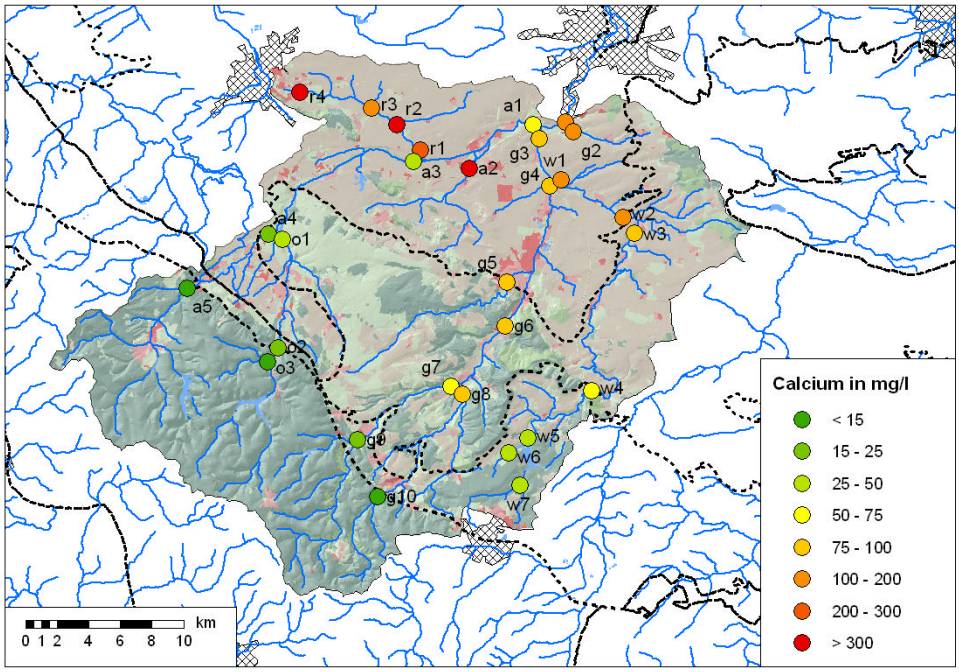


Figure 3. Calcium concentrations at the water quality sample points, 17 January 2008.

In Figure 4, the nitrate concentrations for the water-quality sample points are shown. The concentrations reflect the influence of the land use more explicitly than the calcium concentrations (Figure 3). In the forested area, the nitrate concentrations at the sample points (a5, o2, o3, g9, g10, w7, w6 and w5) are all quite low. For this region, the nitrate is originated mainly from atmospheric deposition.

At the reservoir outlet w5 the nitrate concentration is lower than in comparison to the inflows w6 and w7. This could be caused by denitrification processes or by dilution within the reservoir. At w4, the concentration increases again and indicates the influence of the

increased percentage of arable land. With the measured nitrate concentrations at w3, w2 and w1, this effect could be seen more clearly. In the Gera River, the points g8, g7 and a6 exhibit a moderate increase of nitrate concentrations. The higher concentration at g5 is caused by very shallow soils in its tributary and the result of no water from landscape form 1 diluting the water in the Gera River (cf. Figure 2). The points g4 and g3 also indicate relatively high nitrate concentrations, because of greater level of intensively used agricultural land. The Apfelstädt River has at a4, a3 and a1 comparably low values. The large portion of the landscape form 1 is the reason for relatively low values in the Apfelstädt. The smaller creeks in landscape form r1 – r3 and g2 show the highest concentrations in the entire catchment. In the catchments of these tributaries the portion of arable land is very high and is causing this high nitrate concentration.

The data results indicated for January 2008 (Figures 3 and 4) are quiet similar to results from the more recent sampling survey conducted in February 2008. This would indicate that the above-stated results have a certain degree of reliability for catchment characterization.

The results show that the most important influences of landscape on the water quality could be identified with the introduced water quality network in a sufficient manner.

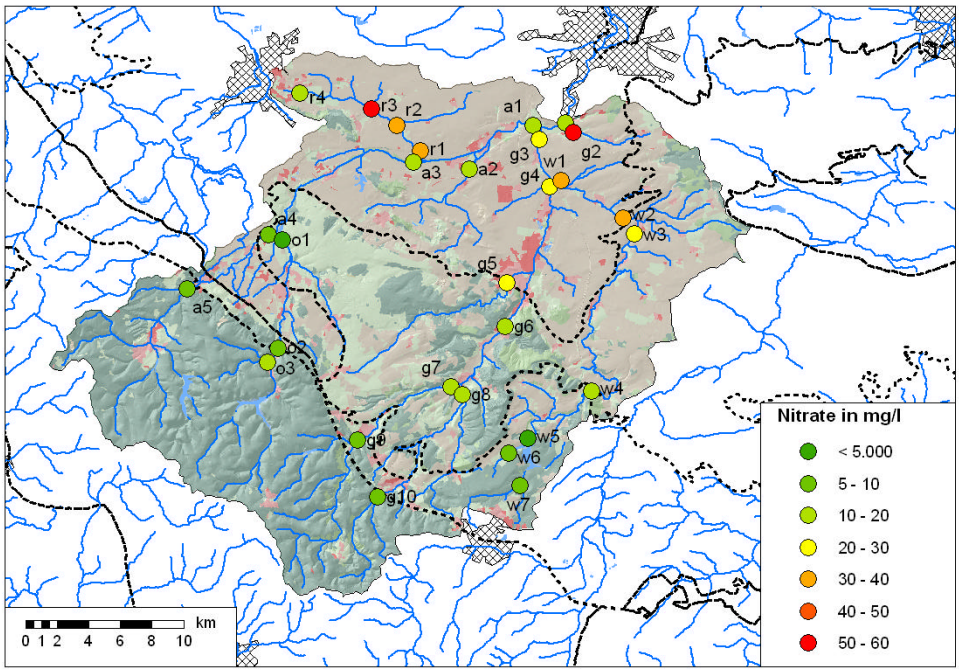


Figure 4. Nitrate concentrations at the water-quality sample points, 17 January 2008.

5. CONCLUSIONS

The data gathered from this catchment's measurement network will be used in the ILMS-Project for further model development and validation using hydrological and nutrient transport models. These data will be in particular used for the J-2000-S nutrient transport model (Bende-Michl et al., 2006, Fink et al., 2007), an extension of the J2000 hydrological model (Krause, 2001; 2002). In addition to the hydrological part of J2000, methods were implemented for the description of the land-use management, plant growth, and nitrogen balance. These methods are extracted from the Soil Water Assessment Tool (SWAT) model (Arnold et al., 1998).

In the ILMS-Project, we seek to improve the existing components hydrological, nitrogen and land-use management of the J-2000-S along with the gathered data. Additionally, these data will help to implement an erosion-phosphorus component and a stream water quality component. We expect from the use of these field data a substantial improvement of the quality and the reliability for our process-modelling system and associated applications.

ACKNOWLEDGEMENTS

The authors wish to thank the German Federal Ministry of Education and Research for funding this work in the course of the program ”Entrepreneurial Regions – the BMBF Innovation Initiative for the New German Länder.

REFERENCES

Ad-hoc AG Boden. Bodenkundliche Kartieranleitung, 5th Edition, edited by: Staatliche Geologische Dienste and Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, 2005.

Arnold, J., Srinivasan, R., Muttiah, R., and Williams, J., Large Area Hydrologic Modelling and Assessment. Part I, Model Development, J. Am. Water Resour. Assoc., 34(1), 73–89, 1998.

Bende-Michl, U., Krause, P., Kralisch, S., Fink, M., and Flügel, W.-A., Current development and application of the modular Java based model JAMS to meet the targets of the EU-WFD in Germany, Proceedings of the iEMSs Third Biennial Meeting: “Summit on Environmental Modelling and Software”, edited by: Voinov, A., Jakeman, A., and Rizzoli, A., International Environmental Modelling and Software Society, Burlington, USA, 2006.

Dijkhuis, A. H., N. K. Hoekstra, L. Lijklema and S. E.A. T. M. van der Zee, Origin of peak concentrations of phosphate during high discharges in a rual watershed. In: Hydrobiologia 235/236, S. 257 – 260, 1992.

Fink, M., P. Krause, S. Kralisch, U. Bende-Michl and W.-A. Flügel, Development and Application of the Modelling System J2000-S for the EU-Water Framework directive, Advances in Geosciences 11, p. 123 – 130, 2007.

Hiekel, W., F. Fritzlar, A. Nöllert and W. Westhus, Die Naturräume Thüringens, Naturschutzreport 21, Jena, 2004.

Krause, P., Das hydrologische Modellsystem J2000; Schriften des Forschungszentrums Jülich, Reihe Umwelt/Environment, Band 29, Jülich, 2001.

Krause, P., Quantifying the impact of land use changes on the water balance of large catchments using the J2000 model, Phys. Chem. Earth, 27, 663–667, 2002.

Seidel, G. (Ed.): Geologie von Thüringen, Stuttgart, 2003.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Insights into catchment behaviour for water harvesting assessment in the East India Plateau

B. Croke[a], S. Norrish[b], A. Kumar[c], A. Islam[d], P.Dey[e], P. Cornish[b], S. Kumar[e], J. Ghosh[e]

[a] *Department of Mathematics, and The Fenner School of Environment and Society, The Australian National University, Canberra ACT 0200, Australia (barry.croke@anu.edu.au)*

[b] *University of Western Sydney, Australia*

[c] *PRADAN, Purulia, West Bengal, India*

[d] *ICAR-RCER, Patna, Bihar, India*

[e] *HARP, Ranchi, Jharkhand, India*

Abstract: As part of an Australian Centre for International Agricultural Research (ACIAR) funded project investigating water harvesting techniques in tribal villages in West Bengal, India, a network of monitoring sites has been set up in the ~2 km^2 Headwater sub-catchment of the Chapai Nula (referred to as the Pogro sub-catchment after one of the villages located in the sub-catchment), with a smaller 0.5 km^2 sub-catchment operating as a control. During the wet season, there is a significant excess of rainfall (~300 mm in 2006), a significant fraction of which remains in the catchment. Consequently, there is potential for planting of dry season crops in low-land parts of the catchment. Due to the high rainfall intensity (up to 240 mm/hr), low infiltration rates (<40 mm/hr) and low vegetation density in the upland areas, a significant fraction of the rainfall runs off to the lower areas of the catchment. The monitoring has also shown that the ponds in the Pogro study site are porous, and act as a source of recharge the groundwater system (typically 5-10 mm/day) as well as being a surface storage.

Keywords: Water harvesting; Hydrologic response; East India Plateau; NGO.

1. INTRODUCTION

Watershed development (WSD) is a major strategy for livelihood improvements, but often the strategy is not underpinned by sound hydrology, both in terms of the local and downstream impact of water harvesting structures. As part of a project in the high rainfall East India Plateau evaluating existing water harvesting techniques employed by PRADAN (an Indian NGO), a program of monitoring the water storages and fluxes in a small (2 km^2) catchment located in the Damodar sub-basin of the Lower Ganga Basin. The study sites are in a tribal village region of the Purulia district of West Bengal. Currently, in many fields only one wet season crop is grown each year, with the male villagers migrating in the dry season to find work. The goal of the WSD is to enhance the capacity for the villagers to grow an additional crop using the residual moisture available after harvesting the wet season crop. The aim of the hydrology component of the project is to evaluate the impact (including out-of-catchment) and effectiveness of the water harvesting techniques used by PRADAN, and explore possible means of increasing the efficiency of the techniques. The monitoring strategy is mostly focused on using techniques that are available to PRADAN staff so that they may continue to monitor the impact of their activities have in the region. To this end, some of the data (piezometers, wells, ponds, streamflow gauges and the manually read weather station) are being recorded by one of the villagers. In addition to the hydrological monitoring work, the project also focuses on potential agronomic improvements, and uses a participatory action-learning approach. The focus of WSD carried out by PRADAN is on small, distributed works that are aimed at assisting the owner of each field, and are hence focused on surface water and shallow groundwater

resources. While the focus is on distributed WSD, the interactions between interventions are also considered in the design of the development of a catchment.

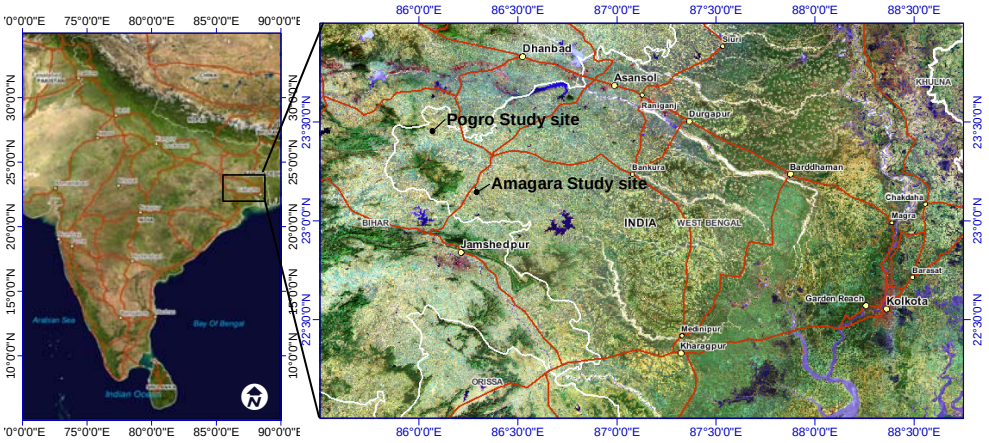


Figure 1. Location of study area

2. STUDY SITES

There are two sites under study in the project. The main study site (Pogro) is an approximately 2 km^2 sub-catchment (see Figure 1) that has a well defined discharge point at a culvert under the railway line that forms the eastern boundary of the sub-catchment (see Figure 2). There is a second culvert which is the outlet from a smaller sub-catchment to the north-east of the study site, and this is being used as a reference. The site is called the Pogro study site after one of the villages located in the sub-catchment. The second study site (Amagara) is an approximately 1 km^2 sub-catchment (see Figure 1). PRADAN has carried out a considerable amount of watershed development in this sub-catchment, but due to the lack of a defined outlet, this sub-catchment is not suitable for monitoring the hydrological response. Subsequently, the Amagara study site is being used as a study site for the agronomic and participatory aspects of the project. The Pogro study site drains into the Panchet Dam on the Damodar River, while the Amagara site drains to the Kangsabati Dam (Kangsabati River). The climate of the region is monsoonal, with an average rainfall of approximately 1.3m falling predominantly in the wet season (July to September).

Figure 2 also shows the different land classes within the Pogro study area, and the digital elevation model (DEM) derived from the differential GPS data collected during the electromagnetic survey. The land classes are Bohal (lowland - located in the drainage areas of the sub-catchment), Kanali (mid-lowlands), Baid (mid-uplands – main agricultural land owned by the local villagers) and Tarh (uplands – mostly grazed land). The interpolation of the data points was constrained using streamlines also mapped as part of the data collection process. Due to the patchy nature of the data, the use of differential GPS data and the highly modified structure of the landscape, there is considerable uncertainty in the derived

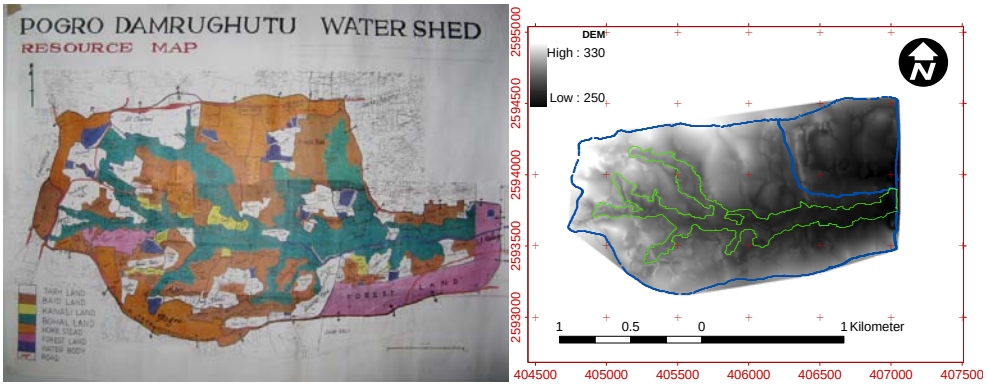


Figure 2. Resource map for the Pogro study site (showing the distribution of land classes and water bodies), and DEM showing elevation in metres.

DEM. Areas with a higher density of points have a smaller uncertainty than more sparcely covered regions, but the influence of extensive bunding in the area introduces error in the interpolation as this assumes a smoothly varying land scale and does not take into account the presence of terraces.

PRADAN employ a range of water intervention techniques intended to work locally as well as in conjunction with other interventions. The aim is to maximise the impact of the interventions while reducing the volume of earth work involved. In the Tarh land, the focus is on reducing runoff and increasing infiltration (using a combination of bunds and pits/trenches), thereby increasing the availability of water for agriculture (e.g. agroforestry and fruit orchards). In the Baid land, pits are used to store water or increase recharge (5% of paddy used for pits). In the wetter parts of the catchment (i.e. Bohal and Kanali land), the focus is on improving access to the subsurface water through use of seepage pits. The focus of this paper is on monitoring the hydrologic response of the study sites, and detailed information of individual interventions will not be given.

3. MONITORING NETWORK

Monitoring of the water fluxes within the Pogro study site has been done at the point scale as well as mapping the variation in conductivity across the catchment (see Figure 3). In addition to water fluxes, measurements of the soil properties (both hydrologic and agronomic) have been made. The initial goal of the monitoring work is to gain understanding of the hydrology of the site, as well as generating the data needed for the development of hydrological models.

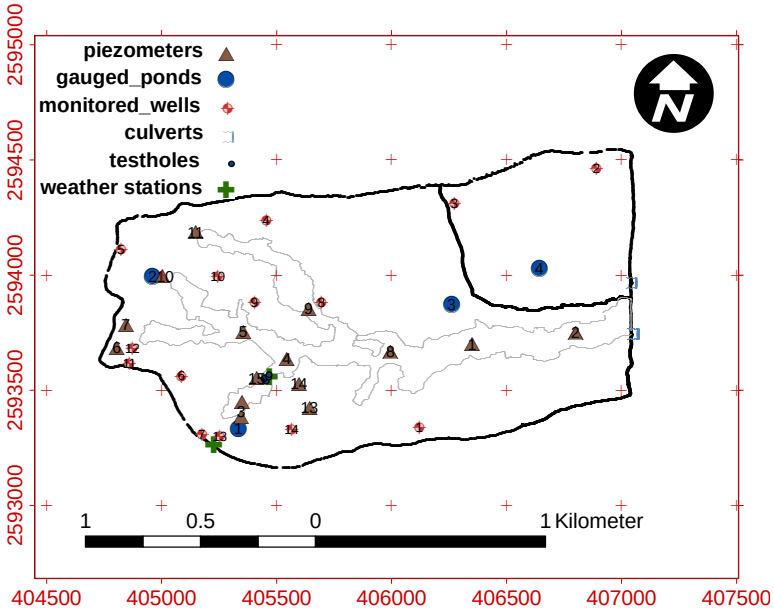


Figure 3. Monitoring of water resources in the Pogro study area (UTM coordinates, with catchment boundaries (black) and bohal area (grey).

3.1 Surface water

Both culverts at the outlets of the sub-catchments in the Pogro study area have had staff gauges installed since 2005, with water level recorded daily by one of the villagers (more frequently during rain events). To date, rating curves for the culverts have not been available, though water level loggers have been installed in

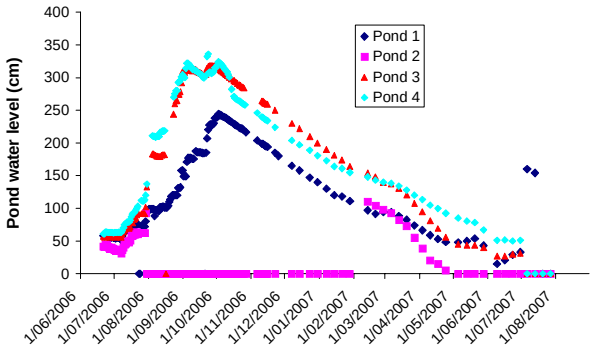


Figure 4. Water level in the 4 monitored ponds.

both culverts, and a campaign to collect rating curve data is underway, permitting closure of the water balance for the study site.

In addition to the culverts, water levels in 4 ponds in the Pogro study area are being manually recorded by one of the trained villagers on a daily to fortnightly basis. Three of these ponds lie inside the main study sub-catchment, the other is in the reference sub-catchment. The observed water levels for the 2006-2007 hydrological year are shown in Figure 4. Two of the ponds (pond 2 and 3) show a rapid decrease in water level from mid-March suggesting a significant increase in the net discharge rate; either a decrease in the inflow from subsurface sources, or an increase in discharge due to declining groundwater levels in the receiving aquifer (unfortunately, the staff gauge for pond 2 was broken at the start of the wet season and repair was not possible until February 2007). The other ponds do not show a significant variation in the rate of decline in the water level. Before March 2007, the rate of decline for ponds 1, 3 and 4 was roughly 11 mm/day, increasing for ponds 2 and 3 from mid-March to approximately 20 mm/day.

3.2 Groundwater

There are 14 wells (see Figure 3) in the Pogro study site for which water level is being recorded on a weekly to fortnightly timescale. The wells are typically about 8 m deep, with a variation in water level through the year of between 4 and 7 m. In the north of the catchment, there are some tube wells that had been dug to a depth of about 80 m without finding a significant source of water. These data give information on the behaviour of the deeper groundwater systems, which is still important in terms of understanding the hydrology of the study site, even though the focus of the interventions is on the shallow groundwater systems.

A network of 15 piezometers have been installed in the Pogro study site, though there has been continued interference with the piezometers (from filling with rocks and branches to removal of the pipes) so the number of active sites at any time is typically less than this. The piezometers are mostly located in the Bohal regions of the sub-catchment, and monitor the shallow groundwater levels. A line of piezometers extends from the Tarh land in the south of the sub-catchment, through a region of Baid land to the Bohal land in the middle of the sub-catchment.

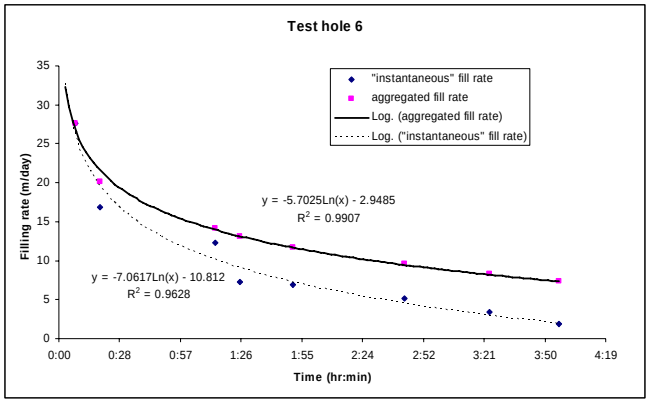


Figure 5. Filling rate for test hole 6

In addition to the piezometers, a number of test holes have been dug in both study sites to rapidly assess the water level at a number of sites. A total of nine test holes were made in Amagara during April 2006. While the one test hole located in the upper part of the sub-catchment was dry, all other test holes at least partially filled with water over a period of time, ranging from 1 hour to fill to within 10 cm of the surface, to requiring up to 24 hours to fill to within 40 cm of the surface. Five test holes were more intensively monitored over a period of a few hours (e.g. see Figure 5), with the filling rate being determined with respect to the preceding data point ("instantaneous" fill rate) and with respect to the first data point (aggregated fill rate). This shows that the shallow groundwater system in this part of the catchment has a very low transmission rate, enabling the system to maintain a supply of water throughout the dry season. These test holes were downhill from a pond located near the Amagara village, and indicate that the pond is likely to be the dominant source of shallow groundwater. This hypothesis is reinforced by observations in the Pogro study site, where all of the Bohal areas extend up towards the main ponds in the catchment, indicating that these are a significant source of recharge to the shallow groundwater system.

The data for Amagara the test holes suggests that, given the conditions at the time (late dry season, with only a few seepage pits) a square pit 3 m wide should fill to within 0.5 m of the surface in approximately 1 day assuming that the infill rate will be proportional to the perimeter (i.e. ignoring the influence of depletion of the aquifer) and that the aquifer extends to the base of the pit. With increased use of the water resource, the available water will be reduced, and the fill times increased. In the extreme case, the pits will dry up. Evaluation of the volume of available water is not possible at this time as the extent of the storage and the effective porosity is not yet known.

3.3 Electromagnetic survey

Multiple electromagnetic induction (EMI) surveys of sub-surface apparent conductivity (ECa) have been undertaken, employing Geonics® EM31 and EM38 instruments. The surveys helped to characterise subsurface properties and some of the variation in soil water content, which had agronomic significance for cropping options. Measurements with both EMI instruments were taken in transects across the catchment and geo-referenced with a differential GPS. EMI surveys were undertaken after the monsoon (November/December), which was assumed to coincide with the wettest state of the catchment. Surveys were also undertaken at the end of the dry season in May/June. ECa surveys were made with the EMI instruments in the vertical mode of operation, measuring to an integrated signal of subsurface conductivity to an approximate depth of 1.5 and 6.0 m for the EM38 and EM31 respectively [McNeill, 1980]. Spatial variability in differences in conductivity between the wet and dry season was evaluated for identifying areas with potential for growing crops after the monsoon. The catchment appears to have very low background electrical conductivity. The highly disturbed nature of the landscape and depth of soil profiles tended to be the primary determinant of variation in EMI signal.

EMI measurements with both instruments were also taken at or near the piezometers during the wet and dry periods. Measurements were made in both the vertical and horizontal modes at 0, 0.5, 1.0 and 1.5 m above the ground for the EM38 and at 0, 1.0, 2.0 and 3.0 m for the EM31. Differences in ECa with height and instrument mode of operation were used for indicating height to water table. Inconsistent results between piezometer sites were assumed to reflect variation in profile depth to underlying strata.

Further EMI measurements were carried out in individual paddy fields that were representative of areas (‘patches’) that local farmers had differentiated in the catchment. Measurements in the wet and dry seasons were made to determine if variation in ECa correlated with patch differentiation by farmers. A change in ECa between seasons was also assumed to be correlated with soil water content in each monitoring site. The degree of

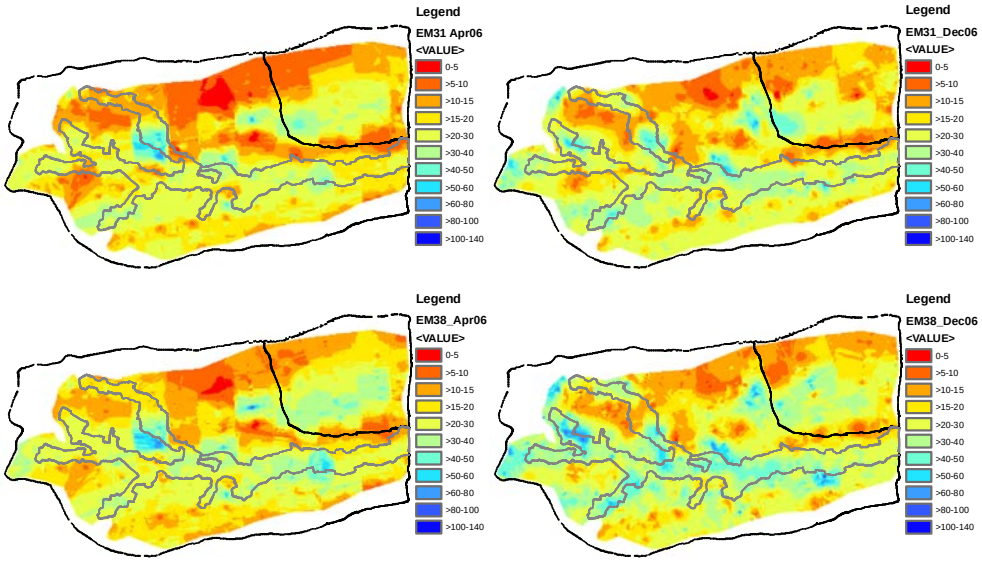


Figure 6. Interpolated EMI data for Pogro (units are mS/m). Images on the left are from April 2006 (late dry season), while the right side correspond to December 2006 (early dry season). Top images are from the EM31, bottom images from the EM38.

change in seasonal ECa was different between patches. Some sites showed large differences in mean ECa between seasons, which appear to be strongly correlated with soil water content. In other monitoring sites, there was very little difference in mean seasonal ECa between seasons. This appeared partially due to the very large differences in ECa within single paddies in both the wet and dry seasons. The large intra-paddy ECa variation was most evident in the upper parts of the landcape. ECa in the lower part of the landscape (bohal and kanali) tended to be more consistent across a single paddy.

The observations of the conductivity were interpolated using kriging (see Figure 6) as well as thin plate smoothing splines [Hutchinson, 1993]. Figure 6 shows the spatial and temporal variations in the conductivity observed by the two instruments, with higher conductivity typically found in the lowland (Bohal) areas, as well as increased conductivity in the early dry season (December) compared to the late dry season (April).

3.4 Soil data

As part of the study, thirteen pits (selected based on a participatory landscape assessment involving both scientists and farmers, with sites randomly selected to represent all land classes (Figure 2) so that the information obtained could be readily used by the farmers) were dug to study the soil profiles, pH, organic carbon, nitrogen and phosphorus in the Pogro catchment. To supplement this, soil pH and fertility measurements have been made for sites involved in agronomy studies. The soil profile descriptions, coupled with the derived DEM have been used to map soil types in the Pogro study sub-catchment.

3.4.1 Infiltration experiments

A series of infiltration experiments were conducted in both the Pogro and Amagara sub-catchments using a double ring infiltrometer (e.g. see Figure 7). In Pogro, infiltration rates in paddy fields located in the Baid areas of the catchment were found to have saturated hydraulic conductivities of about 1 mm/hr or less. This is likely to be an upper limit of the infiltration rate as the evaporative loss may have been significant (the experiments were covered to reduce the evaporative loss). In comparison, experiments conducted in the Tarh land gave K_{sat} rates of up to 40 mm/hr, reflecting the lower clay content of the soils in the Tarh land. The infiltration was also measured at the base of one of the soil pits in the Tarh land, with a rate of 4 mm/hr found at a depth of 1.5 m, while surface rates near the pit were 27 and 15 mm/hr. If these rates are indicative of the area, then there is potential for a subsurface flow down the slope at a depth of about 1m.

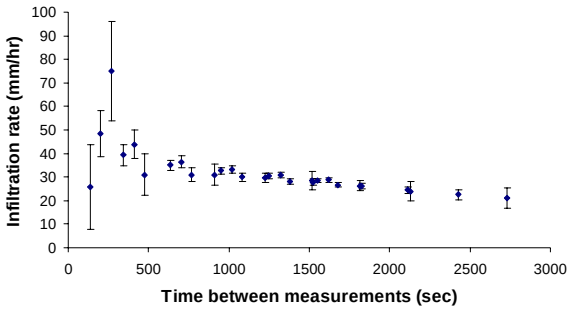


Figure 7. Infiltration rate for the first experiment in Amagara.

The experiments in Amagara were mostly conducted in the Tarh land to the north-west of the village. The infiltration rate in the treated area (bunded 9 by 12 m plots, each with a 3 m^2 pit designed to capture runoff and increase infiltration) was found to be approximately 2.4 mm/hr, while values of 3, 8 and 17 mm/hr were found in the untreated Tarh lands. An experiment was also conducted in a field that had been planted with mustard and recently irrigated. The infiltration rate in the field was found to be 21 mm/hr.

3.5 Climate data

An automatic weather station (AWS) has been installed in the Pogro village, giving minutely values of temperature, relative humidity, wind speed, rainfall and solar radiation. In addition rainfall, minimum and maximum temperature, as well as wet and dry bulb temperature are manually collected as a cross check of the AWS data, as well as providing information if there is a problem with the collection of data from the AWS.

3.5.1 AWS data

The AWS records minutely values of the rainfall, average and maximum wind speed, solar radiation, air temperature and relative humidity. The resolution of the rain gauge is 0.2 mm, and the maximum rainfall intensity recorded is 3.6 mm/min, or 216 mm/hr. Figure 8 shows the rainfall intensity exceedence curve derived using the raw data, and the data resampled to a time step of one hour. The hourly data underestimates the rainfall intensity by an order of magnitude. Modelling the response of the catchment using hourly data will result in significant underestimation of runoff for most sites within the study catchment as the maximum hourly rainfall is a little more than 10 mm/hr. Figure 8 also shows the relationship between peak rainfall intensity within an hour period and the corresponding hourly rainfall. While there is considerable scatter in the data, typically, the peak rainfall intensity is approximately 5 times larger than the hourly rainfall amount.

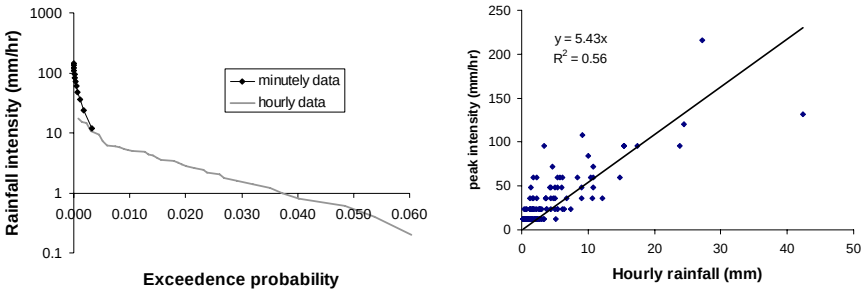


Figure 8. Rainfall intensity exceedence curve, and relationship between hourly rainfall and peak rainfall intensity

The potential evaporation has been estimated using of a model for estimating the short-wave and thermal energy budget [Vardavas, 1987]. Preliminary results give a lake evaporation rate of approximately 8 mm/day in August, decreasing to about 3 mm/day in November. During the 2006 wet season (below average rainfall) the rainfall exceeded the potential evaporation by approximately 300 mm (or about 500 ML). While measurements of the streamflow out of the catchment are not currently available, the recorded water levels at the culverts indicate that most of this residual water remained within the catchment (in ponds, groundwater and in the soil). Soil moisture measurements show that a significant fraction of this water remains in the soil profile, providing a significant water source for rainfed aor supplementary-irrigated dry season (rabi) crops.

4. MODELS

Modelling will be carried out at three scales in the project: intervention, sub-catchment, and catchment scales. Models at the intervention scale will be mostly process-based models with the focus being on understanding the behaviour of individual interventions (or a small number of interventions) in terms of their influence on the flux of water in the immediate vicinity. Initial models have been developed in the spreadsheet application Excel to facilitate adoption by PRADAN.

The sub-catchment scale corresponds to the entire study site (~2 km^2), and is focused at the combined effect of a suite of interventions distributed across the study site. The discharge from the study site is through the culverts under the railway line which defined the lower boundary of the catchment. The second culvert will be used as a comparison (no further WD work will take place in this catchment) in order to help overcome the limited period of data. The models used at the sub-catchment scale will be IHACRES [Croke *et al.*, 2006], PRMS [Leavesley and Stannard, 1995] and WASIM-ETH [Niehoff *et al.* 2002]. These models cover a broad range of modelling approaches, from parsimonious, data-based modelling (IHACRES) to more physics-based, distributed modelling (WASIM-ETH), enabling the impact of the availability of spatial datasets to be assessed through comparison of the outputs of the models.

The modelling at the catchment scale (~100 km^2) is focused on estimating the impact of watershed development across a broad area on down stream users. This will be simulation modelling only as there will not be observations made at this scale in the vicinity of the

study sites. Due to a lack of local streamflow gauges, storage data for 3 neighbouring dams (catchment areas range from 20 to 100 km^2) will be used for testing models applied at the catchment scale near the study sites. Streamflow data for a small number of catchments >~100 km away will be used for testing of models in the region. The IHACRES and PRMS models will be used at the catchment scale (data for the WASIM-ETH model is not available). The lack of gauged streamflows means that a regionalisation approach will be required, though this will not be a comprehensive study due to the lack of gauged sites.

5. CONCLUSIONS

Due to the high rainfall intensity common during storms in the wet season (over 200mm/hr), and the low infiltration rates (< ~20mm/hr) in many areas of the catchment, there is considerable potential for infiltration excess overland flow. This is reduced in most of the study area due to the presence of bunded paddy fields. In the Tarh land however, there is typically little vegetation due to grazing pressure, and a considerable fraction of the rainfall immediately runs off to lower areas of the catchment.

The rate of decline in the water level in 4 ponds supports the hypothesis based on the land class map that these are acting as sources of recharge to the shallow groundwater system. In comparison, the infiltration rate in some of the paddy fields was very low, showing that the spatial distribution of soils is an important factor in understanding the hydrology of the study site. The EM survey work shows that the EM data can be used to indicate the variations in subsurface moisture between early and late in the dry season, with the Tarh land having typically low conductivity and the Bohal land higher conductivity. Coupled with the soil map and climate data being collected, these results will be used to develop and test the three models being explored in this project.

The purpose of the monitoring described in this paper is to understand the response characteristics of the study site, and assist with development of models of the various water harvesting structures that PRADAN use. These models will be used to develop guidelines for the design of water harvesting measures, in order to improve the efficiency of such work (reducing the cost per unit of water captured) as well as increasing the understanding of the impact of water harvesting on downstream users.

ACKNOWLEDGEMENTS

This project has been funded by the Australian Centre for International Agricultural Research (project LWR/2002/100). The authors would also like to thank the staff of PRADAN, and the villagers in the Pogro and Amagara study sites for their assistance in field campaigns, and in monitoring the catchment response.

REFERENCES

Croke, B.F.W., Andrews, F., Jakeman, A.J., Cuddy, S.M., and Luddy, A., IHACRES Classic Plus: A redesign of the IHACRES rainfall-runoff model, *Environmental Modelling and Software*, *21*, 426-427, 2006.

Hutchinson M.F., On thin plate splines and kriging. In: Tarter ME and Lock MD.(eds). *Computing and Science in Statistics* 25. University of California, Berkeley: Interface Foundation of North America, 55-62, 1993.

Leavesley G.H., and Stannard L.G., The precipitation runoff modeling system – PRMS, in: Singh V.P., ed. Chapt 9. Highlands Ranch, CO: Water Resources Publications, 281-310, 1995.

McNeill, J.D., Electromagnetic Terrain Conductivity Measurement at Low Induction Numbers, Technical Note TN-6, Geonics Ltd (www.geonics.com), 1980.

Niehoff D., Fritsch U., and Bronstert A., Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany. *Journal of Hydrology*, *267*, 80-93, 2002.

Vardavas, I.M., A simple model for rapidly computing terrestrial flux, solar flux and global mean surface temperature, *Ecological Modelling*, *35*, 189-210, 1987.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Understanding riverine wetland-catchment processes using remote sensing data and modelling

A. van Griensven[a], Y. Xuan [a] D. Haguma [b], W. Niyonzima [a]

[a] *UNESCO-IHE Institute for Water Education, Department of Hydroinformatics and Knowledge Management, P.O. Box 3015, 2601 DA DELFT, The Netherlands (E-mail A.vanGriensven@UNESCO-IHE.org)*
[b] *KIST, P.O. Box 3900 Kigali, Rwanda*

Abstract:

Many wetlands are dependent on river basin processes, and at the same time, scientists have recognised the important role of wetlands in the river basin: they are important zones for groundwater recharge, they temper high flows and are therefore very important to limit flooding. By purifying the water, wetlands also play a very important role for water quality. However, the underlying processes are often not very well known or are poorly quantified. This is especially true for many wetland rich basins in Africa that are poorly gauged or even ungauged. In this study we analyse to what extend remote sensing data can be of use to analyse and model ungauged or poorly gauged wetland rich river basins, taking the Kagera river basin as a case study by using existing maps with information on the basis of soil water of the toplayer, evapotranspiration maps, DEM and land use maps and satellite born rainfall estimates. A river basin model is built using the "Soil and Water Assessment Tool". The freely available GIS maps for DEM, Soil Maps and land use maps are very useful for setting up catchment models but they are not very helpful for wetland characterisation. The satellite rainfall data are still not well replacing ground data to model the rainfall-runoff processes. The generated maps for daily soil water values are promising to derive information on the dynamics of remote riverine wetlands. With this information, a better integration of wetland and catchment models should be obtained.

Keywords: Wetlands; catchments; remote sensing; SWAT

1. INTRODUCTION

Wetlands are buffer zones between water and land that play important roles in controlling flooding, reducing pollution, providing habitats for fauna and biota and maintaining livelihoods. Wetlands constitute only around 1% of Africa's total surface area, (excluding the smaller seasonal wetlands), and relatively little scientific investigation has been undertaken in comparison to other ecosystems such as forest or to wetlands in other parts of the world (in America and some part of Asia). Africa's many wetland types support a great diversity of plants and animals, and livelihoods.

But, wetlands suffer from over-extraction of fresh water, overuse of their resources, drainage and pollution. They are blocked by dams or turned into agricultural land. Already 50% of the world's wetlands have been lost in the past century (van den Bergh et al., 2004). Also Rwanda's largest population (80%) lives in wetland reclaimed areas. Actually, it is needed to make a serious attention on the remaining ones. In order to increase

understanding of the role of wetlands for humans and nature, and in order to follow the wetland area in the world, global identification and characterisation techniques are needed.

Many wetlands are dependent on river basin processes, and in the same time, scientists have recognised the important role of wetlands in the river basin: they are important zones for groundwater recharge, they temper high flows and are therefore very important to limit flooding. By purifying the water, wetlands also play a very important for river water quality. However, the underlying processes are often not very well known or are poorly quantified. This is especially true for many wetland rich basins in Africa that are poorly or ungauged basins. Therefore we aim to analyse to what extend remote sensing data can be of use to analyse and model ungauged or poorly gauged wetland rich river basins, taking the Kagera river basin as a case study by using existing maps with information on the basis of soil water of the toplayer, DEM and land use maps and satellite born rainfall estimates.

2. THE KAGERA CATCHMENT AND WETLANDS

The Kagera River basin is distributed in four countries: Burundi, Rwanda, Tanzania and Uganda The study area covers an area of approximately 58,000 km². The Kagera River is the largest of the 23 rivers that drain into Lake Victoria and it is the most remote head stream of Nile River. It is formed at the confluence of two rivers: Nyabarongo (Rwanda) and Ruvubu (Burundi). The elevation in the basin varies from around 1,100 m above mean sea level in east to 4500 m in the Northern-west.

Rainfall varies from less than 1000mm over the eastern part of the basin up to 1,800 mm and above in the west, where most of the runoff is generated. There are two rainfall seasons, with the longer south-easterly monsoon bringing rain between about February and May, and the shorter north-easterly monsoon from about September to November. The runoff responds to the rainfall with a higher peak in May and a smaller peak in November. However, the river flows are attenuated by a number of lakes, and wetlands and associated lakes (Holmberg et al., 2003). Heavy rainfall during rainfall seasons and steep slopes cause erosion hence degradation of ecosystem. On the other hand, there are also droughts during prolonged dry seasons. Moreover, Kagera Basin faces problems related to water resources management of trans-boundary watersheds. Wetlands have a significant influence on hydrologic processes in Kagera River basin and they occupy in average 6% of Burundi and Rwanda surface area (FAO-SAFR, 1998).

3. REMOTE SENSING AND GIS DATA

The following data have been used:

Digital Elevation Model (DEM)

- HYDRO1k: the U.S. Geological Survey's (USGS) public domain geographic database HYDRO1k (http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html), which is derived from their 30 arc-second digital elevation model of the world GTOPO30. HYDRO1k has a consistent coverage of topography at a resolution of 1 kilometre.

- SRTM (Shuttle Radar Topography Mission) is a joint project between NASA and NGA (National Geospatial-Intelligence Agency) to map the world in three dimensions (http://www2.jpl.nasa.gov/srtm/index.html). The SRTM digital elevation data has a spatial resolution of 3 arc-second for Global coverage of latitude and longitude (approximately 90 meters). The SRTM "finished" data meet the absolute horizontal and vertical accuracies of 20 meters and 16 meters respectively.

Digital stream network (DSN):

- HYDRO1k: The USGS' HYDRO1k database is derived from the flow accumulation layer for areas with an upstream drainage area greater than 1000 km2.

Soil map:

- FAO: Food and Agriculture Organization of the United Nations (FAO, 1995) provides almost 5000 soil types at a spatial resolution of 10 kilometres with soil properties for two layers (0-30 cm and 30-100 cm depth). Further soil properties (e.g. particle-size distribution, bulk density, organic carbon content, available water capacity, and saturated hydraulic conductivity) were obtained from Reynolds et al. (1999) or by using pedotransfer functions implemented in the model Rosetta (http://www.ars.usda.gov/Services/docs.htm?docid=8953).

Landuse map

- GLCC: the USGS Global Land Cover Characterization (GLCC) database (http://edcsns17.cr.usgs.gov/glcc/glcc.html) has a spatial resolution of 1 kilometre and 24 landuse classes. The parameterization of these classes (e.g. leaf area index, maximum stomatal conductance, maximum root depth, optimal and minimum temperature for plant growth) is based on the available SWAT landuse classes and literature research.

Weather data

- TRMM (Tropical Rainfall Measuring Mission). TRMM is joint project of NASA and National Space Development Agency of Japan, which objectives are observing and understand the tropical rainfall and how this rainfall affects the global climate. TRRM is interested in tropical rainfall because most of the rainfall is found in tropical zones and have a significant influence on the global climate change. Daily accumulated rainfall or rain rate data are available for the period from 1998 to present, with a spatial resolution varying from 0.25 degree to 5.0 degree resolution. (http://disc2.nascom.nasa.gov/Giovanni/tovas/TRMM_V6.3B42_daily.shtml)

- NCEP/NCAR: The NCEP/NCAR Reanalysis Project is a joint project between the National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The NCEP/NCAR Reanalysis data are divided into 6 different sections: pressure level, surface, surface fluxes, tropause sections, derived data and spectral coefficients. Precipitation data are in surface fluxes section. This section comprises 4-times daily, daily and monthly time series from January 1st, 1948 to present. The data sets spatial coverage is T62 Gaussian grid with 192x94 points. Other climate variables available through this dataset are air temperature, surface lifted index, pressure, relative humidity, and wind velocity. The levels of measurement are surface or near the surface, or entire atmosphere. (http://www.cdc.noaa.gov)

Wetlands data:

- WETLANDS-FAO: information on wetlands is gathered at the site http://www.fao.org/DOCREP/003/X6611E/x6611e00.htm#TopOfPage

Soil Water data

Figure 1 shows the daily Soil Water content map with grids of a resolution of 0.25 degree that containes soil water content in percentage as well as identifiers for grids with water surface > 5% of the area (Owe et al., 2008). As can been seen in the area that zooms in onto the Kagera basin, some pixels referring to water surface or the areas that contain wetlands. Each day contains strips of missing data as can be seen in Figure 1 but those strips shift. Around 30-40% of the values of the constructed time series for the Kagera basin are missing data.

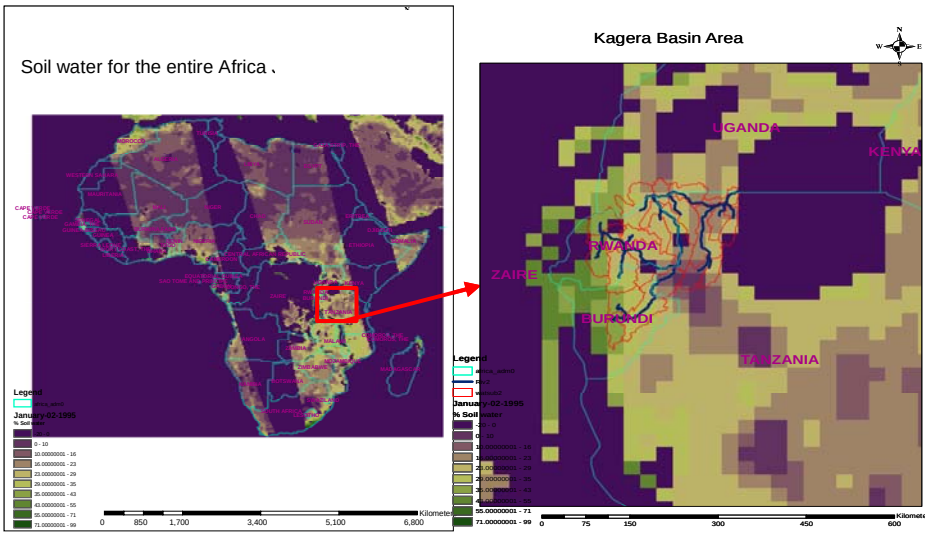


Figure 1 Soil water content map for entire Africa and for Kagera

4. CATCHMENT MODEL

Wetlands play a very important role in the catchment hydrology. Since catchments contain wetlands and not vice versa, the catchment model SWAT (Soil and Water Assessment Tool) (Arnold et al., 1998) was chosen for this study.

SWAT

SWAT is a conceptual model that operates on a daily time step. The objectives in model development were to predict the impact of management on water, sediment and agricultural chemical yields in large basins. To satisfy these objectives, the model (a) uses readily available inputs for large areas; (b) is computationally efficient to operate on large basins in a reasonable time, and (c) is capable of simulating long periods for computing the effects of management changes.

Model set-up

The watershed delineation of Kagera River basin was done using a DEM of spatial resolution of around 90m x 90m from SRTM, and the basin was subdivided into 12 sub-catchments, whose surface areas range from 11,900 to 16,745 km². The other data used are soil map of 1km x 1km resolution produced by FAO and land use map from Global 2000 land cover with a resolution of 1km x 1km. Monthly discharge from "The Global River Discharge Database" for 5 gauging stations was used for model evaluation. Several sources of weather data were compared.

The model calibration period was from 1974 to 1979. The following period up to 1985 was used to validate the calibrated model. The calibration process was done in three phases: resizing of subbasins channels, specification of wetlands, modification of ground water flow parameters and then auto-calibration with sensitive parameters. An automated base flow separation filter (Arnold, 1995) was used to determine the recession constant and ground water delay for the subbasins from stream flow data.

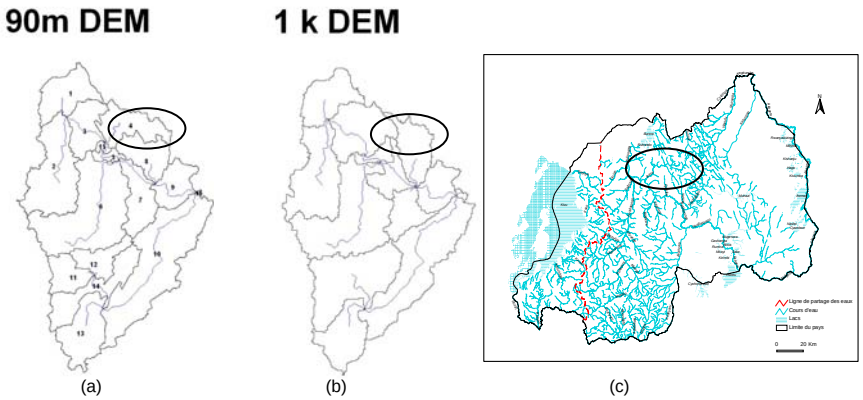


Figure 2 Delineation of Nyabarongo and Rubuvu catchement using 90m DEM (a) and 1 kilomter DEM (b) and the hydrological map of Rwanda (c)

DEM evaluation

A SWAT model was built for the Nyabarongo and Ruvubu sub-catchments that are located in the South–West of the Kagera basin, within the countries of Rwanda and Burundi. In Figure 3, it is clear that the difference of DEM resolution is mainly apparent at smaller scales. The shape of the watershed is very much alike, but several defined subbasins, like sub-basin 4 get a very different shape. Due to this, the runoff water generated in the circeled area of Figure 2 will enter the Nyabarongo river at different locations. From the hydrological map of Rwanda it can be concluded that the 90 m DEM produces the right delineation. Also very clear differences are apparent in the calculated slope that are 4 times higher in the mountaineous subbasin 2 when using the 90 m DEM compared to the 1 k DEM.

Rainfall input evaluation

Different sources of rainfall derived from remote sensing data are compared to groundwater data (Haguma, 2007). The evaluation is based on the coefficient of determination (R^2), Nash-Sutcliffe efficiency coefficient (Ns) and Deviation of stream flow volume (Dv) to measure the model performance.

The simulated stream flow using TRMM gave an underestimation, where the best simulation gave only 18% of the total volume. Due to underestimation of precipitation, the simulated stream flow was very low so that no calibration was done, except manual adjustment of ground water parameters. A calibration and validation was performed for the CRU/dGen and the NCEP/NCAR giving reasonable results, but poor results for predictions of the validation period (Table 1).

Table 1. Performance of the Kagera model for different sources of rainfall data (Ns=Nash-Sutcliffe efficiency, R^2= coefficient of determination, RMSE=Root mean squared error)

	CRU/dGen		NCEP/NCAR		Ground data	
	Calibration (1974-1979)	Validation (1980-1985)	Calibration (1974-1979)	Validation (1980-1984)	Calibration (1974-1979)	Validation (1980-1985)
Ns	0.41	0.01	0.43	-21.03	0.63	-1.36
R2	0.48	0.26	0.45	0.03	0.65	0.08
Dv	-7.67%	-10.83%	-6.96%	39.49%	-7.31%	11.24%

Wetland incorporation

Wetlands were included on the basis of the previously described FAO site. The poor validation results can be related to the lack of information on the wetland hydrology in the

model. Promising results were obtained for the Nyarabongo watershed at Kigali station that does not contain large wetlands (Figure 3). The Nash-Sutcliffe efficiencies of the calibration and validation were 0.65 and 0.78 respectively.

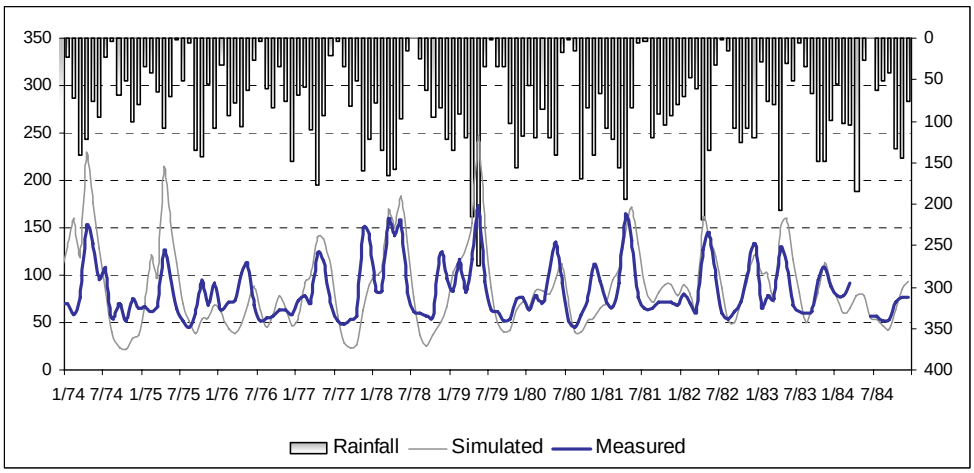


Figure 3 Simulations for the Nyarabongo watershed, validation period using (Kigali station)

4. WETLAND IDENDTIFICATION AND CHARACTERISATION

For the cells 6 (wetland area) and 18 (agricultural area) of Figure 4, time series have been computed (Figure 5) based on the daily soil water data maps for the year 2005.

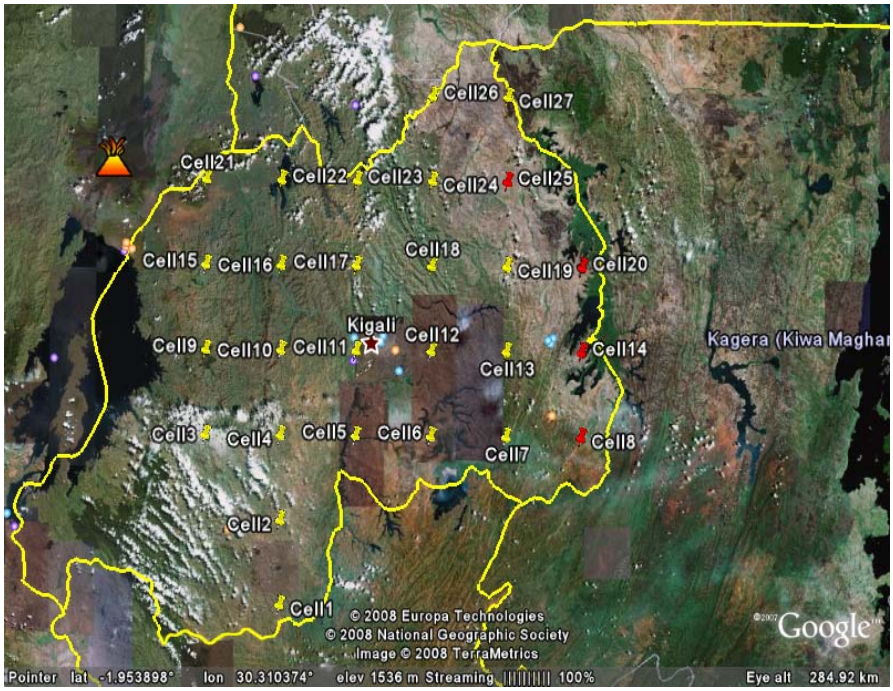


Figure 4 location of the grid centers of the soil water maps for Rwanda

The time series generated for the wetland and agricultural cells are presented in Figure 5. Cells that contain more of 5% water surface got separate indicators and are represented by the value 110% (over-saturation). In the year 2005 60% of the days contained values of which around 60% where identified as water surface. The remaining 82 days with soil water values for both cells show a relative similar pattern (r^2=0.45) and average value with

less then 3%. This gives the indication that the wetland is flooded during around 60% of the time and that the behaviour of the soil water during the periods where the wetland is not flooded, is quite similar to the analysed agricultural land area grid.

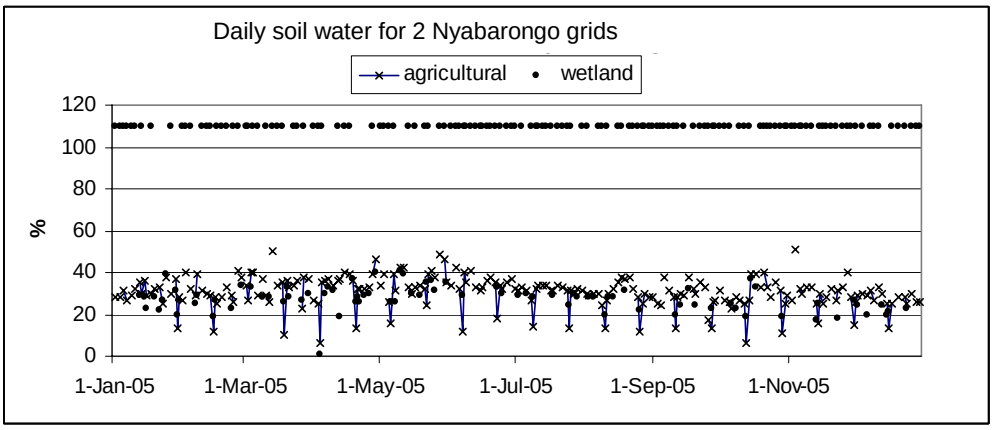


Figure 5 Soil water content for a agricultural grid and a wetland grid (110% means water surface)

5. CONCLUSIONS AND RECOMMENDATIONS

Riverine wetlands and river basins have very strong influences on each other hydrology and functioning. It is therefore important to do a joint analysis of both. However, wetlands are often located in remote areas or poorly gauged what is certainly true for many wetlands in Africa. But, there is a lack of ground data to do a proper understanding of the complex relations between the wetland and the catchment. Therefore a couple of remote sensing born data have been collected and processed for the Kagera catchment in the form of DEM, soil maps, land use maps, daily soil water maps and (sub)-daily rainfall data. The DEM, soil maps and land use maps seemed are extremely useful for catchment modelling tools that are nowadays by default imbedded in GIS based preprocessors, such as the Soil and Water Assessment tool. The delineated (sub)basins and calculated slopes are however very much dependent on the resolution of the DEM. Even thought the Kagera basin is a poorly gauged basin for precipition, it seemed that rainfall estimates derived from satellite images are not very useful up to date for this case.

The automated preprocessing of the GIS data towards a SWAT model does not consider any wetlands or wetland processes. In the Kagera, as well as many other African river basins, wetlands are important landscape features that strongly influence the hydrology. For the identification of wetlands and their flooding periods, an analysis was done of daily soil water data maps. The results show indeed that grids with wetlands are indeed temporally identified as water surface.

Further step to be done is to get riverine wetlands descriptions and processes included into catchment models. Such an integration should consider the contribution of surface water and groundwater to the wetlands, while the wetlands themselves are impacting the groundwater recharge, the river runoff processes and the river hydraulics.

ACKNOWLEDGEMENT

The authors wish to thank the Nile Basin Initiative for supporting the research of William Niyonzima and the EU funded WETwin project. Special thanks for the soil water data we received from the Univesiteit Amsterdam.

REFERENCES

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and J. R. Williams, Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association*, 34(1), 73 89, 1998.

Hails, A. J., Wetlands, Biodiversity and the Ramsar Convention: The role of the Convention on the Wetlands in the Conservation and Wise Use of Biodiversity, Ramsar Convention Bureau, Gland, Switzerland, 1996.

Owe, M., De Jeu, R.A.M., and T.R.H., Holmes, Multi-Sensor Historical Climatology of Satellite-Derived Global Land Surface Moisture, *J. Geophys. Res*. 113,F01002, doi:10.1029/2007JF000769., 2008.

FAO (Food and Agriculture Organization), The digital soil map of the world and derived soil properties. CD-ROM, Version 3.5, Rome, 1995.

FAO-SAFR, Wetland characterization and classification for sustainable agricultural development, Harare, Zimbabwe, 1998.

Haguma, D., Development of hydrologic model of Kagera River basin using Remote sensing data, thesis, Masters of Water and Engineering, UNESCO-IHE, Delft Netherlands,2007.

Holmberg, J. and T., Lilja, Kagera River basin integrated water resources management project, WSP International Sweden AB, 2003.

Kalnay E., Kanamitsu, M., Chelliah, M. Ebisuzaki, W. and R., Jenne. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3):437-471, 1998.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M., and M., New, A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100).Tyndall Centre Working Paper 55, 2004.

NCDC (Nation Climatic Data Centre), Global daily summary: Temperature and precipitation 1977-1991. CD-ROM, Version 1.0. , 1994.

NCDC (Nation Climatic Data Centre). Global daily climatology network (GDCN). CD-ROM, Version 1.0., 2002.

New, M., Hulme, M., and P., Jones, Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *J. Clim.* 13(13), 2217-2238, 2000.

Schuol, J. and K.C., Abbaspour, Using monthly weather statistics to generate daily data in a SWAT model application to West Africa. *Ecol. Model.* 201, 301-311, 2007.

Reynolds, C.A., Jackson, T.J. and W.J. Rawls. Estimating available water content by linking the FAO soil map of the world with global soil profile database and pedo-transfer functions. Proceedings of the AGU 1999 spring conference. Boston, MA, 1999.

van den Bergh, J. C. J. M., Barendregt A. and A.J., Gilbert, Spatial Ecological-Economic Analysis for Wetland Management: Modelling and Scenario Evaluation of Land Use, Cambridge University, 2004.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Effects of Rainfall and Soil/Land Use Spatial Distribution on Hydrological Response at Different Scales

N. Berni[a], A. Viterbo[a], C. Pandolfo[a], M. Stelluti[a], S. Barbetta[b] and L. Brocca[b]

[a] *Functional Centre – Umbria Region - Via Romana Vecchia, 06034 Foligno (Pg), Italy (centrofunzionale@regione.umbria.it)*

[b] *National Research Council, Research Institute for Geo-Hydrological Protection, Via Madonna Alta 126, 06128 Perugia, Italy*

Abstract: The influence of spatial variability of rainfall and soil/land use on the accuracy of runoff simulation was investigated in order to define the optimal spatial discretization of the basin to be adopted in rainfall-runoff modeling. In particular, the effects of a uniform versus distributed spatial representation of both rainfall field and basin properties was analyzed by using a semi-distributed model. These effects were evaluated by comparing the observed and simulated flood hydrographs for different sub-basins of the Upper Tiber River basin at Ponte Felcino river section (drainage area of ~1800 km^2), located in Central Italy. On the basis of the obtained results, a dimension of about 300 km^2 for the homogeneous elements of the basin can be considered adequate for model applications addressed to flood forecasting and warning activities.

Keywords: Spatial variability, scale, rainfall-runoff modelling, semi-distributed scheme; flood forecasting.

1. INTRODUCTION

Several authors have emphasized the role of rainfall spatial distribution on estimating peak discharge and discharge hydrograph shape [Julien & Moglen, 1990; Beven & Hornberg, 1982; Syed et al., 2003]. In particular, Michaud & Sorooshian [1994] showed that significant errors in peak discharge can be found by using rainfall data of an inadequate hydrometeorological network. Ogden et al. [1995] and Singh [1998] investigated the effects of storm direction and velocity on peak discharge and discharge hydrograph shape. Syed et al. [2003] showed the influence of the interaction of basin area and shape with rainfall spatial structure in the runoff generation process.

The spatial distribution of soil/land use characteristics can also affect the hydrograph shape [Merz and Plate, 1997; Merz and Bardossy, 1998; Singh and Woolhiser 2002]. This issue has been investigated by using the Soil Conservation Service method for abstraction for which the soil/land use characteristics are parameterized through a dimensionless Curve Number (CN). Grove et al. [1998] underlined that runoff depth estimates using a distributed CN are as much as 100% higher than that obtained considering a uniform CN. Underestimation of runoff depth due to CN compositing is most severe for wide CN ranges, low CN values and low precipitation depths. For two watersheds in South Africa, Hope and Schulze [1982] concluded that a distributed CN approach furnished estimates of stormflow more accurate than those derived by adopting a lumped CN method.

The selection of the optimal spatial discretization for the above mentioned quantities represents a key factor in the development of reliable rainfall-runoff models [Hellenbrand and van den Bos, 2007]. This issue becomes fundamental when the models are involved in

real-time flood forecasting systems, such those developed and managed by the Italian network of regional "*Functional Centres"*, coordinated by the National Civil Protection Department in Rome. In fact, the availability of accurate forecasts allow to minimize false/missed alarms [Ravazzani et al., 2007].

Based on the above issues, the main purpose of this study is to analyze the influence of rainfall and soil/land use spatial representation on the estimation of the hydrologic response for sub-basins of the Upper Tiber River (Central Italy), that is concern of Umbria region Functional Centre. Specifically, the effects of a uniform versus distributed spatial account of these quantities are analyzed through the comparison of simulated flood hydrographs for different river sections. At the purpose a freely available event-based hydrological model of semi-distributed type (Hydrologic Engineering Center - Hydrologic Modeling System, HEC-HMS) is applied to simulate four significant flood events occurred in the last five years. Moreover, the spatial discretization of the basin is addressed in view of model applications for flood warning purpose.

2. SEMI-DISTRIBUTED RAINFALL-RUNOFF MODEL

The freely available semi-distributed event-based hydrological model HEC-HMS, coupled with HEC-GeoHMS module, was used for the analysis. HEC-GeoHMS is a public-domain extension of ESRI ArcMap 9.1 GIS program which is useful to perform terrain pre-processing, delineate sub-basins and streams, compute hydrologic parameters and directly produce the geographical input data for the HEC-HMS model.

HEC-HMS is the updated version of the United State Army Corps of Engineering (USACE) rainfall-runoff model HEC-1 and it was chosen for this study because of its flexibility. In fact, many of the most common formulations used in hydrologic engineering for runoff volume, base flow, direct runoff and channel flow assessment, are included in the software [USACE, 2000]. The geometric representation of the basin is based on the main elements implemented in HEC-HMS: subbasin, reach, junction and source. In the following, a brief description of the model components formulation adopted in this study is given.

Specifically, losses are estimated by using the Soil Conservation Service - Curve Number (SCS-CN) for which the direct runoff depth, P_e, is:

$$P_e = \frac{(P - I_a)^2}{(P - I_a + S)} \qquad P \geq I_a \tag{1}$$

where P is the rainfall depth, S is the potential maximum retention and I_a is the initial abstraction which can be expressed as a function of S. The SCS expressed $I_a = 0.2S$ on the basis of the results obtained for several experimental watersheds [Ponce and Hawkins, 1996]. The potential maximum retention, S, is related to a dimensionless Curve Number (CN) defined as a function of land use, soil type and antecedent wetness conditions (AWC).

The rainfall-runoff transformation is represented by using the SCS - Unit Hydrograph (SCS-UH) method incorporating only one parameter. This parameter is assumed to be the basin lag time, defined as the time shift between the centroids of effective rainfall and direct runoff [Singh, 1975; 1988]. Specifically, for each homogeneous elements (sub-basin), the lag time, L, is estimated by using the lag-area relationship proposed by Melone et al. [2002]:

$$L = \eta 1.19 A^{0.33} \tag{2}$$

with L basin lag time (in hours), A drainage area (in km^2) and η a parameter to be calibrated. Equation (2) with $\eta=1$ was obtained considering 26 watersheds in Central Italy ranging in area from 12 km^2 to 4147 km^2. However, this result refers to the effective rainfall hyetographs determined by the extended form of the two-term Philip infiltration equation and the use of the geomorphological unit hydrograph for the rainfall-runoff transformation. Therefore, η is considered here as a calibration parameter to take into account of the differences due to the use of the SCS-CN method and the SCS unit hydrograph.

Finally, flood routing along the natural channels is simulated through the lag routing method which represents the simple translation of flood waves neglecting the attenuation or diffusion processes. The method requires the estimation of only one parameter, the channel lag time, L_C, that can be determined as a function of reach length and slope and mean flow velocity, v_m.

3. STUDY AREA AND DATA SETS

The model was applied to the Upper basin of the Tiber River located between the hydrometric sections of Gorgabuia and Ponte Felcino with an area of ~1800 km^2 (see Figure 1). The catchment has a complex topography that can significantly enhance the widespread frontal rainfalls causing the major flood events. A geolithological map (scale 1:100000) and the soil/land use map defined in the CORINE – LAND COVER European project are available. Accordingly, the study area is characterized by terrigeneous facies and flysch deposits mainly consisting of clayeyschistose and clayey-marly sediments; 57% of the total watershed area is covered by wood, the remaining area is 37% agricultural crop and 5% pasture. The forests are generally located in headwater areas and cropping in valley floors. The area is affected by Mediterranean climate with average annual precipitation of about 900 mm and mean annual temperature of 11 °C; the mean annual potential evapotranspiration, computed with the Thornthwaite formula [1948], is almost 800 mm. The 50 years return period peak discharge is equal to 490 m^{3}s^{-1} and 1020 m^{3}s^{-}1 at Santa Lucia and Ponte Felcino section, respectively.

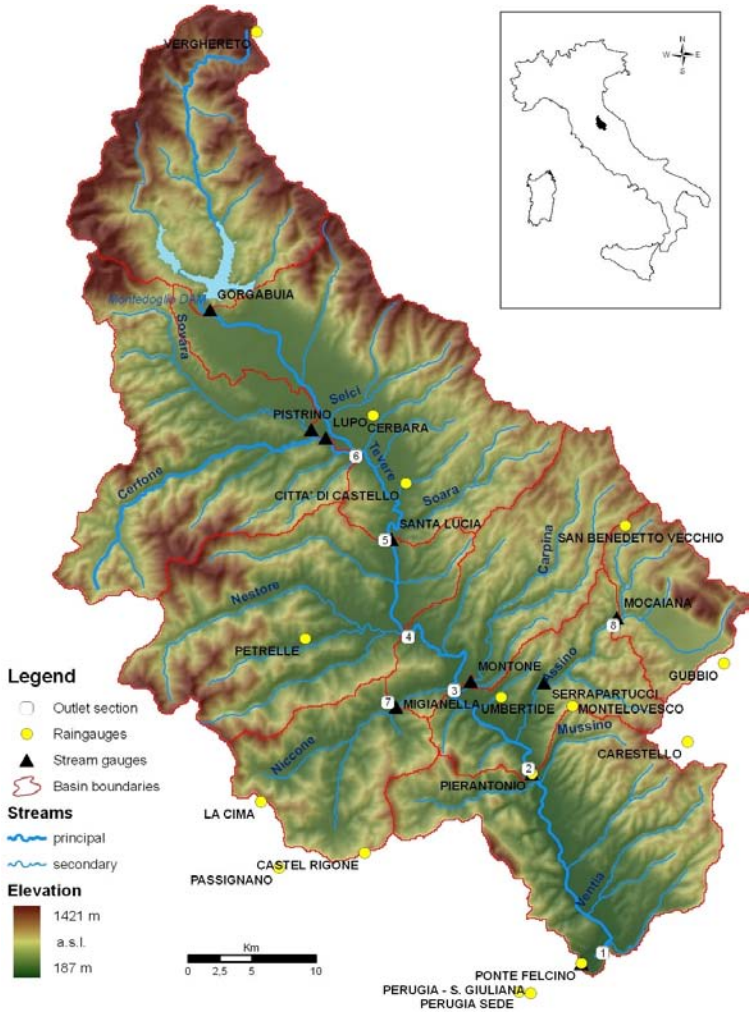


Figure 1. Study area: topographic characteristics and operating hydrometeorological network. The eight selected sections are also shown as white circles.

A dense real time hydrometeorological network (1 station every 150 km^2) has been operating in the Upper Tiber River basin for more than 20 years and the data are recorded with a time interval of 30 minutes. The hydrometric sections are located along the main channel and the secondary streams (see Figure 1), the corresponding stage-discharge relationship is frequently updated so allowing reliable discharge estimates. For this study the data recorded during the period December 1998 - December 2003 by 13 rain gauges and 10 hydrometric gauges were considered. In this period the four more significant flood events were selected, their main characteristics are summarized in Table 1.

Table 1. Main characteristics of selected flood events for the Upper Tiber basin at Ponte Felcino.

Event Date	Rainfall Depth (mm)	Initial Base Flow ($m^3 s^{-1}$)	Observed Direct Runoff Peak ($m^3 s^{-1}$)
December 14, 1996	31.4	26	315
June 01, 1997	98.2	10	399
December 04, 1998	33.9	43	254
December 27, 2000	53.8	148	365

4. METHODOLOGY

In order to define the optimal rainfall and soil/land use spatial distribution, the hydrological model was implemented considering two different configurations for both rainfall and CN parameter. Table 2 summarizes the four different configurations considered for the analysis. In particular, for the distributed Cases (Case 1-3) the rainfall and the CN values were assumed variable from one sub-catchment to another according to the rainfall spatial pattern and soil/land use characteristics. For Case 4 the model can be considered as a lumped model. It has to be noted that the comparison of Case 1 and Case 2 (distributed versus uniform rainfall) and between Case 1 and Case 3 (distributed versus uniform CN) is used to highlight, separately, the effects related to the assumption on the rainfall and the CN spatial distribution, respectively. From the other hand, the comparison between Case 1 and Case 4 allow us to define the optimal spatial discretization.

Table 2. Case study combinations for different rainfall and CN spatial distribution.

		Rainfall	
		Distributed	*Uniform*
Soil/Land Use	*Distributed*	Case 1	Case 2
	Uniform	Case 3	Case 4

To implement the model the study catchment has been subdivided into 70 homogeneous elements (sub-basins) characterized by a drainage area ranging from few square kilometers to 120 km^2 and a mean basin slope ranging from 4% to 28%. It has to be underlined that there is an artificial reservoir in the upper part of the selected catchment formed by the Montedoglio dam (Figure 1). Therefore, the upstream drainage area has been simulated by using a source element in the HEC-HMS model represented by the discharge observed at Gorgabuia hydrometric station.

For the analysis, eight river sections, whose location is shown in Figure 1, were selected. Table 3 summarizes the main properties of each sub-catchments subtended by the eight cross sections including the drainage area, the main basin slope and length and the CN values for intermediate antecedent wetness conditions, CN_{II}.

The four selected flood events were simulated by the HEC-HMS model considering all the combinations of Table 2 and the flood hydrographs estimated at several river sections within the basin were compared. For each flood event the antecedent wetness conditions were assessed scaling by a correction factor, C_f, the CN_{II} values in order to reproduce the observed direct runoff volume at the outlet river section. In this way, the CN spatial distribution was kept on and only the absolute CN values were modified without loosing the spatial variability described by the soil/land use characteristics [Frances et al., 2007].

Table 3. Main characteristics of the sub-catchments selected for the analysis. CN_{II} represents the Curve Number value for intermediate antecedent wetness conditions.

ID	Outlet Section	Drainage Area (km^2)	Mean slope (%)	River length (km)	CN_{II}
1	Tiber R. at P.Felcino section	1791	1.00	107	70.6
2	Tiber R. at Pierantonio section	1511	1.19	93	71.4
3	Tiber R. at Niccone R. confluence	1280	1.34	83	71.6
4	Tiber R. at Néstore R. confluence	948	1.50	71	71.7
5	Tiber R. at S. Lucia section	652	1.83	63	72.0
6	Cerfone Str. at Tiber R. confluence	301	3.13	34	72.0
7	Niccone Str. at Migianella section	134	2.35	16	70.5
8	Assino Str. at Mocaiana section	99	0.83	15	71.0

5. RESULTS AND DISCUSSIONS

The hydrological model was calibrated for the Case 1, i.e. distributed rainfall and CN values. The discharge data of the two first events observed at the outlet station Ponte Felcino were used to estimate model parameters. In particular, the model parameters related to the rainfall-runoff transformation, η, and channel routing, v_m, were optimized considering the agreement of the overall shape of the discharge hydrograph. Then, the optimal correction factor, C_f, for the CN values was determined for each event in order to reproduce the direct runoff volume at Ponte Felcino section; the C_f values are equal to 1.23 for December 1996, 0.55 for June 1997, 1.11 for December 1998 and 1.12 for December 2000. Figure 2 shows the comparison between observed and simulated flood hydrographs for the Ponte Felcino section and for the upstream section of Santa Lucia.

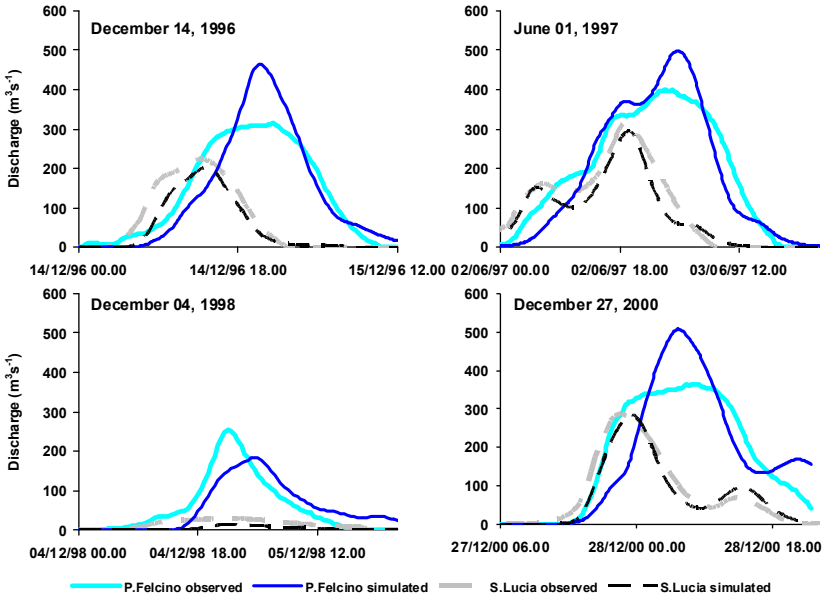


Figure 2. Comparison of the observed and simulated flood hydrographs for the Santa Lucia and Ponte Felcino river sections.

As the model parameters are estimated through the flood hydrograph observed at the outlet of the catchment, the results obtained for Santa Lucia section have to be considered for model validation. As it can be seen in Figure 2, the model predicts a narrow flood hydrograph shape for Ponte Felcino section both for calibration and validation events. This can be due to the fact that the model does not take into account of flooding, that occurred for the flood events of December 1996 and 2000, and of the diffusion process in the channel routing module. However, the model was able to reproduce with a fair accuracy the hydrograph shape and timing both for Ponte Felcino and Santa Lucia river section. Similar results were obtained for other equipped river sections, not shown here for sake of brevity.

In order to investigate the effects of rainfall and CN spatial distribution on hydrologic response, the hydrological model was then applied considering the different configurations of

rainfall and soil/land use characteristics summarized in Table 2. For the flood occurred on June 1997 Figure 3 shows the discharge hydrographs estimated at the eight selected river sections for the four different configurations and, if available, the observed discharge hydrograph.

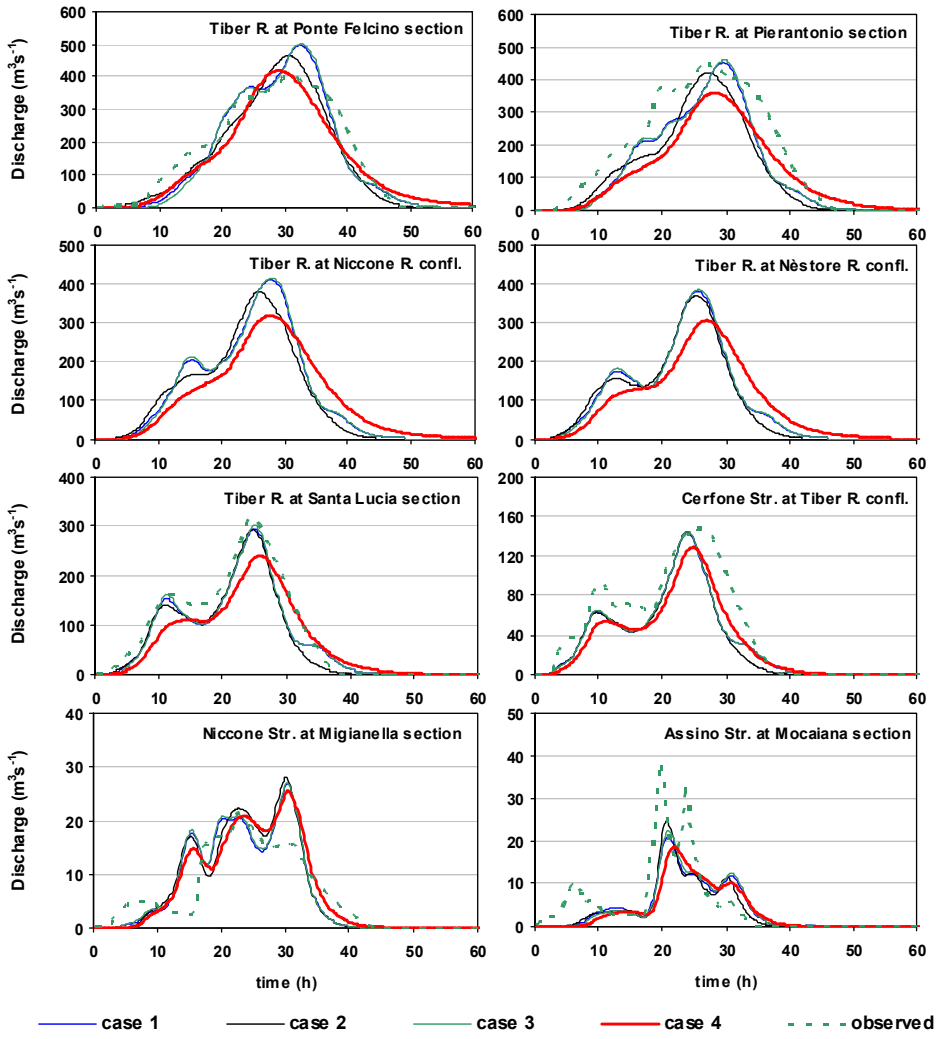


Figure 3. Event of June 1997: comparison between simulated discharge hydrographs for different rainfall and soil/land use spatial pattern (Case 1-4 of Table 2). The observed discharge, if it is available, is also shown.

As it can be seen, Case 1 and Case 3 furnish very similar results for all sections denoting that, for the study area, the CN representation does not significantly affect the hydrologic response. This result is in accordance with Grove et al. [1998]. In fact, the CN values vary in a limited range (Table 3) and the flood events are characterized by high rainfall depths. Moreover, Cases 1 and 3 reproduce the observed discharge hydrograph more accurately than Cases 2, particularly for the Tiber River sections. From these observations it can be derived that, for our study area, the knowledge of the rainfall spatial pattern is more important than the soil/land use spatial distribution. Moreover, for all the selected river sections the discharge hydrographs simulated by the lumped approach (Case 4) were significantly different from those of the other three Cases with a general underestimation of peak flow.

To define the optimal rainfall and soil/land use spatial distribution it was used the Root Mean Square Error (RMSE) and the Nash-Sutcliffe coefficient (NS) [Nash & Sutcliffe, 1970] computed between simulated and 'observed' discharge hydrographs. In particular, the 'observed' hydrograph refers to discharge computed by the hydrological model under the hypothesis of heterogeneity on rainfall and CN values (Case 1), whereas the simulated hydrographs correspond to Case 2-4. For the four selected events, Figure 4 shows the RMSE

as a function of contributing drainage area. As it can be seen, Case 3 provided simulated hydrographs very similar to those 'observed' for every basin scale. On the contrary, Case 2 and mainly Case 4 furnished quite different discharges with error values increasing significantly with the contributing drainage area. Analyzing the slope of the RMSE curve for Case 4, it can be inferred that, for most of the investigated flood events, the RMSE starts to increase quickly when the basin drainage area becomes greater than ~300 km^2. Analogous considerations can be inferred for the NS values which is greater than 82% for basin drainage area smaller than ~300 km^2 and rapidly decrease for larger areas. This threshold value can be assumed as the minimum areal extension of the homogeneous elements for partitioning the basin in order to maintain the spatial variability of rainfall and soil/land use which can affect significantly the basin response. This result was corroborated by the RMSE computed between the simulated hydrographs and those observed at the eight gauged sections, subtending a drainage area from 100 to 1800 km^2 (not shown here for sake of brevity). In fact, the RMSE of Case 4 starts to move away from that computed for Case 1-3 for a drainage area of about 300 km^2.

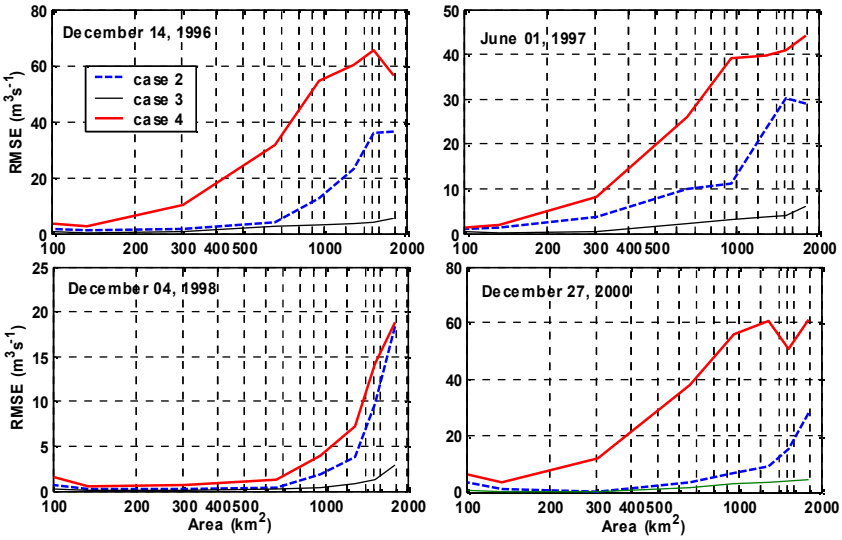


Figure 4 Root Mean Square Error (RMSE) computed between the discharge hydrograph of Case 2-4 and that of Case 1 for the four selected events as a function of contributing drainage area.

6. CONCLUSIONS

In this study a semi-distributed event-based hydrological model was applied to four significant flood events occurred in the Upper Tiber River basin. Model performance was found fairly accurate in reproducing both flood hydrograph shape and timing at the basin outlet and for different equipped sub-basins.

The comparison between the discharge hydrographs provided by different configuration of rainfall and soil/land use spatial pattern was carried out in order to analyze the effect of heterogeneity of these quantities on the hydrological response at different scales. In particular, a uniform versus distributed spatial representation of these quantities was analyzed. The obtained results highlighted the importance of the rainfall spatial distribution knowledge to determine the runoff response in our study area and, the low influence of the soil/land use spatial characterization due to its low variability in the study area. Moreover, a threshold value for the sub-catchment area of 300 km^2 was identified as the optimal spatial discretization to be adopted for rainfall-runoff models designed for flood warning purposes. However, further investigations on other basins with different areal and physiographic characteristics should be carried out in order to verify the reliability of the results obtained in this study.

REFERENCES

Beven, K.J., and G.M. Hornberger, Assessing the effect of spatial pattern of precipitation in modeling streamflow hydrographs, *Water Resources Bulletin*, 18, 823-829, 1982.

Frances, F., Velez, J.I., and Velez, J.J. Split-parameter structure for the automatic calibration of distributed hydrological models, *Journal of Hydrology*, 332 (1-2), 226-240, 2007.

Grove, M., J. Harbor and B. Engle, Composite vs. distributed curve numbers: effects on estimates of storm runoff depths, *Journal of the American Water Resources Association*, 34(5), 1015-1023, 1998.

Hellenbrand, H., and R. van den Bos, Investigating the use of spatial discretization of hydrological processes in conceptual rainfall runoff modelling: a Case study for the meso-scale, *Hydrological Processes*, published online Dec 31 2007, 2007.

Hope, A.S., and R.E. Schulze, Improved estimates of stormflow volumes using the SCS curve number method. In: V.P. Singh (ed.), Rainfall-runoff relationships, *Water Resources Publications*, Littleton, 419-431, 1982.

Julien, P.Y., and G.E. Moglen, Similarity and length scale for spatially varied overland flow, *Water Resources Research*, 26, 1819-1832, 1990.

Melone, F., C. Corradini and V.P. Singh, Lag prediction in ungaged basins: an investigation trough actual data of the upper Tiber River valley, *Hydrological Processes*, 16, 1085-1094, 2002.

Merz, R., and A. Bardossy, Effects of spatial variability on the rainfall runoff process in a small loess catchment, *Journal of Hydrology*, 212-213, 304-317, 1998.

Merz, R., and E.J. Plate, An analysis of the effects of spatial variability of soil and soil moisture on runoff, *Water Resources Research*, 33(12), 2909-2922, doi: 10.1029/97WR02204, 1997.

Michaud, J.D., and S. Sorooshian, Effect of rainfall sampling errors on simulation of desert flash floods, *Water Resources Research*, 30, 2765-2775, 1994.

Nash, J.E., and J.V. Sutcliffe, River flow forecasting through conceptual models, Part I: A discussion of principles, *Journal of Hydrology*, 10(3), 282-290, 1970.

Ogden, F.L., J.R. Richardson and P.Y., Julien, Similarity in catchment response: 2. moving rainstorm, *Water Resources Research*, 31(6), 1543-1547, 1995.

Ponce, V.M. and R.H. Hawkins, Runoff curve number: Has it reached maturity? *Journal of Hydrologic Engineering, 1*, 11-19, 1996.

Ravazzani, G, M. Mancini, I. Giudici and P. Amadio, Effects of soil moisture parameterization on a real-time flood forecasting system based on rainfall thresholds, *Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007*, 2007.

Singh, V.P., Derivation of surface water lag time for converging overland flow, *Water Resources. Bulletin*, 11, 505-513, 1975.

Singh, V.P., *Hydrologic Systems: Rainfall-Runoff Modeling*. Prentice Hall: Englewood Cliffs, New Jersey; 480, 1988.

Singh, V.P., Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrological Processes*, 11, 1649-1669, 1997.

Singh, V.P., Effect of the direction of storm movement on planar flow, *Hydrological Processes*, 12, 147-170, 1998.

Singh, V.P., and D.A. Woolhiser, Mathematical modeling of watershed hydrology, *Journal of Hydrologic Engineering*, 7(4), 270-292, 2002.

Syed, K.H., D.C. Goodrich, D.E. Myers and S. Sorooshian, Spatial characteristics of thunderstorm rainfall fields and their relation to runoff, *Journal of Hydrology* , 271, 1-21, 2003.

Thornthwaite, C.W., An approach toward a rational classification of climate, *Geographic Review*, 38, 55-94, 1948.

U.S. Army Corps of Engineers (USACE) *Hydrologic Engineering Center, Hydrologic Modeling System HEC-HMS Technical Reference Manual*, 2000.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

The Significance of Spatial Variability of Rainfall on Runoff

I.G. Pechlivanidis, N. McIntyre and H.S. Wheater

Department of Civil and Environmental Engineering, Imperial College London, SW7 2AZ, London, UK (i.pechlivanidis05/n.mcintyre/h.wheater@imperial.ac.uk)

Abstract: A key issue in rainfall-runoff modelling is to assess the importance of the spatial representation of rainfall on streamflow generation. Distributed models have the potential to represent the effects of spatially variable inputs such as rainfall making them an appropriate tool to investigate the role of spatial rainfall on runoff. This paper explores the importance of spatial rainfall representation for rainfall-runoff modelling as a function of catchment scale and type. The study investigated the effect of catchment scale and type using 9 gauged catchments ranging in size from 30 to 1040 km^2. Regional relationships between known catchment characteristics and model parameters have been developed to overcome the task of estimating model parameter values at ungauged subcatchments. The results indicate the importance of considering the effect of spatial rainfall in most of the catchments with the significance of spatial effects increasing at small spatial scales. Finally, the importance of spatial variability is enhanced when impermeable areas are investigated.

Keywords: Spatial rainfall; Calibration; Semi-distributed modelling; Rainfall-runoff models.

1. INTRODUCTION

One question that arises in rainfall-runoff studies is: "How important is the spatial nature of rainfall to runoff response?" In a recent review, Singh [1997] concluded that the significance of spatial rainfall varies as a function of catchment rainfall properties, catchment scale and type and antecedent conditions. Much effort has further been devoted to gain a better understanding of the role of spatial rainfall on runoff [Bell and Moore, 2000, Smith et al., 2004, Segond et al., 2007] concluding that the spatial rainfall can influence considerably the volume of runoff, peak runoff and time to peak. However, although the literature on the relationship between spatial rainfall and runoff generation is extensive, results have been contrasting and sometimes contradictory.

Hydrological models can be classified based on the spatial variation of inputs such as rainfall and model parameters into: (a) lumped; and (b) distributed [Beven, 2001]. Spatially distributed models, compared to lumped, have the potential to represent the effects of spatially variable inputs making them an appropriate tool to investigate the role of spatial rainfall on streamflow estimation [Arnaud et al., 2002, Tetzlaff and Uhlenbrook, 2005]. However, calibration of distributed models is rarely a straightforward task, due to the large number of parameter values that may need to be estimated [Carpenter and Georgakakos, 2006]. A variety of distributed model calibration strategies have been suggested to take into account the spatial heterogeneity of model parameters; most of these involve applying a priori parameter estimates based on regionalisation approaches [Wagener and Wheater, 2006].

This paper presents the initial results of a study on the significance of spatial rainfall on runoff, considering 8 years of hourly historic data from 9 flow gauges within the Upper Lee catchment, UK. From the 8 years, five storm events were analysed in detail. For each of these events, the simulated hydrograph response using three representations of the

spatial variability of the rainfall data (i.e. averaging over four different spatial scales) was investigated. The paper also presents a simple approach to calibrating a distributed model.

2. REVIEW OF THE INFLUENCE OF SPATIAL RAINFALL ON STREAMFLOW

The picture obtained from the literature is that the spatial nature of rainfall is a significant factor controlling runoff response. However, this effect is likely to be less clear-cut when humid temperate catchments are investigated, due to the trade-off between the impact of spatial variability of rainfall and dampening and filtering effect of the catchment [Smith et al., 2004, Segond et al., 2007].

The suitable spatial resolution of rainfall measurements for hydrological purposes is linked to the size of the catchment with the importance of spatial rainfall decreasing at bigger catchment scales [Woods and Sivapalan, 1999]. For small (< 100 km^2) and medium to large (100-2000 km^2) catchments, the spatial resolution of rainfall is important and more precise areal rainfall estimates are required [Arnaud et al., 2002]. Studies show that as the scale increases, the catchment response time distribution becomes the dominant factor governing the runoff generation [Bell and Moore, 2000].

Effects of spatial rainfall are also likely to vary depending on the effect of antecedent catchment conditions [Singh, 1997]. In general, good runoff predictions are obtained with spatially averaged rainfall input under wet conditions. However, for dry catchment conditions, the runoff prediction errors are seen to be considerably higher than for the wet conditions [Arnaud et al., 2002]. This result suggests that there is an interaction between the spatial rainfall and the spatial distribution of soil moisture which controls runoff prediction.

The proportion of rainfall becoming direct runoff is also controlled by the permeability of the catchment and can mask the impact of spatial rainfall [Tetzlaff and Uhlenbrook, 2005]. On pervious catchment areas, rainfall variations are damped by the integrating reaction of the catchment, whereas on impervious areas a high proportion of precipitation becomes effective rainfall. Impervious catchments are fast responding and studies showed that a high density rain gauge network is required for runoff modelling [Berne et al., 2004].

Urbanisation has a great effect on runoff generation. Regarding the significance of spatial rainfall on runoff taking into account that the land use is urban, the conclusion is that small urban catchments require very dense rain gauge networks and weather radar data are an interesting alternative [Dodov and Foufoula-Georgiou, 2005].

Additionally, the runoff response is sensitive to the rainfall type [Koren et al., 1999]. Knowledge of spatial rainfall is important in rainfall-runoff modelling especially when simulating extreme events in the summer [Bell and Moore, 2000]. The error in runoff simulation is decreased for convective rainfall events when spatially high resolution data are applied. However for single frontal events with a longer duration, the spatial distribution of rainfall has less influence on mean catchment rainfall because the low variability of these type of events [Arnaud et al., 2002].

The importance of spatial rainfall is dependent upon the rainfall-runoff modelling strategy. A fine model discretisation is preferable to avoid loss of rainfall information due to averaging at too large scale [Koren et al., 1999]. Although studies found that large errors may arise when a lumped instead of a distributed model is applied to the catchment, the use of the latter type of models introduces large uncertainties in model parameters due to the spatial variability of rainfall [Chaubey et al., 1999].

3. CATCHMENT DESCRIPTION

The significance of spatial rainfall variability was assessed using observed data from the 1040 km^2 Upper Lee catchment (Fig. 1) in the Thames region, UK. The catchment is characterised as humid temperate with mean annual precipitation of 632 mm, while the elevation varies between 20 and 250 meters above UK ordnance datum.

Historical hourly flow and rainfall data were provided by the Environment Agency of England and Wales for the period 1991-1998. The location of the flow stations and raingauges can be found in Fig. 1. In the current analysis only the following gauged catchments are used: Water Hall (150 km^2), Stevenage (31), Hertford (176), Griggs Bridge (50), Wadesmill (136), Mardock (79), Sheering (55), Glen Faba (278), and Feildes Weir (1040).

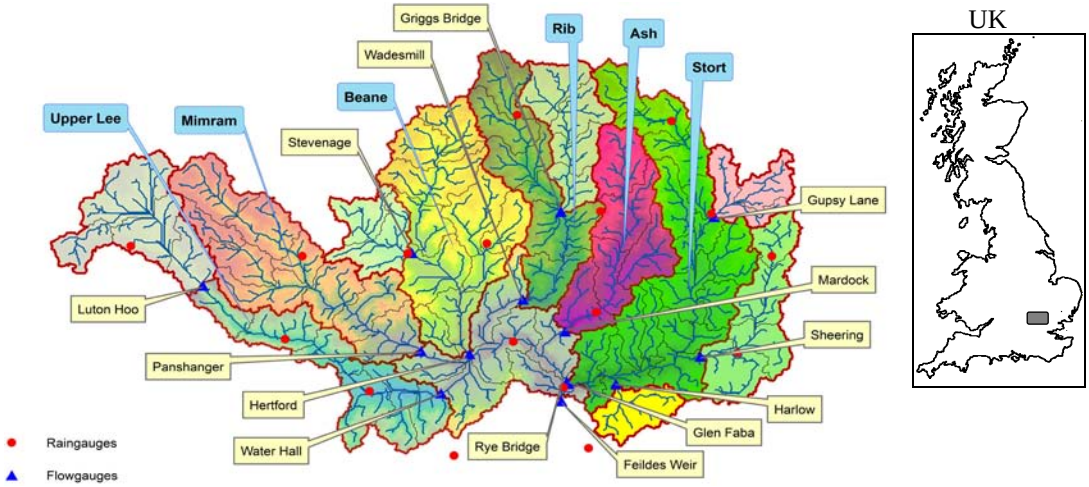


Figure 1. Map of the Upper Lee catchment (main tributaries of the catchment are labelled with blue boxes, and gauged subcatchments with yellow boxes)

The 150 km^2 Upper Lee subcatchment has a flow gauge in Luton Hoo and Water Hall. The Upper Lee at Luton Hoo is mainly chalk with high permeability. Most of Luton Hoo area is urban, followed by arable and horticulture. On the other hand the rest of Upper Lee subcatchment until the Water Hall gauge has a smaller proportion of urban areas. Chalk soil characterizes the 130 km^2 Mimram subcatchment. Although, the catchment is homogeneous and mainly rural, there is a small urban segment due to the city of Welwyn Garden. The Beane subcatchment of 176 km^2 has two flow gauges located on Stevenage and Hertford. It consists of rural and urban area due to the build-area of Stevenage city. The soil type is chalk and clay. The most part of the 136 km^2 Rib subcatchment is rural, with a small proportion of urban. The soil type is chalk which is under laid by extensive deposits of boulder clay. There are two available flow gauges at Griggs Bridge and Wadesmill. The Ash subcatchment is the smallest one with an area of 79 km^2 and is classified as mainly rural. Only a very small proportion of the subcatchment is covered by urban areas. The soil type of Ash is clay. The Stort has the largest subcatchment area of 278 km^2. In the upper part of Stort the area is mainly rural, whereas the valley in the lower part is urban. Three cities of Bishop Stortford, Sawbridgeworth, and Harlow affect the runoff. The geology is chalk and clay. Overall, the 1040 km^2 Upper Lee at Feildes Weir is mainly rural, characterised by arable farming. The area has seen significant growth in housing, with urban areas covering 15 % of the total area.

A relatively fine subdivision of the subcatchments into subunits of about 15-25 km^2 is applied, so that the model spatial discretisation is not the major control on how well the spatial rainfall can be represented. As there is one raingauge per 60 km^2 on average, a smaller subdivision seems unjustified. The locations of the subdivisions are based on the soil and land cover type, and stream network.

4. RAINFALL RUNOFF MODELLING

4.1 Model description

The Probability Distributed Model, PDM [Moore, 2007] assumes that rainfall during each time step accumulates in the soil moisture store, where a specific function is used to describe the distribution of storage capacity over the catchment (Fig. 2).

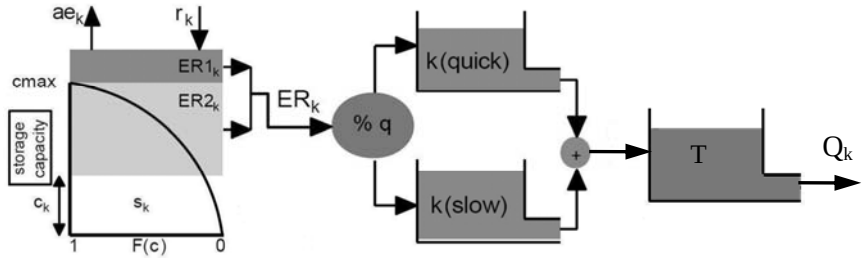


Figure 2. PDM model structure

The soil moisture storage is depleted by evaporation as a linear function of the potential rate and the volume in storage. The soil moisture storage capacity, *c*, is assumed to be described by a Pareto distribution (Eq. 1).

$$F(c) = 1 - (1 - c / c_{max})^{b} \quad (1)$$

where *c* is the storage capacity in the catchment, c_{max} is the maximum capacity at any point in the catchment, and *b* controls the spatial variability of storage capacity over the catchment.

The effective rainfall *ER* is equal to positive values of Eq. 2 integrated over the catchment area.

$$ER_k = r_k - ae_k - c_k + s_{k-1} \quad (2)$$

where *r* is the rainfall (mm) in time step *k*, *s* is the spatially variable soil moisture storage (mm), and *ae* is the actual evapotranspiration which is assumed equal to the potential evapotranspiration multiplied by the relative saturation of the catchment.

The PDM was combined with a routing component consisting of two linear reservoirs in parallel, representing the quick and slow response of the system. The parameter *% q* defines the proportion of total effective rainfall going to the fast response reservoir, while *k(quick)* and *k(slow)* are the time constants of the fast and slow reservoirs respectively. The simulated streamflow at the outlet of the subunit is determined by the combination of the two pathways. The latter is routed to the catchment outlet through a channel routing module consisting of a conceptual linear reservoir with residence time *T*.

4.2 Model identification method

The PDM model parameters for every subunit were first estimated using a regionalisation method. This method is summarised by:

- Calibration and validation (8 and 4 years respectively) of each of the gauged subcatchments in the Upper Lee using a lumped PDM model. The Nash Sutcliffe Efficiency, NSE, was used for the calibration.
- Regression analysis to relate the calibrated model parameters and known catchment characteristics. The catchment characteristics (land cover, soil type, climate and topography) are taken from the Flood Estimation Handbook [details in NERC, 1999].
- Using the regression equations (Table 1), estimate the parameter values for all the subunits within the subcatchments. Hence the equations, developed at the subcatchment scale, are assumed to apply to the smaller subunit scale [details in Pechlivanidis et al., 2007].
- Spatial multipliers are used to optimise the regionalised estimates to gauged streamflow at the case study subcatchments. Whereas the parameter estimates vary spatially, the multipliers do not, hence the values of each parameter are adjusted consistently across the subcatchment. This limits the calibration burden to 6 multipliers per subcatchment, despite the much larger number of parameters. This calibration of the multipliers was done using a uniform random search (URS),

whereby 30,000 sets of the multipliers were randomly sampled from pre-defined ranges and the one which led to the best performance using the NSE for each subcatchment individually was adopted. The initial soil moisture condition was not calibrated. However, the first 20 % of the time-series was used as a warm-up period (during which the discrepancies were neglected in calculation of NSE) in order to remove sensitivity to initial conditions.

Table 1. Relationships of PDM model parameters with catchment characteristics

Equation	Partial R				Total R^2
Cmax = 25093-124.06 10^{-6} URBCONC-6780.56 ln(RMED_2D)	-0.79	-0.94			0.88
b = 0.022 - 202.14 10^{-6} AREA + 3.593 URBEXT1990	-0.92	0.98			0.97
k(quick) = - 843.7 + 24.414 RMED_2D + 17.753 ln(URBEXT1990)	0.76	0.88			0.83
k(slow) = 8755 + 3741.904 ln(BFIHOST) - 18151.5 URBEXT1990	0.62	-0.78			0.64
% q = 1.923 - 0.003 DPLBAR - 0.418 ln(ALTBAR) + 0.013 SPRHOST	-0.83	-0.95	0.97		0.98
T = 30.7+0.074 SAAR -43.811 URBEXT1990+66.538 ln(PROPWET)+4.1 ln(DPLBAR)	0.97	-0.99	0.99	0.99	0.99

4.3 Experiments with spatial rainfall

The investigation of the impact of spatial variability of rainfall on runoff was approached considering three spatial representations of rainfall: 1) No spatial variation in rainfall: rainfall is assumed uniform over the entire Upper Lee catchment; 2) No spatial variation in rainfall at subcatchment scale: rainfall is assumed uniform over each of the 9 subcatchments; and 3) Full spatial variation in rainfall: rainfall is variable within the subcatchments and estimated for each subunit. The arithmetic average of the subcatchment subunits provides the subcatchment and catchment mean areal rainfall. The Inverse Distance Weighting method was used to spatially interpolate rainfall from the 17 raingauges over the entire Upper Lee catchment. A prior statistical analysis did not show strong relationship (R^2 = 0.26) between rainfall distribution and catchment properties (elevation, exposition etc.); therefore the interpolation approach did not considered any topographical aspects.

The PDM model was initially calibrated for each subcatchment and rainfall scenario to assess the effect of spatial rainfall resolution on achievable model performance in terms of NSE over the 8-year calibration period. Then, the model parameter sets were kept constant at the values optimised using rainfall Scenario 3, and the three scenarios of rainfall were applied in turn, to investigate the effect of spatial rainfall on runoff. This was evaluated by measuring the change in NSE, and the percentage change from the observed data in the peak runoff, time to peak, and runoff volume for a rainfall event. Overall, five rainfall events were selected from the simulated period, which represent different types of rainfall event as described in Table 2.

Various spatial properties of rainfall (e.g. variability location, mean, volume, maximum, duration, antecedent conditions, and evaporation) have been estimated for each event. However, the current analysis only focuses on the spatial variability of rainfall. The Spatial Deviation Index (I_{SDI}) [Segond et al., 2007] was defined for each storm event and catchment as a measure of spatial variability of rainfall, with high values of I_{SDI} indicating high spatial variability.

Table 2. Rainfall events and spatial characteristics for the Feildes Weir catchment

Date	Duration (hrs)	I_{SDI}	Volume (10^7 m^3)	Mean (mm/hr)	Maximum (mm/hr)	Peak runoff (cumecs)
20/10/1992	18	0.47	2.94	1.68	4.57	27.48
03/02/1994	7	0.209	1.22	1.68	3.5	36.63
22/01/1995	32	0.231	2.46	1.49	4.9	49.38
08/01/1996	26	0.213	2.25	1.28	3.02	35.44
09/11/1998	12	0.273	0.67	0.59	1.36	14.47

5. RESULTS

The effect of spatial rainfall resolution on achievable NSE performance can be found in Table 3. On average for all subcatchments NSE values of 0.72, 0.75 and 0.76 are observed for Scenario 1, 2 and 3 respectively. Although, a finer resolution rainfall (Scenario 3) performs better on average, the difference in model performance between Scenarios 2 and 3 is not considered significant, while in some cases Scenario 2 outperforms Scenario 3. The available raingauge network structure could be a factor influencing the results as it might not support a finer spatial resolution than Scenario 2 (subcatchment scale).

Table 3. Model performance for the 8 years period based on NSE objective function

NSE	Stevenage	Griggs Bridge	Sheering	Mardock	Wadesmill	Water Hall	Hertford	Glen Faba	Feildes Weir
Area (km^2)	**31**	**50**	**55**	**79**	**136**	**150**	**176**	**278**	**1040**
Scenario 1	0.660	0.663	0.657	0.738	0.738	0.727	0.667	0.824	0.842
Scenario 2	0.690	0.686	0.765	0.761	0.740	0.801	0.681	0.868	0.858
Scenario 3	0.689	0.677	0.763	0.776	0.745	0.798	0.684	0.869	0.859

Fig. 3(a) plots the NSE performance against the I_{SDI} for all five events and all subcatchments using the fixed parameter set, illustrating the impact of the three rainfall representations on the NSE. The PDM model performs equally under all the three rainfall scenarios for less spatially varied events (Feb 1994, Jan 1995, Jan 1996 in Table 2). The importance of considering finer rainfall information is enhanced for the other two events. In the latter case, NSE values improved by up to 15 % when Scenario 3 rainfall was considered rather than Scenario 1. The conclusions from Fig. 3(a) are similar to those from previous case studies [Bell and Moore, 2000, Segond et al., 2007].

On Figs. 3(b-d), the catchments are ordered along the x-axis from smallest to largest scale. For each catchment there are five sets of three points plotted, corresponding to three rainfall scenarios for each of the five events. Each set of three points shows how the peak, volume and timing of each hydrograph change using the different representations of spatial rainfall. The Upper Lee is a relatively heterogeneous catchment; hence conclusions should be derived considering land cover and type properties, as well as catchment scale.

The impermeable clay/urban subcatchment of Sheering is most sensitive to the spatial distribution of rainfall with respect to runoff peak and volume (Figs. 3b,c,e). Results indicate the importance of a fine rainfall resolution when more spatially varied rainfall events occur (the first and last of the five events). An underestimation in the peak and runoff volume of 30 and 40 % respectively could occur in small impermeable catchments if spatial rainfall variability is not considered. However, there is significantly less improvement to hydrograph shape in the less spatially variable events. This was acknowledged in previous case studies [Tetzlaff and Unlenbrook, 2005]. Results from the larger impermeable Glen Faba subcatchment show that there might be a decrease in the significance of spatial information as the scale increases. This conclusion was justified by Segond et al. [2007]. It is interesting to see that the small urban subcatchment of Stevenage shows no significant sensitivity to spatial rainfall. The spatial model discretisation and the lack of raingauge stations near the city of Stevenage could explain this result. No clear relationship could be identified for the time to peak (Fig. 3d).

Some of the more permeable subcatchments (Mardock and Hertford) seem to damp the spatial variability of the rainfall input, based on peak and volume results (Fig. 3b,c,f). However, this is not such a significant effect for other permeable catchments (e.g. Wadesmill and Water Hall). Fine rainfall resolution does not significantly improve performance for the three less spatially variable events, while it slightly does so for the other two.

Although the hydrographs were quite insensitive to the spatial representation of rainfall at the largest two catchments (Glen Faba and Feildes Weir), there was no clear evidence that sensitivity decreases at larger scales, as previously reported by Segond et al. [2007].

Further work is underway to include more catchments and more events, and to consider more of the spatial properties of the rainfall, in order to improve the evaluation.

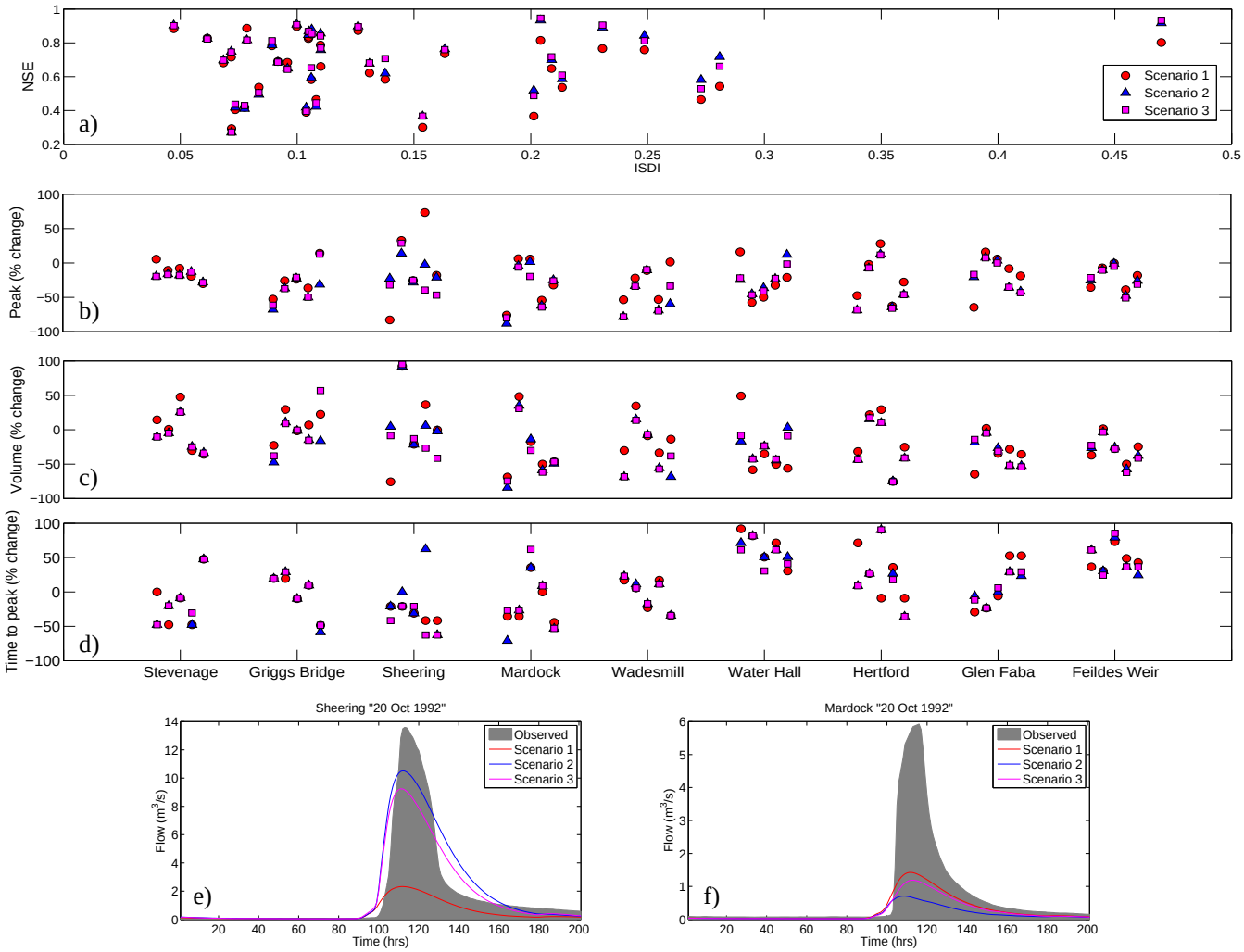


Figure 3. (a) Impact of spatial rainfall representation on NSE model performance; Percentage change from the observed data for each rainfall event in: (b) peak flow, (c) volume, (d) time to peak for each subcatchment ranked according to catchment scale, and (e-f) streamflow simulation for all scenarios for Sheering and Mardock for 20th Oct

6. CONCLUSIONS

The importance of spatial variability of rainfall on runoff as a function of catchment scale and type and rainfall properties was demonstrated in the present study through a semi-distributed rainfall-runoff model. The model was calibrated using calibration multipliers based on a priori parameter values derived from regional relationships between known catchment characteristics and model parameters. Three rainfall scenarios were used corresponding to different degrees of spatial rainfall aggregation over the catchment, while a statistical analysis of five rainfall events indicated the importance of spatial rainfall on model performance and hydrograph properties.

Results show that spatial rainfall can influence the achievable model performance in terms of NSE by up to 15 % when high spatially variable events are investigated. Although it is acknowledged in previous studies, there was no clear evidence that the sensitivity of runoff generation to spatial rainfall is related to catchment scale. Impermeable subcatchments are more sensitive than permeable subcatchments to the spatial variability of rainfall, particularly when spatially varied rainfall events occur. This sensitivity is significantly decreased in the case of less spatially variable events. Two permeable catchments seem to

damp the effect of the spatial variability of rainfall. However, this is not evident for the other permeable catchments. More events should be considered in the future studies to provide more coherent conclusions. Further research should also include use radar data to consider spatial variability of rainfall within the subunits, and also more detailed investigation of the significance of catchment type.

REFERENCES

Arnaud, P., C. Bouvier, L. Cisneros, and R. Dominguez, Influence of rainfall spatial variability on flood prediction, *Journal of Hydrology*, vol. 260, no.1, pp. 216-230, 2002

Boyle, D.P., H.V. Gupta, S. Sorooshian, V. Koren, Z. Zhang, and M. Smith, Toward improved streamflow forecast: value of semi distributed modelling, *Water Resources Research*, vol. 37, no. 11, pp. 2749-2759, 2001.

Bell, V.A., and R. Moore, The sensitivity of catchment runoff models to rainfall data at different spatial scales, *Hydrology and Earth System Sciences*, vol. 4, no. 4, pp. 653-667, 2000.

Berne, A., G. Delrieu, J.D. Creutin, and C. Obled, Temporal and spatial resolution of rainfall measurements required for urban hydrology, *Journal of Hydrology*, vol. 299, no. 3-4, pp. 166-179, 2004.

Beven, K., *Rainfall-runoff modelling - The Primer*, John Wiley and Sons, UK, 2001.

Carpenter, T.M., and K.P. Georgakakos, Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, *Journal of Hydrology*, vol. 329, no. 1-2, pp. 174-185, 2006.

Chaubey, I., C.T. Haan, S. Grunwald, and J.M. Salisbury, Uncertainty in the model parameters due to spatial variability of rainfall, *Journal of Hydrology*, vol. 220, no. 1-2, pp. 48-61, 1999.

Dodov, B., and E. Foufoula-Georgiou, Incorporating the spatio-temporal distribution of rainfall and basin geomorphology into nonlinear analyses of streamflow dynamics, *Advances in Water Resources*, vol. 28, no. 7, pp. 711-728, 2005.

Koren, V.I., B.D. Finnerty, J.C. Schaake, M.B. Smith, D.J. Seo, and Q.Y. Duan, Scale dependencies of hydrologic models to spatial variability of precipitation, *Journal of Hydrology*, vol. 217, no. 3-4, pp. 285-302, 1999.

Moore, R.J., The PDM rainfall-runoff model, *Hydrology and Earth System Sciences*, vol. 11, no. 1, pp. 483-499, 2007.

Natural Environment Research Council (NERC). Soil Biodiversity Programme Data Policy. *NERC*, Swindon, 1999. (http://soilbio.nerc.ac.uk/data_policy_and_strategy.htm)

Pechlivanidis, I.G., N. McIntyre, and H.S. Wheater, Relating conceptual rainfall-runoff model structure and parameters to catchment characteristics: A case study on the Upper Lee catchment, in *HYDRA Postgraduate Research Poster Meeting*, 17 December, UCL, London, UK, 2007. (http://www3.imperial.ac.uk/pls/portallive/docs/1/35545700.PDF)

Segond, M-L., H.S. Wheater, and C. Onof, The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK, *Journal of Hydrology*, vol. 347, no. 1-2, pp. 116-131, 2007.

Singh, V.P., Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrological Processes*, vol. 11, no. 12, pp. 1649-1669, 1997.

Smith, M.B., V.I. Koren, Z.Y. Zhang, S.M. Reed, J.J. Pan, and F. Moreda, Runoff response to spatial variability in precipitation: an analysis of observed data, *Journal of Hydrology*, vol. 298, no. 1-4, pp. 267-286, 2004.

Tetzlaff, D. and S. Uhlenbrook, Significance of spatial variability in precipitation for process-oriented modelling: results from two nested catchments using radar and ground station data, *Hydrology and Earth System Sciences*, vol. 9, no. 1-2, pp. 29-41, 2005.

Wagener, T. and H.S. Wheater, Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *Journal of Hydrology*, vol. 320, no. 1-2, pp. 132-154, 2006.

Woods, R., and M. Sivapalan, A synthesis of space-time variability in storm response: rainfall, runoff generation and routing, *Water Resources Research*, vol. 35, no. 8, pp. 2469-2489, 1999.

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Integration of In-situ and remote sensing Data for Water Risk Management

Bernhard H.C. Sputh[a], **Oliver Faust**[a], **Lasse H. Pettersson**[b], **Torill Hamre**[b], **Damiano Vitulli**[c], **Alastair R. Allen**[a], **and Tim Spracklen**[a]

[a]*School of Engineering, University of Aberdeen, Aberdeen AB24 3UE, UK ({b.sputh, o.faust, a.allen, tim_spracklen}@abdn.ac.uk)*

[b]*Nansen Environmental and Remote Sensing Center, Thormøhlensgt, 47, N-5006 Bergen, Norway ({Torill.Hamre, lasse.pettersson}@nersc.no)*

[c]*Sysmedia S.r.l., Via Paduni 2a, 03012 Anagni (FR), Italy (damiano@sys-media.it)*

Abstract: The combination of environmental measurements from different sensor equipment improves water quality monitoring and associated risk management. The merging of in-situ and remote sensing data is a prerequisite to efficient monitoring of coastal and marine environments, because of their complementary nature and application areas. Both sensor types can monitor the same phenomena, but remote sensing delivers the overall picture and in-situ measuring provides localised, high resolution measurements, over a wide range of parameters. The WARMER (**WA**ter **R**isk **M**anagement in **E**u**R**ope) system combines remote sensing satellite data with in-situ measurements for water quality risk management applications and delivers products to end-users through a web GIS (Geographic Information System) portal. This data combination aspect brings the WARMER system in line with the objectives of the GMES (European Global Monitoring for Environment and Security) initiative.

The ADDP (Aberdeen DISPRO DB Proxy) is a key element in the WARMER system, because it conditions in-situ measurements for the data combination. This paper details the handling of in-situ data within the ADDP, especially its capability to aggregate in-situ data from multiple sources. This aggregation caused problems with data synchronisation and unification, which influenced the overall design of the ADDP.

Keywords: FP6 IST, water quality, risk assessment, remote sensing; in-situ measurements; UMN MapServer; OGC-WMS

1 INTRODUCTION

Coastal waters have many different users, such as: maricultures, fisheries, shipping, and tourism. Each user has specific environmental quality requirements, and is a potential source and victim of pollution. Water quality monitoring is the only way to protect victims and identify the source of pollution. For instance, Pablakis et al. [2001] point out that shipping is a constant source of oil pollution, due to accidental events and deliberate operations. Oil pollution affects fisheries and maricultures by being harmful to or by poisoning sea-living animals and birds. Furthermore, it has a degrading impact on the environment and hence tourism industry. Troell et al. [2003] report that maricultures pollute the environment with feces, antibiotics and nutrients. The high concentration of sea-living animals, inside maricultures, results in a discharge of higher than normal concentration of feces. Furthermore, animals bred in maricultures require additional nutrients. Naturally, not all nutrients get consumed by the caged fish and these surplus nutrients leave the maricultures,

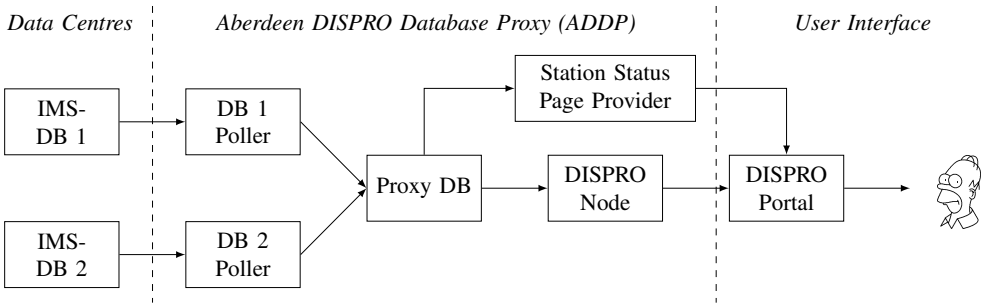


Figure 1: Overview of ADDP and its surrounding elements

either to be consumed by animals outside the mariculture, or to rot on the sea-bed. Finally, maricultures are often mono-cultures, making them vulnerable to diseases and parasites. To cure or protect the animals, antibiotics and other medications are used which leave the mariculture and thus pollute the environment.

The EC WFD (Water Framework Directive) urges the national environment protection agencies to improve the water quality of all their water resources, including coastal waters (see European Commission [2000]). To improve the water quality the level of pollution must be detected and if possibly its origin determined. This can be achieved by regular water quality monitoring. To monitor the quality of coastal waters two methods are available: in-situ measurements using buoys, ships, unmanned vehicles etc. and remote sensing from satellites or aircraft. In-situ measurements are performed by IMSs (In-situ Monitoring Stations) located on buoys or moving platforms such as ships. IMSs monitor the water quality with high resolution, but are limited to the "buoy" location. Monitoring large areas, such as a coastal area, requires a network with a large number of IMSs. This results in high initial deployment and operational maintenance costs. In contrast, remote sensing offers a comparatively cheap way, disregarding the total cost of the satellite, to obtain measurements over large areas and it provides a synoptic mapping of a large surface of the environment. The shortcomings of remote sensing are its limited number of observational environmental parameters and their accuracy, as well as the spatial and temporal resolution of the data.

The WARMER project has developed and implemented advanced IMS sensor technology for high accuracy measurements of a range of "water quality" parameters, using innovative in-situ sensor technologies. Several of the developed sensors are able to measure low concentrations of environmental contaminants present in the natural environment. However, even these low concentrations of environmental contaminants already represent a high risk factor for human health and environmental safety. Several of these environmental contaminants do not have other distinct attributes which would allow indirect detection using e.g. remote sensing — such as water discoloration, which applies for e.g. phytoplankton, sediments and dissolved organic matter, or surface film forming attributes, which applies for e.g. crude oil or biological surfactants.

Apart from advancing the present state of sensor technology, the WARMER project aims to provide a cost effective, all embracing view on the current water quality situation for risk management in aquatic environments, such as coastal areas. We achieve this all embracing view by combining water quality measurements from remote sensing and in-situ sensors. This paper presents our current work on the ADDP (Aberdeen DISPRO[1] DB Proxy), which transforms in-situ measurements to an OGC (Open Geospatial Consortium) compatible format. These data are made available to web-based remote sensing data presentation systems. The presentation of all relevant environmental monitoring information within the same user interface makes this data as easy to access and comprehensible. According to Hepting and Haciag [2006], easy access and comprehensible presentation of environmental data is a vital aspect of such systems. The remote sensing data helps also to identify optimal IMS locations, in order to obtain representative measurements.

[1] DISPRO stands for DISMAR (Data Integration System for Marine Pollution and Water Quality) Prototype.

Accordingly, the system offers cost effective high quality monitoring of coastal areas.

Figure 1 illustrates the information flow within the proposed system. The Data Centres provide in-situ data to the ADDP, see Section 2.1. The ADDP, discussed in Section 3, retrieves the in-situ data, processes it, and provides it to the User Interface. The User Interface (Section 2.3) accesses the remote sensing data held by a DISPRO Node and fuses it with the processed in-situ data.

2 PREREQUISITES

This sections details the sources of water quality information within WARMER: in-situ measurements and remote sensing. The section closes by detailing the DISPRO system we chose to perform the data combination and which also provides the user interface.

2.1 In-situ Measurements

In WARMER a network of instrumented stations obtain in-situ measurements. These IMSs (In-situ Monitoring Stations) measure water quality parameters, such as: temperature, salinity, chlorophyll, oxygen content, and aquatic pollutants. These measurements are taken at fixed locations, given depths, and in specific time intervals. Each measurement is a tuple which consists of timestamp, location, measured parameter, and measurement value. After obtaining a measurement the IMS transfers it to a data centre, which acts as a central repository and includes automatic quality control routines for the first data assessment. From there the measurement data can be retrieved and assessed by multiple users and applications.

In-situ measurements have a high temporal resolution, with a sampling period from seconds upwards, however at a fixed location. This high temporal resolution prevents missing or misinterpreting, important events, such as high levels of pollution over a short time period, caused by a discharge of pollutants. IMSs are located at critical or representative locations, for instance potential water pollution sources (factories, maricultures, etc.), or where water pollution may cause severe harm (maricultures, water intakes etc.). The IMS measurements may also provide the vertical structure of the water column.

The highly localised monitoring provided by IMSs makes it a challenge to monitor large areas, because the spatial resolution of the field measurement is directly proportional to the number of stations. For example, one possible WARMER field trial is to monitor the water quality of the Boknafjord in Norway. The Boknafjord has roughly $1,000\mathrm{km}^2$ of water surface. Therefore, to achieve a spatial resolution of 1 measurement per km^2 (a typical resolution for remote sensing satellite data) we would need to deploy around 1000 IMSs. This is an impossibly high number of IMS to be deployed, for at least two reasons: (1) high cost of both initial deployment and operational maintenance; (2) deploying a large number of IMSs effectively prevents shipping, which is an undesirable impact on the usability of the area. To monitor an area such as the Boknafjord one is typically only permitted five to ten station locations.

Different applications require access to different measurement parameters. A data centre stores the measurements in a dedicated DB (database), the so-called IMS-DB. In the WARMER system there is only one data centre operated by Sysmedia (Sysmedia [2008]), based in Italy. Sysmedia is responsible for quality control and validation of the stored information.

2.2 Remote sensing

Only a subset of parameters measured by the WARMER buoy system is directly fused with any of the existing (and planned) satellite remote sensing technologies and applications. Accordingly, an integrated monitoring approach using a facet of optical, infrared and microwave satellite remote sensing sensors will be implemented for the demonstration areas. The integrated use of these information sources will form the basis for the assessment of emergency / environmental situations.

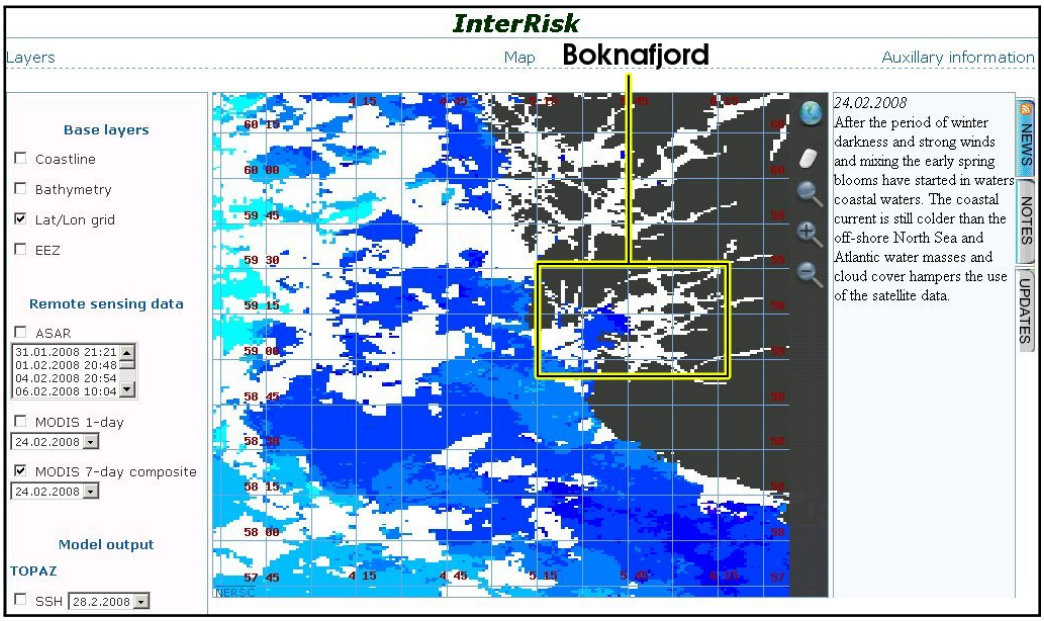


Figure 2: Screenshot of the DISPRO Client user interface, to be used by the WARMER project

The proposed suite of EO (Earth Observation) sensors includes the European Space Agency Envisat sensors - MERIS FR (Medium resolution Imaging Spectrometer 10's meters). These EO sensors however, meet neither the requirements of frequent spatial coverage, nor near real-time availability of the data.

Accordingly, the project demonstrations will take place at coastal locations where one may expect gradients in the level of contaminants. These demonstrations are used to assess the project development. In addition, the information should be supplemented with higher spatial resolution satellite remote sensing information from optical sensors, such as the Ikonos sensor.

Figure 2, shows a 7-day composite MODIS sea surface temperature (in February) of the Boknafjord and abutting coastal areas. Establishing even this, rather low, spatial resolution of the water quality situation in the Boknafjord with IMSs is impractical due to the high cost and its previously mentioned impact on shipping. This clearly demonstrates why combining remote sensing with in-situ measurements is essential for the efficient monitoring of large aquatic environments.

2.3 Data Combination and User Interface

In WARMER we use the DISPRO system to combine remote sensing and in-situ data. DISPRO has been developed by the DISMAR Consortium, coordinated by the Nansen Center (Tuama and Hamre [2007]). We have extended the DISPRO functionality to suit the special needs of the WARMER project.

The DISPRO system includes a set of independent DISPRO Nodes. Each node provides an OGC-WMS (Web Map Service) interface. The DISPRO Portal communicates with the nodes and provides various client applications, such as map viewer, profiler, and metadata catalogue. Figure 1 illustrates the relationship between DISPRO Portal, DISPRO Node and users, in the context of the WARMER application.

The implementation of the DISPRO system relies heavily on Java and XML technologies, with the Open Source UMN (University of Minnesota) MapServer (UMN [2007]), running as a CGI application under the Apache web server, providing the core web GIS (Geographic Information System) client and OGC WMS (both client and server) functionality. The implementation of the main DISPRO components is described in the following.

The metadata catalogue is at the heart of the DISPRO system. It stores data derived from capabilities files returned by participating nodes in response to a GetCapabilities request issued by the portal aggregator. A capabilities file consists of information about the WMS node itself and about each map layer that it can serve. This aggregated dataset underpins DISPRO, and is updated every 15 minutes (configurable). The distributed data structure assures that the latest data from each provider is available in the system.

The web GIS interface (Figure 2) allows a user to view the map layers in a predefined area of interest and map projection. Standard functions, such as select/deselect layer, pan, zoom, and centre are provided, and rely on basic web browser capabilities (no plug-ins required). The DISPRO user interface will be extended with in-situ monitoring specific features, such as the SSP (see Section 3.4).

3 THE ABERDEEN DISPRO DB PROXY

The block diagram in Figure 1 shows that the ADDP acts as an intermediate service between Data Centres and User Interface. The DISPRO Portal provides user interface and data combination capabilities. To increase system performance and reliability the ADDP includes an internal storage service for in-situ data.

The ADDP aggregates in-situ measurement data from the IMS-DBs (In-situ Monitoring Stations-Data Bases) within the data centres. For this aggregation the ADDP utilises multiple DB Pollers, discussed in Section 3.1, which periodically check the IMS-DBs for new in-situ data. The DB Pollers store new measurement data in the Proxy DB, see Section 3.2. The DISPRO Node projects the data stored in the Proxy DB onto a map, which is then fused by the DISPRO Portal with remote sensing data, Section 3.3 gives more details. The S2P2 (Station Status Page Provider), detailed in Section 3.4, provides a SSP (Station Status Page) for each IMS, which, among other information, presents time series plots of in-situ measurements. This allows operators to assess the wider range of in-situ data. These SSPs are available from within the user interface provided by the DISPRO Portal.

3.1 DB Pollers

The DB Pollers have a number of different functions. They obtain all relevant information from the IMS-DB of a Data Centre and store it in the Proxy DB. This involves transforming the obtained information into a format suitable for the Proxy DB. A typical example of such a transformation is the conversion of the station location from latitude/longitude format to PostGIS geometry format (Ramsey [2001]).

In relational databases, such as Proxy DB and IMS-DBs, data sets refer to one another using each others unique IDs. The Proxy DB allows data aggregation from multiple Data Centres by assigning new unique IDs to the data sets. The DB Pollers translate the IDs within the data sets retrieved from the IMS-DBs into Proxy DB IDs. For this purpose each DB Poller has a set of tables which contain the relationship between the IDs of IMS-DB and Proxy DB.

Another problem is that we must assume that the database layouts of the external DB and the Proxy DB differ. To overcome this, we use a special event driven technique to import data from an IMS-DB into the Proxy DB. One aspect of this technique is that the data handling is first sensor and then measurement centric. That means, the DB Poller first determines the available sensors within an IMS-DB. In a second step it checks whether or not these sensors have any

measurements unknown to the Proxy DB. When this is the case, these measurements get added to the Proxy DB. The data set that represents a measurement is not atomic, i.e. it refers to other data sets. The measurement data set refers to the data set describing the sensor that took it. The data set, describing the sensor, has therefore to be present in the Proxy DB as well. The system handles this when a DB Poller tries to add a new data set. During this process the system checks whether or not all data sets the new data set refers to are available within the Proxy DB. If a missing data set is discovered the system requests the DB Poller to add it, before the initial data set may be added. This guarantees the consistency of the Proxy-DB contents.

Apart from updating the Proxy DB with new information from the IMS-DBs, the DB Pollers also trigger the generation of mapfile and XML metadata files. This happens once a new sensor-type gets added to the Proxy DB. This includes the generation of a legend image for this sensor type, which maps colours to measurement values, using the same colour scales for similar parameters.

3.2 Proxy DB

The Proxy DB is a spatial enabled (PostGIS, see Ramsey [2001]) data base (PostgreSQL, see Pos [2008]). The ADDP uses it to store all water quality information that the DB Pollers retrieved from the IMS-DBs. This local storage of water quality information increases system performance and reliability.

Presently, the Proxy DB uses four tables to hold the retrieved data: Measurements, Sensors, SensorTypes, and Stations. The rows of the Measurements table hold individual measurements. Each row consists of: Measurement-ID, time, location, value, and the Sensor-ID of the sensor that performed the measurement. The Sensor-ID refers to a row of the Sensors table, where each row has the following columns: Sensor-ID, SensorType-ID, and Station-ID. The table SensorTypes contains detailed information for each different sensor type, the main elements are: SensorType-ID, name, unit, precision, minimum value, and maximum value. The last table of the Proxy DB is the table Stations which contains entries consisting of: Station-ID, name, location. Furthermore, the Proxy DB automatically generates a table which contains the most recent measurement of each sensor. This is achieved using a stored procedure within the Proxy DB. This procedure is called by the PostgreSQL server once a new measurement is added to the Measurements table.

3.3 DISPRO Node

The DISPRO Node component of the ADDP interfaces directly to the DISPRO Portal. It provides the portal with the latest in-situ measurements, available in the Proxy DB. The data gets served by UMN MapServer running as a web application within the Apache HTTP server. The UMN MapServer provides specifications of available data layers within a so-called mapfile. In the case of in-situ measurements, each sensor measurement is represented by a coloured symbol, at the location of the IMS. The symbol colour represents the values of the various measured parameters. A mapfile consists of one or more layers. Each layer groups similar information, allowing for different parameters to be presented within the same mapfile. The ADDP uses this feature by grouping all measurements obtained by the same sensor-type within one layer, i.e. all temperature measurements in one layer, all chlorophyll measurements in another layer, and so on. This allows operators to select between the different parameters.

The UMN MapServer obtains the data it projects directly from the Proxy DB. This is possible with UMN MapServer if the DB is spatially enabled. One example for such a spatially enabled data base is PostGIS (Ramsey [2001]). Using UMN MapServer in conjunction with a spatially enabled DB makes it unnecessary to generate a new mapfile every time new data becomes available. In such a configuration it is only necessary to generate a new mapfile when a new layer gets added. This is the case when a DB Poller adds a new sensor-type to the Proxy DB. This triggers the generation of the new mapfile.

For each mapfile layer the ADDP DISPRO Node provides additional information, such as: title, legend image, and a XML metadata file. The legend image relates station symbol colours and measurement values within the user interface. The XML metadata contains, among other information, source and contact information related to the sensors represented within this layer.

3.4 Station Status Page Provider

The S2P2 (Station Status Page Provider) provides a SSP (Station Status Page) for each IMS in the system. This page contains information about the station, i.e. location, operational status, available sensors, and other parameters. Furthermore, it provides a time series plot over past measurements for each sensor. These time series plots allow operators to assess what happened in the past.

User interfaces such as the SSPs should provide a responsive look and feel. To achieve this the SSPs are implemented using Ajax (Asynchronous Javascript And XML) technology. The Ajax technology allows us to develop the SSP as a GUI application executing within standard web browsers. Similarly to normal applications, the SSP communicates with the S2P2 backend server to obtain or commit information. We implement the S2P2 using the GWT (Google Web Toolkit), (Goth and Costlow [2007]). The GWT allows us to develop the whole application with the Java programming language. The GWT translates the resulting Java program into an AJAX enabled web page, the SSP. The SSP communicates with the S2P2 via a dedicated protocol, which gets generated by the GWT as well.

4 Conclusion & Further Work

Coastal waters have many different users, each user is a potential source and "victim" of pollution. Water quality monitoring is the only way to protect this environment in order to preserve its usability. The WARMER project integrates in-situ and remote sensing data to monitor water quality. After introducing both types of environmental monitoring, the paper discusses how both types of monitoring complement each other for the monitoring of large aquatic environments. The proposed system represents an all embracing tool for water monitoring and risk management of coastal and marine environments. Based on the ADDP structure this paper has outlined a system which combines in-situ and remote data.

The ADDP system is composed of independent components which are arranged around a central data base. This data base aggregates the contents of multiple Data Centres. The data base unifies the data access for the different system components. This makes the DISPRO Node and S2P2 components independent from the data layout of the different data centres. The result is an easy to integrate, extendible, responsive and reliable information handling and assessment system.

Its easy integrability and extendability allows its use outside the WARMER project boundaries, thus allowing others to take advantage of the combination of in-situ and remote sensing data. The ADDP will be released as Open Source, thus a community of ADDP users and developers could emerge. Over time this community will shape the ADDP to suit the different data presentation and data aggregation needs and thus turn the ADDP into an industrial strength tool for environmental monitoring.

The system presented in this paper is only one step towards the combination and presentation of remote sensing and in-situ monitoring measurements. It will be tested in the WARMER project field trials, during 2008. The next ADDP development step is to integrate mechanisms to present and handle alarms generated by IMSs. This will require additions to most ADDP components. Furthermore, to demonstrate the power of the divide and conquer approach behind the ADDP design more ways to access the data of the Proxy DB will be developed. One example would be to implement an OGC-SOS (Sensor Observation Service) component, following OGC-SWE (Sensor Web Enablement) recommendations, detailed in OGC [2006].

ACKNOWLEDGEMENTS

This work has been supported by the EC FP6 project WARMER (**WA**ter **R**isk **M**anagement in **E**u**R**ope) project (contract no.: FP6-034472). The DISPRO system was first developed in the EC FP5 DISMAR (Data Integration System for Marine Pollution and Water Quality) project (contract no.: IST-2001-37657) and is currently being enhanced in the EC FP6 InterRisk (Interoperable GMES Services for Environmental Risk Management in Marine and Coastal Areas) project (contract no.: FP6-035253).

REFERENCES

European Commission. Directive 2000/60/EC of the European Parliament and of the council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities*, L327(1), 22 December 2000.

Goth, G. and T. Costlow. Currents: The Google Web Toolkit Shines a Light on Ajax Frameworks. *IEEE Software*, 24(2):94–98, 2007. DOI: http://doi.ieeecomputersociety.org/10.1109/MS.2007.39.

Hepting, D. H. and T. Haciag. Opening Access to Environmental Software Systems. In *Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software"*. International Environmental Modelling and Software Society, July 2006.

OGC. OGC® Sensor Web Enablement: Overview And High Level Architecture. Internet, July 2006. http://portal.opengeospatial.org/files/?artifact_id=15540 (last visited 04.03.2008).

Pablakis, P., D. Tarchi, A. J. Sieber, G. Ferraro, and G. Vincent. EUR 19906 EN: On the Monitoring of Illicit Vessel Discharges — A Reconnaissance Study in the Mediterranean Sea. Technical report, EC DG - Joint Research Centre & EC DG - Environment, 2001.

PostgreSQL Global Development Group. *PostgreSQL 8.3.0 Documentation*, 2008. http://www.postgresql.org/files/documentation/pdf/8.3/postgresql-8.3-A4.pdf (last visited 27.02.2008).

Ramsey, P. PostGIS Manual. Technical report, Reflection Research Corporation, 2001. http://postgis.refractions.net/docs/postgis.pdf (last visited 10.02.2008).

Sysmedia. Sysmedia S.r.l. — Home Page. Internet, March 2008. http://www.sys-media.it/ (last visited 07.03.2008).

Troell, M., C. Halling, A. Neori, T. Chopin, A. Buschmann, N. Kautsky, and C. Yarish. Integrated mariculture: asking the right questions. *Aquaculture*, 226(1-4):69–90, 2003.

Tuama, É. Ó. and T. Hamre. Design and Implementation of a Distributed GIS Portal for Oil Spill and Harmful Algal Bloom Monitoring in the Marine Environment. *Marine Geodesy*, 30(1):145 – 168, January 2007. DOI: http://dx.doi.org/10.1080/01490410701296671.

UMN. UMN Mapserver. Internet Homepage, December 2007. http://mapserver.gis.umn.edu/ (last visited 14.12.2007).

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A rainfall-runoff database to support flood risk assessment

S. Liersch and M. Volk

Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
(stefan.liersch@ufz.de)

Abstract: A comprehensive rainfall-runoff database for the Mulde catchment in Central Germany was developed to support flood risk assessment and flood management. A large number of randomly generated 20-day rainfall scenarios and several combinations of model initialization states are the basis for the simulation of basin response by applying a metric conceptual rainfall-runoff model. The rainfall scenarios are divided into three periods; a five day initialization period to represent pre-conditions, a two day storm period representing a rainfall extreme event, and a final period characterized by low rainfall volumes to study streamflow recession behaviour. The rainfall-runoff database can be used as an effective tool to easily assess possible streamflow situations assuming different rainfall volumes for the following days. The benefit of the database approach is that it can be easily used by persons who are not necessarily familiar with hydrologic modelling, because the modelling step has already been accomplished. The study area was strongly affected by the extreme flood event in August 2002. In order to study the rainfall-runoff models' capabilities to capture this extreme event it was calibrated to training sets of input data including and not including this event. The model was calibrated using different objective functions, the Nash-Sutcliffe efficiency and a modified Nash-Sutcliffe efficiency adapted to high flow conditions. Only simulations with model settings calibrated to time series including the extreme event were capturing the flood event sufficiently. Due to the growing complexity in rainfall-runoff modelling during seasons affected by snowmelt processes, the applicability of the database in the current state is limited in snow affected catchments to the warm season.

Keywords: Rainfall-runoff database; Flood risk management; Random rainfall generator; Parsimonious.

1. INTRODUCTION

The frequency of large floods has increased substantially during the twentieth century and has a significant positive trend (Milly et al., 2002). Moreover, floods have the greatest damage potential of all natural disasters worldwide (Smith and Ward, 1998). Recent large floods, such as in the Elbe basin in summer 2002 lead to increased interests in research and highlight the necessity of improved flood forecasting techniques and flood risk assessment. To account for this development programs like the European Flood Directive (Directive 2007/60/EC) have been implemented to assess and manage flood risks.

In this paper we present the development of a comprehensive rainfall-runoff database as a tool for flood risk management. For this purpose a rainfall scenario generator was developed generating flood-relevant rainfall events. These events serve as input for the simulation of runoff by using the metric conceptual rainfall-runoff model IHACRES (Identification of unit Hydrographs and Component flows from Rainfalls, Evaporation and Streamflow data) (Jakeman et al., 1990; Jakeman and Hornberger, 1993). The IHACRES model was calibrated to the Mulde catchment in Central Germany that was strongly affected by the

devastating flood event in August 2002. IHACRES was chosen because of its parsimonious approach to model parameterization and because of the simplicity to initialize the model. The rainfall-runoff scenarios are directly exported to an object-relational database management system (PostgreSQL, www.postgresql.org) in order to handle the data in a proper way.

Flood risk management includes flood forecasting with oftentimes only sparse information available such as observed streamflow and precipitation data of the last days as well as the weather forecast. The rainfall-runoff database can be used here as an effective tool to easily assess possible streamflow situations assuming different rainfall depth for the following days. First, a database query must be performed to select all scenarios where the five day initialization rainfall patterns or depth are similar to the real rainfall pattern of the previous five days. The next step is scenario based where the water manager selects all simulations where the rainfall storm event is in a certain range (rainfall depth in mm). Due to the large number of available simulations in the database the result is a variety of simulated hydrographs, reflecting the uncertainties of rainfall forecasts and model results. A benefit of this approach is the minimal time required by the database queries to achieve the desired results. The database was tested for several flood events that occurred during the period 1983 to 2002 achieving good results.

2. STUDY SITE

The Mulde river basin in Central Germany is one of the major tributaries of the Elbe River with an area of approximately 5,400 km² (at gauge Golzern) and consists of three larger sub-basins: The Freiberger Mulde draining the Central Ore Mountains, the Zwickauer Mulde which drains the western Ore Mountains and the Vereinigte (unified) Mulde. The altitudes in the basin reach between below 50 m to above 1200 m.a.s.l. 60% of the basin is used as farmland with high proportions of drainage-tiled areas, followed by forests with 17%, urban areas with 10%, pasture with 10% and 3% for others. Several flood protection measures have been implemented in the basin because of an increasing number of catastrophic floods during the last decades (especially in the year 2002). Extremely high discharge events, originating in the large mountainous parts of the upper Mulde catchment area after intensive rainfall events, have been always the basis of hydrological and landscape-forming development of the river course and its adjacent countryside.

3. METHODS

3.1 Rainfall scenario generator

The basis of the development of the rainfall-runoff database for flood events are rainfall scenarios of a length of 20 days. Hlavcova et al. (2005) state that there is no real preference for a certain method to estimate design rainfall events until now. Furthermore, the rainfall-runoff model used in our study requires only average catchment time series of precipitation and temperature to simulate streamflow. Thus, it was not necessary to apply sophisticated methods to generate distributed rainfall events as has been accomplished for instance by Gabellani et al. (2007). For this reason we used a random number generator to produce the required rainfall scenarios. The scenarios are divided into three periods where the length of the periods and the rainfall amounts can be defined by the user. We chose a length of five days for the first period - the initialization period - and allowed rainfall depth to range between 0 and 80 mm (representative values for the study area). The rainfall initialization period is not to initialize the model, as described in the following section, but to capture uncertainties in rainfall patterns and measurements. The random generator first defines for each scenario a rainfall depth for the initialization period between 0 and 80 mm and distributes the value randomly over the five days. Thus, a five day rainfall period is generated, representing a variety of artificial (or not yet occurred) as well as real rainfall events, capturing a large spectrum of different volumes and intensities. Following the initialization period a two day storm event is calculated randomly based on the same approach. We defined possible rainfall depth for the storm event in the range of 40 to 180 mm to be distributed

over the two days. Instead of two days the user could also define a different duration of the storm event. An analysis of the rainfall time series in the Mulde showed that the duration of two days for the storm event is adequately representing the natural conditions. In order to study the streamflow recession behaviour a third period was introduced characterized by low rainfall events. We chose the duration of 13 days with rainfall depth of 0 to 20 mm over this period. The rainfall scenario generator was implemented in the programming language C++ and was equipped with a database interface. The 20-day rainfall scenarios are directly exported to an object-relational database management system (PostgreSQL).

Due to the large number of rainfall scenarios, both artificial rainfall patterns and natural rainfall patterns are captured, although the scenarios are generated randomly: Figure 1 shows an example of the observed rainfall event during the flood in August 2002 as well as two generated scenarios from the database capturing the pattern of the observed event.

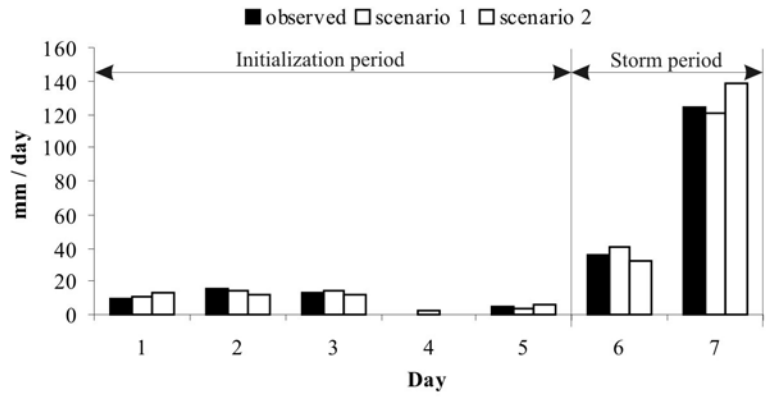


Figure 1. Randomly generated rainfall scenarios capturing the observed rainfall event in August 2002 in the Mulde catchment

Additionally a 20-day time series of daily mean temperature is produced by the generator. Here we focus on flood events during the summer period and calculate a time series representing typical temperatures in the months June, July, August and September.

3.2 Rainfall-runoff modelling

The IHACRES metric conceptual rainfall-runoff model (Jakeman et al., 1990; Jakeman and Hornberger, 1993) has a parsimonious approach to model parameterisation. The version used in this paper has six free calibration parameters. IHACRES has been applied to catchments with a wide range of climatologies and sizes (Croke et al., 2004). It has been used to predict streamflow in ungauged catchments (Kokkonen et al., 2003; Post and Jakeman, 1999; Post et al., 1998), to study land cover effects on hydrologic processes (Croke et al., 2004; Kokkonen and Jakeman, 2002), and to investigate dynamic response characteristics and physical catchment descriptors (Kokkonen et al., 2003; Sefton and Howarth, 1998).

Due to the minimal data requirements IHACRES can be applied in many catchments without spending a long time to prepare necessary input data. The model requires only time series of precipitation and temperature to simulate catchment runoff. Observed streamflow data are used for calibration. As illustrated in figure 2 a rainfall (r_k) time series is converted to effective rainfall (u_k) in the non-linear loss module (Jakeman and Hornberger, 1993). In order to achieve this, a catchment wetness index or antecedent precipitation index, representing catchment saturation is calculated for each time step. In the linear routing module the effective rainfall is converted to streamflow (x_k). We used a storage configuration of two parallel storage components, a quick (x^q) and a slow component (x^s). The parameter (α) is the recession rate for the quick and slow storage component and parameter (ß) represents

the fractions of effective rainfall (u_k) for peak response. A comprehensive description of the model functions used in our study is given in Jakeman and Hornberger (1993). In order to avoid long model warm-up periods, the streamflow simulations based on each rainfall scenario are calculated starting with different initial model states. These model states represent a variety of possible pre-conditions (wet to dry) at time step zero (t_0). The parameters required for model initializations are the *antecedent precipitation index* (API) and *streamflow* (Q). Initial combinations of these parameters are stored in a table where for instance the first combination is Q = 10 m^3/s and API = 0.1, the second combination Q = 10 m^3/s and API = 0.2 and so on. In our case we used 220 pre-condition combinations. Thus, for each rainfall scenario 220 streamflow simulations were performed. The model initialization should not be mixed up with initialization period of the rainfall scenario described in the previous section.

The IHACRES model was implemented as a module for the open source GIS SAGA (www.saga-gis.org) and has been equipped with a calibration tool based on the Monte Carlo approach which is appropriate to calibrate six free parameters in a short time. Due to the parsimonious parameterization of the model and the minimal data requirements, a simulation of a long time series can be calculated in a few seconds for instance. These are optimal conditions to realize thousands of simulations within a short time which was necessary to develop the rainfall-runoff database.

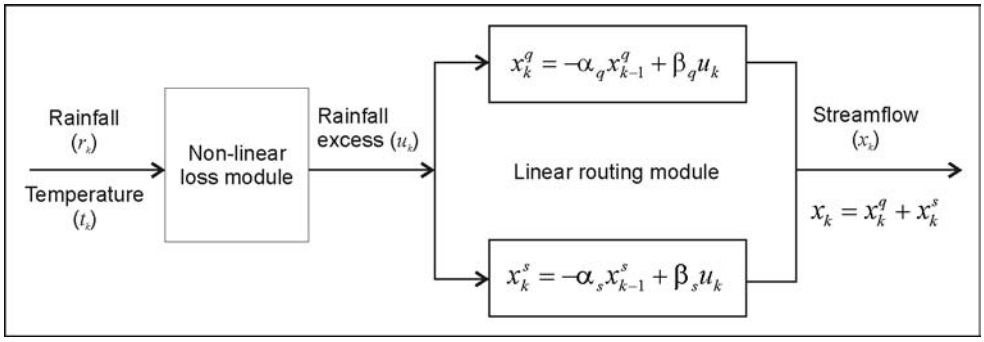


Figure 2. IHACRES model (after Jakeman and Hornberger, 1993; Croke et al., 2004)

Calibration

The model was calibrated at gauge Golzern with a catchment size of 5400 km^2. Arithmetic mean time series of all available precipitation and temperature gauges in the catchment served as model inputs. The period 1995 to 2002 (period 1) was chosen, because it includes the extreme flood event in August 2002 and consists of a variety of hydrologic conditions, namely wet, dry and normal years, as shown in table 2. In order to learn about the models' capabilities to capture extreme events with magnitudes not included in the training set, period 2 (1995 to 2001), not incluing the 2002 extreme event, was used to calibrate the model with a second set of parameters adjustments. For model validation the period from 1983 to 1994 was selected. The performance of the model calibration was measured on a daily time step using three objective functions, the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), a modified Nash-Sutcliffe efficiency (Hoffmann et al., 2004) adapted to high flow conditions (ANSE) and the PBIAS in [%].

For each calibration period the model was calibrated focussing on the objective functions NSE and ANSE. Hence we obtain four simulations characterized by different model parameter settings. To distinguish the different model versions the nomenclature shown in table 1 is used in the following.

Table 1. Nomenclature of model versions

	NSE	**ANSE**
Period 1	NSE_P1	ANSE_P1
Period 2	NSE_P2	ANSE_P2

Model performance

Reasonable model performance for all model versions was obtained for the calibration period in all years (Table 2). Only in the years 1996 and 1997 the efficiency dropped below 0.6. In the validation period the performance is also reasonable except for the years 1983-85 and 1990 where the Nash-Sutcliffe efficiency is below 0.5. Noticeable is that these years can be characterized as dry years. Generally the model performs much better in the Mulde catchment in wet years than in dry years. During the wet years mean Nash-Sutcliffe efficiencies of 0.83 were obtained (NSE_P1), however in dry years the mean NSE was with 0.54 considerably lower. It should be emphasized here that the model versions calibrated for period 2 seem to be more robust than the versions calibrated for period 1, particularly in the validation phase. An explanation for this phenomenon could be that the enormous runoff volume during the storm event in 2002 (not included in calibration period 2) corresponds approximately to the volume of 60 to 70 days of mean discharge. Thus, the 2002 flood event is influencing model calibration considerably distorting the parameter settings towards the achievement of a good fit to this event.

Table 2. Model performance and comparison

	NSE_P1			ANSE_P1			NSE_P2			ANSE_P2		
Year	**NSE**	**ANSE**	**pbias**	**NSE**	**ANSE**	**pbias**	**NSE**	**ANSE**	**pbias**	**NSE**	**ANSE**	**pbias**
Cal.												
1995 w	0.79	0.85	-3.9	0.72	0.78	-5.5	0.80	0.80	-2.6	0.81	0.83	-5.9
1996 n	**0.59**	0.74	14.1	**0.41**	0.62	10.0	**0.55**	0.71	16.6	**0.54**	0.70	18.1
1997 d	**0.52**	0.50	5.5	**0.44**	0.41	5.4	0.68	0.69	6.1	0.69	0.71	5.8
1998 w	0.77	0.81	2.9	0.72	0.78	0.1	0.77	0.80	2.8	0.84	0.86	1.2
1999 n	0.84	0.86	3.5	0.85	0.90	5.5	0.91	0.94	1.8	0.85	0.89	5.4
2000 w	0.88	0.91	-18.3	0.88	0.91	-16.9	0.89	0.90	-12.5	0.91	0.92	-8.3
2001 d	0.60	0.68	-7.7	**0.47**	0.57	-11.3	0.64	0.72	-8.1	0.71	0.77	-4.5
2002 w	0.93	0.96	6.3	0.95	0.98	2.8	0.74	0.72	12.5	0.77	0.73	9.6
Mean	**0.74**	**0.79**	**7.8**	**0.68**	**0.74**	**7.2**	**0.75**	**0.79**	**7.9**	**0.77**	**0.8**	**7.4**
Val.												
1983 d	**0.37**	0.49	-35.3	**0.13**	0.24	-37.1	0.76	0.87	-14.9	0.76	0.89	-18.6
1984 d	**0.37**	0.52	12.3	**0.25**	0.45	11.7	**0.24**	0.46	12.6	**0.34**	0.45	19.9
1985 d	**0.44**	0.56	7.5	**0.34**	0.50	6.1	**0.24**	0.37	4.9	**0.40**	0.53	14.5
1986 d	0.66	0.69	-12.4	0.65	0.74	-13.4	0.66	0.73	-5.2	**0.47**	0.41	-1.5
1987 w	0.76	0.80	1.3	0.81	0.85	1.5	0.73	0.76	1.7	0.70	0.70	13.9
1988 w	0.83	0.84	-3.2	0.78	0.79	-5.4	0.86	0.88	-4.0	0.82	0.83	12.3
1989 d	0.77	0.82	-5.2	0.63	0.70	-6.4	0.83	0.87	-3.6	0.80	0.83	1.0
1990 d	**0.23**	0.37	-6.7	**0.06**	0.23	-7.6	**0.49**	0.62	-1.7	**0.58**	0.66	-2.2
1991 d	**0.51**	0.57	-3.2	**0.32**	0.41	-6.4	**0.57**	0.66	-1.1	**0.59**	0.68	4.1
1992 d	0.67	0.69	-13.3	0.60	0.61	-12.7	0.76	0.77	-10.4	0.77	0.76	8.3
1993 d	**0.54**	0.61	-9.5	**0.37**	0.46	-9.8	**0.51**	0.65	-11.6	0.63	0.67	-2.3
1994 d	0.80	0.84	-11.1	0.77	0.83	-10.6	0.80	0.83	-11.8	0.83	0.85	-11.6
Mean	**0.58**	**0.65**	**10.1**	**0.48**	**0.57**	**10.7**	**0.62**	**0.70**	**7.0**	**0.64**	**0.69**	**9.2**

w = wet, n = normal, d = dry

Flood event modelling

With regard to the objective of our study it was important to analyze the model performance for flood events. During the period 1983 and 2002 three floods in 1983, 1995 and 2002 occurred in the summer periods with daily discharge values exceeding 500 m^3/s. The hydrographs shown in figure 3 and 4 are not simulation results obtained from the rainfall-runoff database, but from a simulation period starting in 1975. Figure 3 is illustrating the model results for the flood events in 1983 and 1995 with maximum discharge values of about 700 m^3/s. The model versions calibrated to period 1 tend to overestimate runoff peaks and the versions calibrated to period 2 are underestimating peak flow. Both events are represented best by averaging the four simulations, an approach comparable to the multi-model approach proposed for example by Goswami and O'Connor (2007). The simulation results of

the extreme flood event in 2002 are different. Version ANSE_P1 captures the flood peak perfectly and NSE_P1 adequately. Both versions calibrated to period 2 are underestimating the extreme flood peak dramatically, simulating only 50% of peak discharge, see figure 4. This leads us to the assumption that the IHACRES model is not able to capture events of extreme magnitudes if no comparable event occurred in the calibration period or training set, respectively. Thus, from the modellers point of view we are "fortunate" that an extreme event with a recurrence interval of 500 to 1000 years (LfUG, 2002), such as the flood in 2002, occurred and data are available from several streamflow gauges. Hence it was possible to train the model with this information and to gain the capability to better represent comparable events in the future. This is an important prerequisite for the development of a "reliable" rainfall-runoff database for the Mulde catchment, because at least statistically the probability of recurrence of a larger event is relatively low.

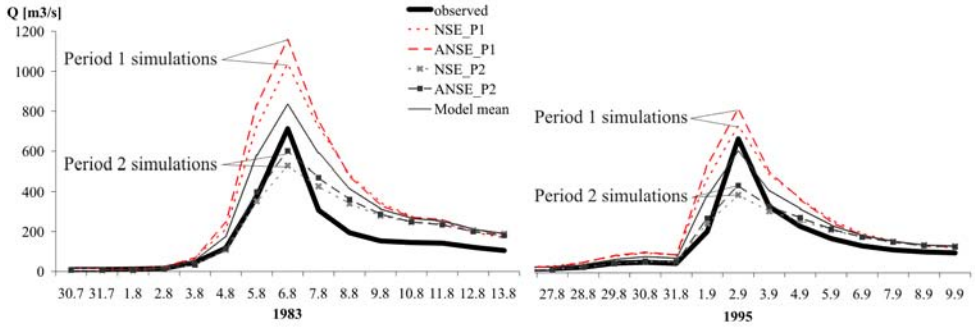


Figure 3. Flood event simulations of the years 1983 and 1995

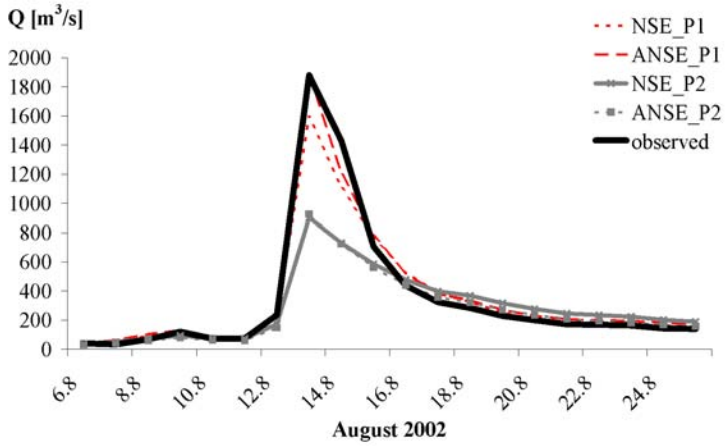


Figure 4. Extreme flood event in the Mulde catchment at gauge Golzern in August 2002

3.3 The rainfall-runoff database

With the development of the rainfall-runoff database our objective was to support applied flood risk management providing a large set of 20-day rainfall-runoff simulations. The number of runoff simulations depends on the number of rainfall scenarios and the number of combinations of initial model states (parameters API and Q). We generated 10,000 rainfall scenarios and 220 API-Q-combinations. Hence the number of runoff scenarios in the database is 2.2 Million. Three categories of tables exist in the database: rainfall and temperature scenario tables, tables with initial model conditions (initial API and Q values), and tables containing the runoff simulations.

Figure 5 shows an example runoff result set, where only runoff simulations, based on rainfall scenarios that are comparable to the 1995 flood, are displayed. Therefore, a database query in the form of:

SELECT *all simulations*
FROM *runoff simulation table*
WHERE *rainfall volume in the initialization period is* > 45 and < 50
AND *rainfall volume in the storm period is* > 90 and < 100
AND *initial streamflow is* < 30 m^3/s
AND *initial antecedent precipitation index is* > 0.1 and < 0.4

was performed.

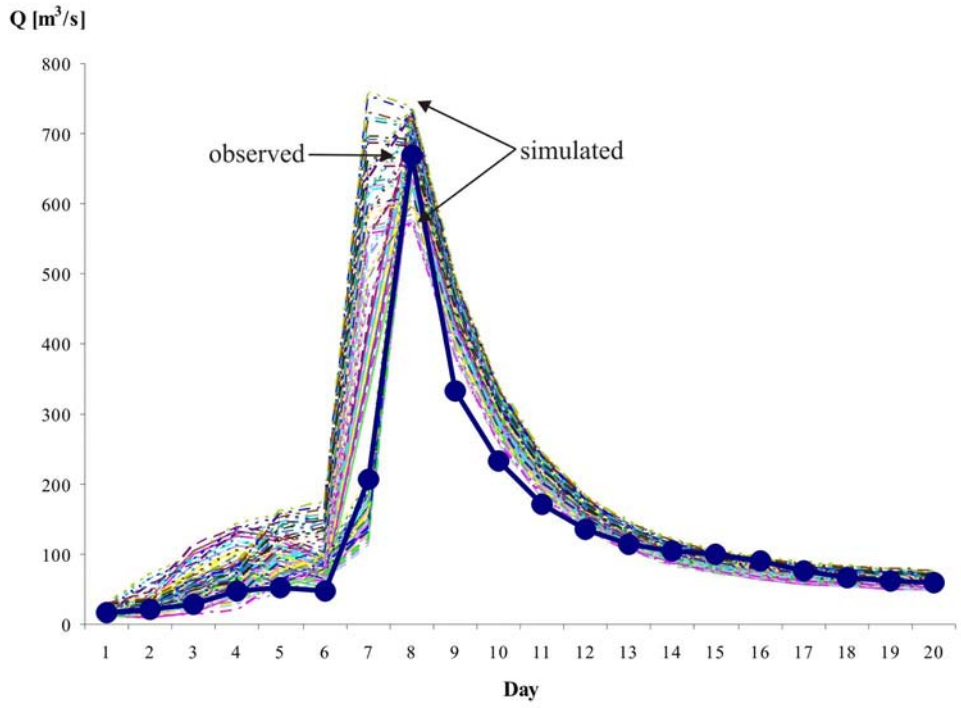


Figure 5. The 1995 flood event captured by the database simulation result set

The example shows that the runoff simulations capture the 1995 flood event, where some simulations are underestimating and some are overestimating the flood peak. The differences between the lowest simulated flood peak and the highest simulated flood peak can be considered as an uncertainty range.

4. CONCLUSIONS

Sources of uncertainties in the flood forecasting process are manifold, starting from uncertainties in precipitation measurements, followed by the uncertainty attributable to the internal states of the hydrological and hydrodynamic models, the initial conditions and relevant process parameterizations (Arduino et al., 2005). With the development of the rainfall-runoff database we are able to show the uncertainties related to precipitation measurements and forecasts as well as uncertainties related to initial conditions. The results of our studies in the Mulde catchment showed that the generated rainfall scenarios are representing a variety of artificial (or not yet occurred) as well as real rainfall events, capturing a large spectrum of different volumes and intensities. Moreover, we demonstrated that flood events can be captured by the range of the result set of runoff simulations.

By using database queries the water manager is able to produce flood risk scenarios, depending on short-term rainfall forecast. The database can be used by persons who are not

familiar with hydrological modelling, because the modelling step has been accomplished already. By linking the simulation results (minimum and maximum peak flow) with a hydrodynamic model a water manager is able to delineate flood inundation areas.

Due to the growing complexity in rainfall-runoff modelling during seasons affected by snowmelt processes, the applicability of the database in the current state is limited in snow affected catchments to the warm season. In order to better account for seasonality in climate scenarios we propose to use more sophisticated weather generators, such as the EARWIG (Kilsby et al., 2007) based on a Neyman-Scott point process model (Cowpertwait, 1991), or stochastic models as the WGEN (Richardson and Wright, 1984) or LARS-WG (Semenov et al., 1998; Semenov and Brooks, 1999), to produce design rainfall and related temperature events.

REFERENCES

Arduino, G., Reggiani, P. and E. Todini, Recent advances in flood forecasting and flood risk assessment, *Hydrology and Earth System Sciences*, 9, 280-284, 2005.

Cowpertwait, P.S.P., Further Developments of the Neyman-Scott Clustered Point Process for Modeling Rainfall, *Water Resources Research*, 27, 1431-1438, 1991.

Croke, B.F.W., Merritt, W.S. and A.J. Jakeman, A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments, *Journal Of Hydrology*, 291, 115-131, 2004.

Gabellani, S.; Boni, G.; Ferraris, L.; von Hardenberg. J. and A. Provenzale, Propagation of uncertainty from rainfall to runoff: A case study with a stochastic rainfall generator, *Advances in Water Resources,* 30, 2061-2071, 2007.

Goswami, M. and K.M. O'Connor, Real-time flow forecasting in the absence of quantitative precipitation forecasts: A multi-model approach, *Journal of Hydrology*, 334, 125-140, 2007.

Hlavcova, H., Kohnova, S., Kubes, R., Szolgay, J. and M. Zvolensky, An empirical method for estimating future flood risks for flood warnings, *Hydrology and Earth System Sciences*, 9, 431-448, 2005.

Hoffmann, L., Idrissi, A. E., Pfister, L., Hingray, B., Guex, F., Musy, A., Humbert, J., Drogue, G. and Leviandier, T., Development of regionalized hydrological models in an area with short hydrological observation series, *River Research and Applications*, 20, 243-254, 2004.

Jakeman, A.J., Littlewood, I.G. and P.G. Whitehead, Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, *Journal of Hydrology*, 117, 275, 1990.

Jakeman, A.J. and G.M. Hornberger, How Much Complexity Is Warranted in a Rainfall-Runoff Model?, *Water Resources Research*, 29, 2637-2649, 1993.

Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., James, P., Smith, A. and R.L. Wilby, A daily weather generator for use in climate change studies, *Environmental Modelling and Software*, 22, 1705-1719, 2007.

Kokkonen, T.S., Jakeman, A.J., Young, P.C. and H.J. Koivusalo, Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, *North Carolina Hydrological Processes*, 17, 2219-2238, 2003.

Kokkonen, T.S. and A.J. Jakeman, Structural Effects of Landscape and Land Use on Streamflow Response, in: *Environmental Foresight and Models*: A Manifesto, 303-321, 2002.

LfUG, Vorläufiger Kurzbericht über die meteorologisch-hydrologische Situation beim Hochwasser im August 2002. Version 5, 2.12.2002. Sächsisches Landesamt für Umwelt und Geologie, 2002.

Nash, J. and Sutcliffe, J., River flow forecasting through conceptual models, Part 1 - a discussion of principles, *Journal Of Hydrology*, 10, 282-290, 1970.

Post, D.A. and A.J. Jakeman, Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model, *Ecological Modelling*, 123, 91, 1999.

Post, D.A., Jones, J.A. and G.E. Grant, An improved methodology for predicting the daily hydrologic response of ungauged catchments, *Environmental Modelling and Software*, 13, 395-403, 1998.

Richardson C.W. and D.A. Wright, WGEN: A model for generating daily weather variables, US Department of Agriculture, Agricultural Research Service, ARS-8, USDA, Washington, DC, 1984.

Sefton, C.E.M. and S.M. Howarth, Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, *Journal of Hydrology*, 211, 1, 1998.

Semenov, M.A., Brooks, R.J., Barrow, E.M. and C.W. Richardson, Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates, *Climate Research*, 10, 95-107, 1998.

Semenov, M.A. and R.J. Brooks, Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain, *Climate Research*, 11, 137-148, 1999.

Smith, K. and R. Ward, Floods: *Physical Processes and Human Impacts*. John Wiley, Chichester, 1998.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Bottom-up methodologies for the modeling and upscaling of farm Nitrogen losses in European landscapes

T. Dalgaard[a], **C. Kjeldsen**[a]

[a] *University of Aarhus. Department of Agroecology and Environment, Faculty of Agricultural Sciences (DJF). P.O. Box 50. DK-8830 Tjele, Denmark*

Abstract: This paper presents bottom-up scaling methodologies, implemented in the MEA-Scope strategic research project, and in the NitroEurope landscape level component. Both projects are based on a bottom-up approach, where farm information are collected for landscapes in Germany, Slovakia, Poland, France, Hungary, Italy, Scotland, The Netherlands and Denmark; and landscape level impacts of the Model for European Agriculture (MEA) are assessed. This paper is about the upscaling from farm to landscape level, and focuses on the modelling of Nitrogen surplus from agriculture as an indicator for water pollution. It is demonstrated in detail how farm information from the Danish landscape is upscaled for such landscape level analyses, using the EU Integrated Area Control System (IACS) and GIS. Subsequently, farm N-surpluses, upscaled for each of the other MEA-scope landscapes, are also presented, and different upscaling pathways are reviewed. Based on the results, advantages in the bottom-up approaches applied are emphasized. It is concluded, that bottom-up methods for upscaling are needed to convey information from research to decision-makers, and that it is important to specifically address the scale issue within the cycle of strategic research, where an iterative interaction between researchers and decision-makers is carried out.

Keywords: Scaling, farm, landscape, Land Parcel Information System (LPIS), Geographical Information System (GIS), bottom-up modeling.

1. INTRODUCTION

Strategic research is characterized by an iterative interaction between researchers and decision-makers [Bierkens *et al.* 2000]; denoted *"The Cycle of Strategic Research"* (Figure 1). The MEA-scope strategic research project (www.MEA-scope.eu), from which results are presented in this paper, is a good example of such interaction. In this project landscape level impacts of the Model for European Agriculture (MEA) are assessed, and within this context, the present paper focus on one of the major challenges within the cycle of strategic research, namely the problem of scaling research results to the scale, where information is needed by decision makers [Dalgaard *et al.* 2003]. This is exemplified with results from the MEA-scope project, as presented in Dalgaard et al. [2008], and with perspectives for the NitroEurope Integrated EU research project (www.NitroEurope.eu).

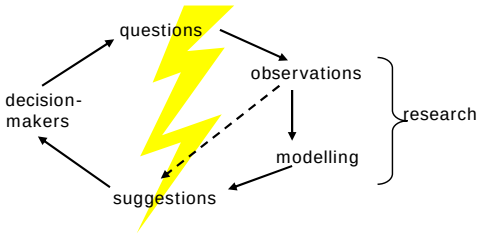


Figure 1. The cycle of strategic research [Bierkens et al. 2000]. The lightening symbolizes the gap between the scale where decision makers operate, and the scale where observations and modeling typically are carried out by researchers.

In MEA-scope, the initial question was formulated by The European Commission, requesting "an integrated framework for the assessment of the multifunctionality impacts of the EU common agricultural and rural development policy reform" (www.MEA-scope.org). The consortium of research institutions behind MEA-scope responded to this question with a project formulation, focussing on an impact assessment of the

agricultural production and its multiple functions in seven landscapes selected. During end-user workshops in Brussels, the research progress was presented. Based on feed-backs from these meetings and from internal project workshops, data collection and modelling approaches were designed [Müller and Piorr 2008].

Since landscapes can be conceived as a conglomerate of different homogenous units [Forman and Godron 1986], of which farming is a very important part, it is vital from the perspective of multifunctionality to represent landscapes in a manner which reflects their multifunctional nature [Brandt and Vejre 2004; Vejre *et al.* 2007].

The upscaling procedure can be said to aim at establishing a representation of landscape functions, building on the farm level simulations, which aims to serve two purposes: adaptation of the knowledge generated in the project to the needs of the potential end users, in addition to pinpointing areas in need of further research. It can be expressed as in the Figure 2, which depicts scaling as a procedure which aims at balancing both research and policy needs.

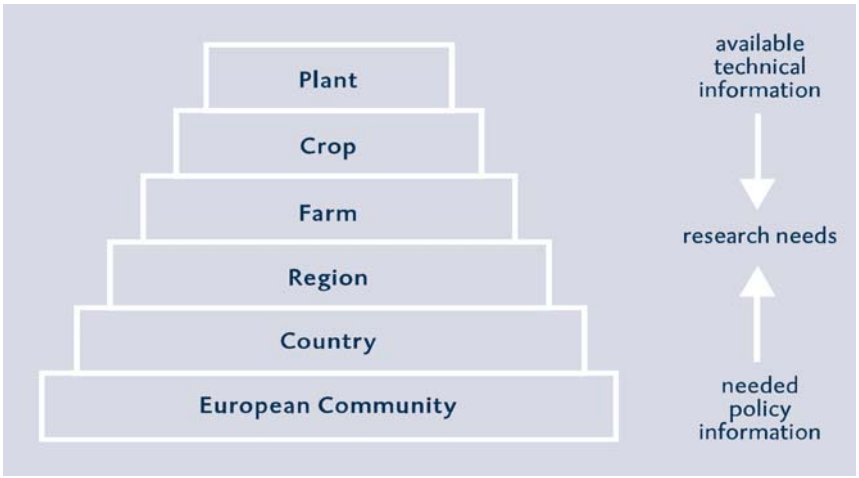


Figure 2. Levels of scale and research needs [van Latesteijn 1998, 1999]

Dalgaard *et al.* [2003] has dealt with the interactive process of setting the appropriate scale relative to the needs of decision-makers. The framework derived from this work is summarised in the table below.

Table 1. General upscaling framework to support and evaluate the conveyance of information between science and decision-makers [Dalgaard et al. 2003]. See the text for further explanation

Criteria 1. Define the decision-maker and the problem and the scale at which the decision-maker needs information.
Criteria 2. Determine on which scales information regarding this problem is available and collect the relevant information.
Criteria 3. Create a hypothesis of how existing information, identified in criteria 2, can be transformed to the scale needed for decision-making, identified in criteria 1. First try with simple linear scaling procedures, and after having tested them in criteria 4, try more complicated, non-linear or hierarchical scaling procedures.
Criteria 4. Test the hypothesis of criteria 3 with independently sampled decision-maker scale information. If the hypothesis is rejected, try with a new hypothesis or seek new information, which can be transformed to the decision-maker scale.

In the MEA-Scope context, application of this upscaling framework could be expressed with the following four criteria analyses: 1. Identify the most relevant landscape functions, required by decision-makers for the specific landscape: for example the present paper focus on nitrogen surplus and the related provision of clean drinking water and non-eutrophicated surface waters. This is especially an important function in the Danish landscape [Schrader et al. 2008]. 2. Review and collect relevant farm and landscape level data. 3. Scaling

from farm to landscape: Mapping indicators via GIS-data, farm-data geo-coding, farm type modelling and regionalisation 4. Test the upscaled results

As the above framework incorporates insights from systems theory [Spedding 1979; Altieri 1995; Gliessman 1998; Checkland 1999], a distinction is made between different types of upscaling: linear, non-linear and hierarchical [Dalgaard et al. 2003]. Linear upscaling refers to a case where upscaling is simply a matter of aggregation of lower level data, whereas non-linear and hierarchical upscaling takes emergent factors into account. A practical example, which is used in the article quoted above, is the influence of increasing field size on farm level fuel use. Another example is the results presented in the present paper, and the non-linear relationship between nitrogen surplus and livestock density (Figure 8). Non-linear upscaling addresses issues within relative fixed boundaries of the system in question, whereas hierarchical scaling can be considered an extended case of non-linear scaling, since it deals with the consequences of extending system boundaries. One practical example is when scale is increased from farm to landscape level. The consequence is that scaling must be approached in a reflective and iterative way, taking into consideration many different levels of organisation and the different temporal and spatial scales that might be of importance for the long-term sustainability of the system as a whole [Dalgaard et al. 2006; Fresco and Kroonenberg 1992].

This paper presents upscaling methodologies, implemented in the MEA-scope strategic research project, and perspectives for the NitroEurope landscape level component. MEA-scope is based on a bottom-up approach, where farm information are collected for landscapes in Germany, Slovakia, Poland, France, Hungary, Italy and Denmark. The NitroEurope landscape level component is based on data collection from landscapes in Scotland, France, Poland, Italy, The Netherlands and Denmark. This paper is about the upscaling from farm to landscape level, and focuses on the modelling of Nitrogen surplus and Nitrogen losses from agriculture as indicators for water pollution. Based on Dalgaard *et al.* [2008], it is demonstrated in detail how farm information from the Danish landscape is upscaled for such landscape level analysis, using the EU Integrated Area Control System (IACS), The Land Parcel Information System (LPIS) obligatory for all member states, and geographical systems analysis (GIS). Subsequently, farm N-surpluses, upscaled for other landscapes, are also presented, and different upscaling pathways are reviewed. Based on the results, advantages in the bottom-up approaches applied are emphasized.

2. MATERIALS AND METHODS

There are two main approaches to derive landscape level farm information for the use in decision-making (Figure 3). The first approach is a top-down approach where information from national or regional farm statistics are disaggregated (downscaled) to the landscape level. Recent examples of such approaches are outlined in Leip et al. [2007]. In MEA-Scope, we use the second approach, where landscape level farm information is derived bottom-up. This means that the landscape level information is aggregated (upscaled) from farm information required locally within the actual landscape (for example from local farm surveys and detailed GIS land use maps in combination with maps over the placements of specific farms within a landscape, see below).

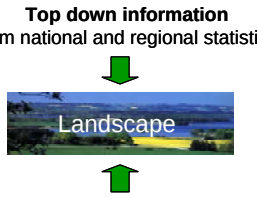


Figure 3. Sources for top-down versus bottom-up derivation of farm information for landscape level studies

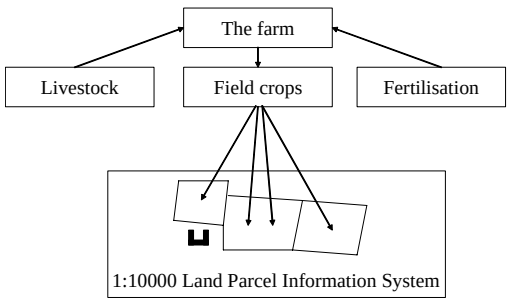


Figure 4. Illustration of the types of digital farm data registered in EU member states for the control of farm subsidies paid. Via the obligatory Land Parcel Information System (LPIS) these data can be geo-referenced and mapped in GIS. It is mandatory to include livestock and field crop registrations, while fertilisation practices are only registered in some member states

In MEA-Scope we apply two different levels of bottom-up landscape farm mapping: 1) The first level includes "real farm maps" required from the mandatory EU digital farm registers, and 2) the second level relies on "proxy farm maps", derived from local farm surveys and GIS-information.

While Ungaro et al. [2008] and Damgaard et al. [2008a] focus on the second level of farm mapping, the present paper uses the first level "real farm maps" to exemplify points regarding scaling. The "real farm maps" are derived from the mandatory EU digital farm registers. According to The European Commission [1992] all EU member states are required to set-up an Integrated Area and Control System (IACS), where subsidy payments are digitally registered. Moreover, a GIS-based Land Parcel Information System (LPIS), to which the subsidy payments can be geographically related, must be established (Figure 4).

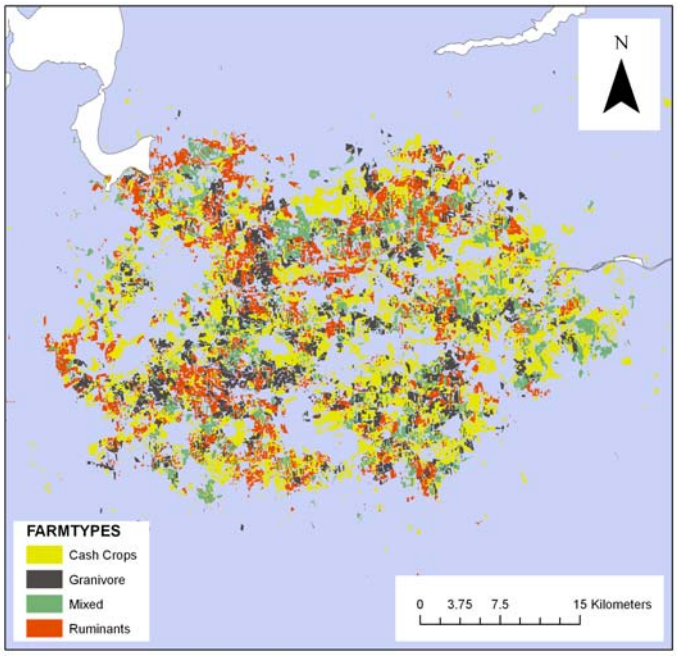


Figure 5. "Real farm map" for the Danish study landscape in year 2002. Each of the in total 1.871 farm has been classified into four main types, according to the EUROSTAT/FADN methods (McClintock 1989; Dalgaard et al. 2002b).

In MEA-Scope, IACS and LPIS data have been available for the study landscapes in Denmark and Slovakia. From this information maps showing the areas belonging to each farm within each land parcel can be constructed. Figure 5 shows an example of such map, where each of the 1.871 farms in the Danish study landscape in year 2002 has been classified into four main types, according to the European Farm Accountancy Data Network, FADN and EUROSTAT methods [McClintock 1989, Dalgaard et al. 2002b].

There are two different pathways for the upscaling of bottom-up farm level information for landscape level modelling [Marshall et al. 1998; Kjeldsen et al. 2006]. In the first pathway, modelling is initially carried out



Figure 6. The geographical location of the seven MEA-Scope study landscapes.

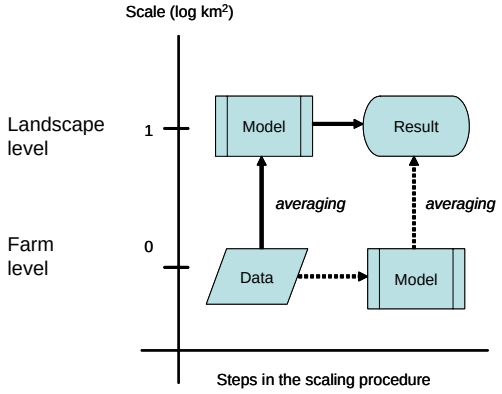


Figure 7. The two pathways for the upscaling of farm information and model results from farm to landscape level: 1) modelling before aggregation, and 2) vice versa (based on Kjeldsen et al. 2006).

on the single farm data, before aggregating the model results to the landscape level, while in the second pathway the farm information is averaged before modelling (Figure 7).

In the present paper it is demonstrated how these two pathways can lead to significantly different results. This is demonstrated using the Farm-N model (www.Farm-N.dk, www.Farm-N.dk/farmNtool) to simulate farm nitrogen (N) surpluses for the Danish study landscape in 2002. With this model, the farm N-surplus is calculated in kg N/ha/yr as N-inputs (mineral fertilizer, manure, feed, straw, seeds and animals bought + N fixed and N deposited from the atmosphere) minus N-outputs (cash-crops, animal prod-ucts, milk, manure, and feed and straw sold). For more details see Dalgaard et al. [2007b].

In total, the MEA-Scope model framework is applied to seven European landscapes (Figure 6). The project focussed on beef farming, and the landscapes were selected to include a significant part of the land with grassing livestock. The hypothesis was that these systems carried a special potential to contribute to a more sustainable development. Additional key figures for agriculture and land use in these landscapes can be found in Dalgaard et al. [2007a] and at http://mea-scope.eu/.

3. RESULTS

Based on the bottom-up farm information available, the nitrogen surpluses are modelled for all the 1.871 farms in the Danish study area, and plotted against the livestock density (Figure 8).

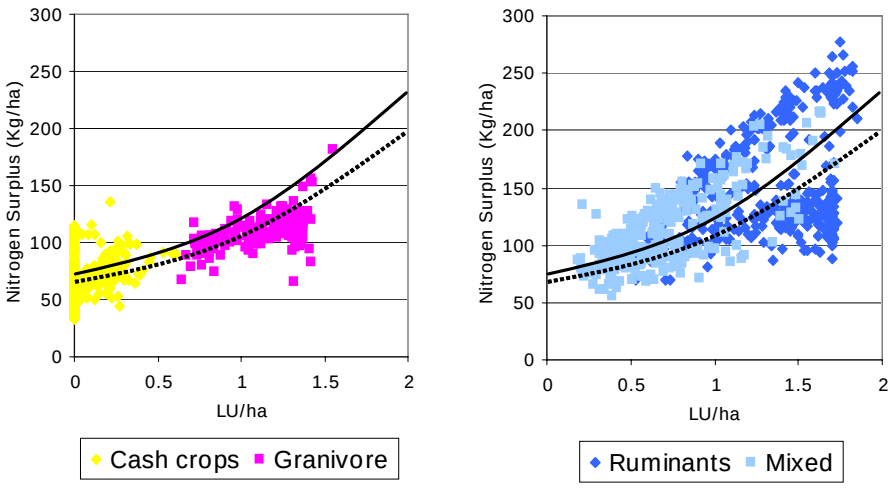


Figure 8. Nitrogen surplus estimated with the Farm-N model for each of the 1.871 farms within the Danish study landscape in year 2002, and plotted against the livestock density in livestock units (LU) per ha. The farms are divided into the four main farm types of Figure 5. The dotted lines show the regression for all farms modelled ($y= 64\ e^{[0.57x]}$), compared to the non-dotted $y= 77\ e^{[0.56x]}$ regression from an independent empirical study of 41 farms within the study area (Dalgaard et al. 2002a).

In correspondence with the statistically significant relationship, derived empirically by Dalgaard et al. [2002a], the model results show a non-linear relationship between nitrogen surplus and livestock density (Figure 8). The model results show lower estimated N-surpluses ($y= 64\ e^{[0.57x]}$) than the empirical study ($y= 77\ e^{[0.56x]}$), a result which could be expected because of increased manure N utilisation rates over time. I.e. the model calculations, which are for the year 2002, assume higher rates for the utilisation of N in livestock manures than those generally anticipated on farms from the empirical study from year 1994-1998. Moreover, the empirical study included farms with higher livestock densities (up to 2.9 LU/ha) than the present study from 2002, and as a consequence of the non-linear relationship, a relatively higher N-surplus for these farms can be anticipated.

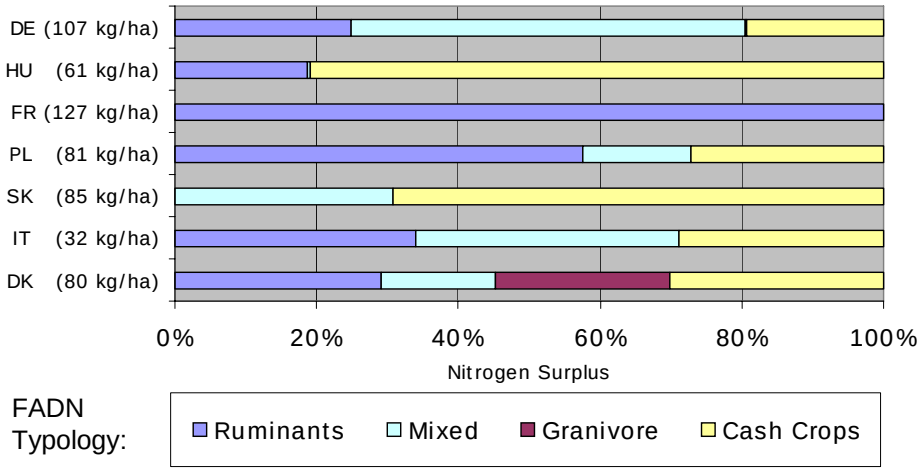


Figure 9. Upscaled farm nitrogen surpluses from the seven MEA-Scope study landscapes in Germany (DE), Hungary (HU), France (FR), Poland (PL), Slovakia (SK), Italy (IT) and Denmark (DK). The N surpluses for all farms in each of the landscapes are modelled with the farm-N model, and the results are upscaled to the landscape level using the pathway 1 approach of figure 7. Finally, the results are summarised in the four main EUROSTAT farm type classes of figure 5, and for each study landscape the average kg N-surplus per ha is specified in brackets.

In summary, the non-linear relationship indicates that the bottom-up information approach applied in the present study results in another N-surplus result than if a top-down approach was applied. If a top-down approach had been applied, the farm N-surpluses would namely not have been modelled for each farm separately (pathway 1 in Figure 7), but for averaged groups of farms (pathway 2 in Figure 7). For example according to the non-linear relationship in Figure 8, farms with 0, 1 and 2 LU/ha would typically yield around 64 $e^{[0.57x0]}$)= 64, 64 $e^{[0.57x1]}$)= 113 and 64 $e^{[0.57x2]}$)= 200 kg N-surplus per ha per year, respectively; according to the regression line (Figure 8). However, the average of (64+200)/2= 132 is not equal to 113, and a top-down approach following the pathway 2 scaling procedure would in this case typically overestimate the total N-surplus from a group of farms.

Another important advantage in using a bottom-up approach based on single farm data is the possibility for detailed mapping and geographical analysis. Figure 10 shows an example of such mapping based on the farm N-surpluses of Figure 8 and the "real farm map" of Figure 5. With such map it is possible to identify nitrogen surplus hot-spots in the landscape, and to make overlay analysis with maps over Natura 2000 sites, groundwater protection areas etc. Finally, explicit mapping of the farming structure enables advanced analyses of the relations between farm structural development and environmental

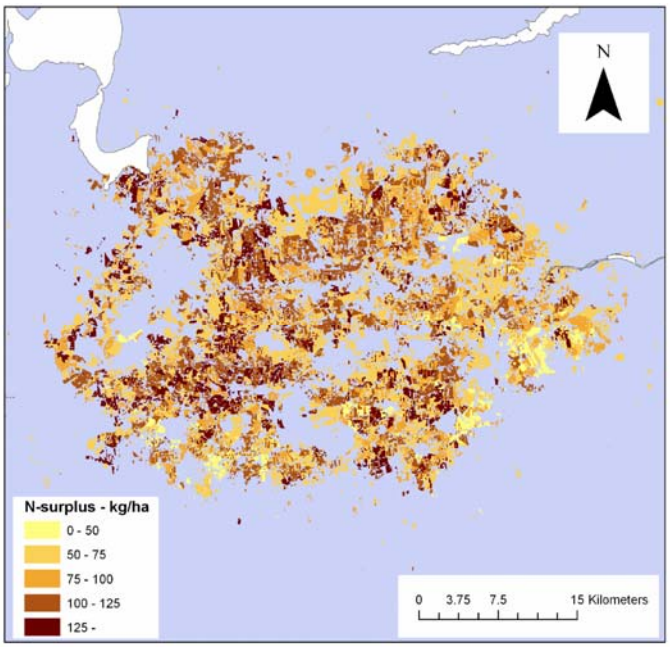


Figure 10. Farm Nitrogen surplus map over the 1.871 farms in the Danish study landscape in year 2002. The mapped N-surplus values are equal to those presented in the scatter plot of figure 8.

effects of agriculture in the form of nitrogen pollution. However, this is out of the scope with the present paper, and must be left for future studies.

Based on the bottom-up farm information acquired in the seven MEA-Scope study landscapes of Figure 6, the N surpluses for all farms in each of the landscapes are modelled with the farm-N model, and the results are upscaled to the landscape level using the pathway 1 approach of Figure 7. Figure 9 shows the summarised nitrogen surplus results from the seven MEA-Scope landscapes in year 2002, distributed on the four EUROSTAT main farm types of Figure 5: ruminants (mainly cattle), granivores (mainly pigs), mixed farms and cash crop farms. In summary, the N-surplus was highest on farms with ruminant and granivores (at average 108 kg N/ha and 121 kg N/ha, respectively), and lowest on cash-crop and mixed farms (at average 74 kg N/ha and 61 kg N/ha, respectively), but with a large variation both within and between landscapes (for example farms with ruminants in the French landscape showed an average N-surplus of 127 kg N/ha, compared to an N-surplus of 85 kg N/ha in the Polish landscape). In Figure 9 the differences between landscapes are indicated with the average kg N-surplus per ha, specified in brackets for each study landscape. In line with the results of Figure 8, the landscapes in France, Poland, Germany and Denmark with a relatively high share of ruminant farms, have a high average N-surplus. With exception of the Italian landscape, where the livestock density of this category is quite low, this is because these farms generally have a high livestock density and consequently also a high Nitrogen surplus. Moreover, in line with the results presented in Figure 8, the variation in N-surplus was highest on ruminant farms, including a significant proportion of grassing livestock. Thereby, because of the large variation in N-surpluses, the potential for reducing the N-surplus also seems to be the highest within this farm category. This corresponds well to the hypothesis of the MEA-scope project, that these systems have a special potential to contribute to a more sustainable development; and in this case a reduction in the N-surplus from agriculture.

5. CONCLUSIONS

The results presented in this paper illustrate clear advantages in using the bottom-up approaches applied in MEA-Scope and Nitro-Europe, compared to the more usual top-down approaches. Moreover, advantages in upscaling of model results from farm to landscape level, using the first pathway of Figure 7 are illustrated, and dangers of the second pathway are emphasized.

In reality geo-referenced bottom-up farm information is often not readily available at the landscape level all over Europe. This makes the creation of detailed maps like Figure 5 and Figure 9 difficult, and a combination of the first and the second upscaling pathway of Figure 7 must be applied. Actually, such aggregation is also applied in MEA-Scope, where bottom-up farm information for “real farm mapping” has only been available for Slovakia and Denmark, and where farm group information has been applied in some part of the modelling instead of farm specific information. It is important to be aware of the potential problems of applying such compromises. It is our hope that the present paper can draw attention to some of the key problems in scaling from farm to landscape, and help to enlighten some of the errors that might appear when doing the needed compromises in the scaling procedures. In Denmark, “real farm maps” are now being used nationally to produce detailed maps over the risk for N-losses from farming, and in the years to come these maps will be used to evaluate the goals setup according the EU Water Framework Directive (Kronvang et al., 2008). Also, the Farm-N model, is now is used nationally to account N-losses from farms which increase their livestock production significantly, and it will be interesting to follow to which degree the issues discussed in this paper will be integrated in the practical procedures for the landscape scale evaluation of farm Nitrogen losses.

To finally conclude on the nature of upscaling, it can be defined as an iterative process, where the actual outcome cannot be seen in isolation from the policy needs formulated by the end users. Thus, end user workshops like the ones carried out in MEA-Scope, and the scaling of information within the Figure 1 cycle of strategic research produce a very important input to this process.

ACKNOWLEDGEMENTS

This work was mainly carried out as part of the EU funded 6th framework project MEA-Scope (Micro-economic instruments for impact assessment of Multifunctional agriculture to implement the Model of European Agriculture, SSPE-CT-2004- 501516), and partly in the Integrated project NitroEurope (The Nitrogen Cycle and its influence on the European greenhouse gas balance, SUSTDEV-2004-3.I.1.1). The authors wish to thank all the consortia members who contributed to the discussion that helped shape this

work. Especially we thank you the co-authors of the paper by Dalgaard *et al.* [2008], from which the contents of this paper originate.

REFERENCES

Altieri, M.A., Agroecology: The science of sustainable agriculture. Westview Press, Boulder, 1995.

Bierkens, M.F.P., P.A. Finke, P. de Willigen, Upscaling and downscaling Methods for environmental research. Developments in Plant and Soil Sciences 88. Kluwer Academic Publishers, Dordrecht, 2000.

Brandt, J., H. Vejre, Multifunctional landscapes – motives, concepts and perspectives. In: Brandt J, Vejre H (eds) Multifunctional Landscapes Volume I: Theory, values and history. WIT Press Southampton, Boston, pp 3-31, 2004.

Checkland, P., Systems thinking, systems practice. A 30-year retrospective. Viley, Chichester, 1999.

Dalgaard, T., C. Kjeldsen, M.S. Jørgensen, N. Hutchings, L. Mogensen, A. Osuch, M. Damgaard, K. Happe, A. Piorr, Scaling from farm to landscape. In: Piorr, A. and K. Müller (eds.) Rural Landscapes and Agricultural Policies in Europe. Springer Verlag, Berlin, 2008, in press.

Dalgaard, T., C. Kjeldsen, N. Hutchings, K. Happe, A. Osuch, M. Damgaard, P. Zander, A. Piorr, Multifunctional farming, multifunctional landscapes and rural development. In: Mander Ü, H. Wiggering, K. Helming (Eds) Multifunctional Land Use: Meeting future demands for landscape goods and services. Springer, Berlin, pp 168-18, 2007a.

Dalgaard, T., C. Kjeldsen, M. Damgaard, N. Hutchings, B.V. Iversen, T. Heidmann, M.S. Jørgensen, et al., Report on apllicability of farm and landscape level MEA-Scope. Deliverable report D5.5. MEA-Scope EU Strategic Research Project SSPE-CT-2004-501516. Cited January 2008 from www.MEA-scope.org. Aarhus University, Faculty of Agricultural Sciences, 2007b.

Dalgaard, T., S. Ferrari, M. Rambonilaza, Introduction: Features of environmental sustainability in agriculture: some conceptual and operational issues. Int. J. Agricultural Resources, Governance and Ecology 5(2/3): 107–115, 2006.

Dalgaard, T., N.J. Hutchings, J.R. Porter, Agroecology, scaling and interdisciplinarity. Agriculture Ecosystems and Environment 100: 39-51, 2003.

Dalgaard, T., T. Heidman, L Mogensen, Potential N-losses in three scenarios for conversion to organic farming in a local area of Denmark. European Journal of Agronomy 16: 207-21, 2002a.

Dalgaard, T., H. Rygnestad, J.D. Jensen, P.E. Larsen, Methods to map and simulate agricultural activity at the landscape scale. Danish Journal of Geography 3: 29-39, 2002b.

Damgaard, M., C. Kjeldsen, F. Ungaro, Recreating context in spatial modelling of agricultural landscapes. In: Piorr, A., K. Müller (eds), Rural Landscapes and Agricultural Policies in Europe. Springer Verlag, Berlin, 2008a.

Damgaard, M., C. Kjeldsen, A. Osuch, K. Happe, T. Dalgaard, Validation of an agent-based, spatio-temporal model for farming in the River Gudenå landscape. In: Piorr, A. and K. Müller (eds), Rural Landscapes and Agricultural Policies in Europe. Springer Verlag, Berlin, 2008b.

European Comission, Council regulation (EEC) No. 3508/92 of November 27, 1992. Establishing an Integrated Area Control System for certainty Community aid schemes. Official Journal L 355 (05.12.1992): 1-5, 1992.

Forman, R.T.T., M. Godron, Landscape Ecology. Wiley, New York, 1986.

Fresco, L.O., S.B. Kroonenberg, Time and spatial scales in ecological sustainability. Land Use Policy 9(3): 155-168, 1992.

Gliessman, S.R., Agroecology. Ecological processes in sustainable agriculture. Ann Arbor Press, Chelsea, 1998.

Kampsax, Report from the pilot project DMK-block (rapport om pilot-projekt DMK Blok). Danish Ministry for Agriculture and Fisheries, Copenhagen, 1994.

Kjeldsen, C., T. Dalgaard, P.K. Bøcher, Methodological issues of modelling farm and landscape indicators for sustainable land systems. Danish Journal of Geography 106(2): 35-43, 2006.

Kronvang, B., H.E. Andersen, C.D. Børgesen, T. Dalgaard, S.E. Larsen, J. Bøgestrand, G. Blicher-Mathiasen, Effects of policy measures implemented in Denmark on nitrogen pollution of the aquatic environment. Environmental Science & Policy, Volume 11, Issue 2, April 2008, Pages 144-152, 2008.

Leip, A., G. Marchi, R. Koeble, M. Kempen, W. Britz, C. Li, Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen losses from cropland soil in Europe. Biogeosciences Discuss., 4, 2215–2278, 2007. www.biogeosciences-discuss.net/4/2215/2007/.

Marshall, B., J.W. Crawford, J.R. Porter, Variability and scaling: matching methods and phenomena. In: Gardingen, P.R., Foody, P.M. and Curran, P.J. (eds) Scaling-up from cell to landscape. Seminar Series 63. Cambridge University Press, Cambridge, pp 253-272, 1998.

McClintock, J., Farm Accountancy Data network. An A to Z Methodology. Report, 1st edition. Comission of The European Communities, Brussels, 1989.

Müller, K., A. Piorr, www.MEA-scope.org. Cited April 2008.

Schrader, C., H. Moschritz, C, Kjeldsen, J. Wasilewski, M. Stolze, Societal demand for commodity and non-commodity outputs – a region perspective. In: Piorr, A., K. Müller (eds), Rural Landscapes and Agricultural Policies in Europe. Springer Verlag, Berlin, 2008.

Spedding, C.R.W., An introduction to agricultural systems. Applied Science Publishers, Essex, 1979.

Ungaro, F., A. Ciancaglini, C. Sattler, S. Uthes, M. Damgaard, K. Happe, A. Osuch, Spatial characteristics of land use patterns in Mugello (Central Italy) and policy impacts on their environmental outputs. In: Piorr, A. and Müller, K. (eds) Rural Landscapes and Agricultural Policies in Europe. Springer Verlag, Berlin, 2008.

van Latesteijn, H.C., A policy perspective on land use changes and risk assessment. Agriculture, Ecosystems & Environment 67(2-3): 289-297, 1998.

van Latesteijn, H.C., Land use in Europe - a methodology for policy-oriented future studies. WRR - Scientific Council for Government Policy. Preliminary and Background Studies V106. Sdu Uitgevers, The Hague, 1999.

Vejre, H., J. Abildtrup, E. Andersen, P.S. Andersen, J. Brandt, A. Busck, T. Dalgaard, B. Hasler, H. Huusom, L.S. Kristensen, S.P. Kristensen, S. Præstholm, Multifunctional agriculture and multifunctional landscapes - land use as interface. In: Mander.,Ü, H. Wiggering, K. Helming (Eds.), Multifunctional Land Use: Meeting future demands for landscape goods and services, Springer, Heidelberg, Berlin,, p. 93-104, 2007.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Simulation of the flow in the Koiliaris River basin (Greece) using a combination of GIS, the HSPF model and a Karstic– Snow melt model

Nektarios N. Kourgialas, George P. Karatzas, Nikolaos N. Nikolaidis

Technical University of Crete – Department of Environmental Engineering, Greece
(nektarios.kourgialas@enveng.tuc.gr, karatzas@mred.tuc.gr)

Abstract: The purpose of this study was to model the Koiliaris River flow where flood phenomena appear from time to time. The Koiliaris River basin is located east of the City of Chania, Crete, (Greece). The basin is extended from the White Mountains (Lefka Ori) to the coastline. The geology of the basin is mainly constituted by carbonate (Karstic area), quaternary-neogenic deposits and flysch formation. The main volume of water is discharged from the karstic system of the White Mountains (Lefka Ori) through springs and temporary rivers. In order to calculate the flow from the springs the Maillet Karstic model is used. The main volume of water that is discharged to the springs through the karstic system comes from snow melt. In order to determine with high precision the rate of snow melt, a Snow melt model and the applications of GIS are combined. Using this approach the discharge from springs (for the time of the simulation) is determined. The time-series of the karstic discharge from the springs is entered in the code of the Hydrological Simulation Program - FORTRAN (HSPF) model, in order to calculate both the surface and subsurface discharge. The HSPF is a set of computer codes that simulate the hydrologic process. The hydrologic model of HSPF functions in the frame of the BASINS 4 model. The final step is the Calibration and the Sensitivity analysis of the HSPF model. The main objective of the present study is to develop a tool for the prediction and management of flood events that occur in the area.

Keywords: GIS; HSPF model; Karstic area; Koiliaris River basin; Snow melt.

1. INTRODUCTION

One of the main disantvantages of the models estimating the river flow is the fact that very often the contribution of the groundwater flow is ignored. The contribution of this kind of flow can be significant especially in cases where the geological formation is characterized as karstic. In the karstic areas most of the surface runoff and the snowmelt water quantity becames groundwater flow which can be realized farther down the flow direction in the form of springs. Also, this groundwater flow in many cases becames a significant contribution to the main river flow. In the present study, the HSPF model in combination with a snow melt model was employed to compute the surface and groundwater flow contributions to the Koiliaris River flow (Jaquet et al., 2004).

2. KOILIARIS RIVER BASIN

2.1 General Characteristics

The Koiliaris River basin is located 15 Km east of the City of Chania, Crete and is extended from the White Mountains (Lefka Ori) to the coastline. The area of the basin has been estimated to be approximately 130 km^2. The elevations of the basin range from zero to 2120 m ASL (Figure 1). The total length of the Koiliaris River network is 36 km. From the

intersection point, where all the streams met, to the outflow point the length of the river is 3.3 km. In 2005, a hydrometric station was installed at the location Agios Georgios in order to monitor the flood events. The rainfall data of the study area have been collected from the raingauge station of Kalibes, which is close to the river and located at an elevation of 24 m ASL (Figure 2).

2.2 Hydro Geological Characteristics

The topography of the area of interest is smooth with a mild topographic slope of 12%. The geology of the basin is mainly constituted by carbonate, quaternary-neogenic deposits and flysch formation. Based on a former research project, conducted by the Ministry of Agriculture of Greece, it has been estimated that 58% of the total land use of the basin is characterized as pasture (public or private), 29.4% as crops, 2.8% settlements and roads, 8.5% as forests, 0.6% as water surfaces and 0.7% as other uses.

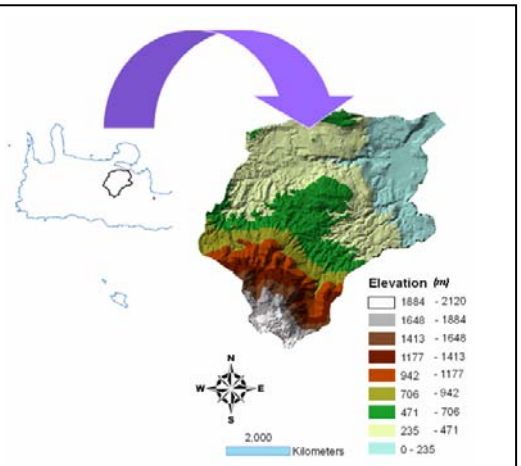


Figure 1. The island of Crete and the Koiliaris River basin.



Figure 2. The hydrometeorological network of the Koiliaris River

2.3 Temperature – Rainfall gradient of the Koiliaris River basin

Until 2007, in the Koiliaris River basin there were not a sufficient number of meteorological stations. Therefore, in the present study the meteorological information was obtained from stations located in the surroundings of the basin area. Since the Koiliaris River basin is small it is considered that data from the surrounding the basin stations can be used for the present study. These stations are even distributed regarding altitude, land-planing and have continuous measurements of the meteorological parameters for a long time period. The nearest meteorological stations used in the present analysis were the meteorological stations of Souda, Kalibes, Askifou, Brises and Chania.

Through the analysis of the meteorological data from the above stations the monthly temperature change and the annual rainfall change were estimate, for the North-East part of the Prefecture of Chania, where the Koiliaris River basin is located. The meteorological data were for the time period between the years 1975 and 2006, (Table 1, Figure 3).

Table 1. Average monthly temperature change with the altitude in the Koiliaris River basin.

Monthly temperature change per 100m altitude											
Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
0.82	0.75	0.74	0.56	0.25	0.26	0.17	0.72	0.62	0.68	0.93	0,7

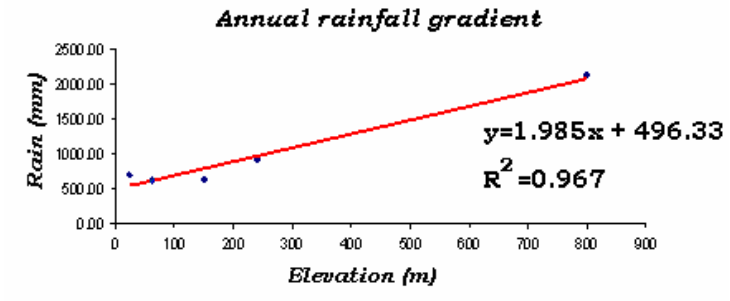


Figure 3. Annual rainfall gradient according to the altitude, for the North-East part of the City of Chania.

3. THE KARSTIC AND SNOW MELT RATE MODEL

In the present study the conceptual model considers the following:
The main volume of water in the area of study is discharged from the karstic system of the White Mountains (Lefka Ori) through springs, streams and temporary rivers. In the Koiliaris River basin three main karstic discharge points exist: the Stilos, the Armenoi and the Zourbos springs. The Stilos spring is one of the main water sources and the major discharge point of the karst system of White Mountains and it is considered to be the main water source of the Koiliaris River.

The Kartic model used was the stochastic Maillet karstic model. The reason for choosing Maillet's model is that it has a physical meaning; it is based on the hydrologic balance equation during the recession period and takes into consideration springs discharges at specific points (the case of Stylos spring). The Karstic system of the Koiliaris River was considered as a two reservoir system (Stamati et al., 2006). The analysis of the above system suggests the existence of "two reservoirs" within the karstic system; the upper reservoir with a faster response (wide conduits), and the lower reservoir with a slower response (narrow fractures). A 'two part' Maillet model was developed representing the upper and lower reservoirs (Figure 4).

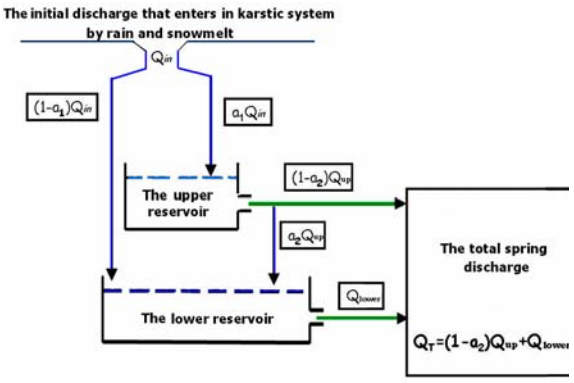


Figure 4. Graphic description of the Karstic model.

The formulation of the karstic model of Maillet is:

$$Q_T = (1-\alpha_2) * Q_{up} + Q_{lower}$$

$$Q_{up} = Q_{up-1} e^{-k_u \Delta t} + a_1 Q_{IN} (1 - e^{-k_u \Delta t}) \quad (1)$$

$$Q_{lower} = Q_{lower-1} e^{-k_i \Delta t} + [a_2 Q_{up} + (1-a_1) Q_{IN}](1 - e^{-k_i \Delta t})$$

Where:

Q_T is the total spring discharge, Q_{up} and Q_{lower} are the initial discharge contributions of the upper and lower reservoirs respectively, Q_{IN} is the initial discharge that enters in karstic system, α_1 is the percent of the water (rain and snow melt) that enters in the upper reservoir and α_2 is the percent of discharge that enters from the upper to the lower reservoir, k_u and k_i (1/d) are the recession coefficients of the upper and lower reservoirs respectively. Moreover, the time step (Δt) of the model is equal to 1 day, (Tzoraki et al., 2006).

The parameter Q_{IN} includes apart from the rainfall the daily rate of snow melt. The daily rate of snow melt is calculated by empirical equations. These equations are developed by the U.S. Army Corps of Engineers (1956) and have been applied in many flow basin models, (Nikolaidis et al., 1988):
A) Snow melt equation during the dry period:

$$M_s = C_{m1} * (1.8 * T_a)^{n+1} / 1000 \quad (2)$$

B) Snow melt equation during the rain period:

$$M_s = C_{m2} * (0.007 * P_w + 0.074) * T_a * 1.8 + 0.05) * 0.254 / 1000 \quad (3)$$

Where:
Ms is the snow melt rate (m/d), C_{m1} is the snow melt factor ($md^{-1}C^{-1}$), C_{m2} is the snow melt factor adjusted to local conditions (dimensionless), n is the coefficient with the value of 0.25, P_w is the daily rate of rainfall in (m/d), and T_a is the average daily air temperature in (^oC).

3.1 The Karstic area and snow melt zones

Stilos spring discharges at 17 m ASL and have an intense seasonal fluctuation in its discharge. The average annual discharge is $73*10^6$ m^3/yr (hydrologic years 1970-2006). In addition to this the average annual discharge of Armenoi and Zourbos springs were taken into consideration which are $27*10^6$ m^3/yr and $26*10^6$ m^3/yr respectively, (Figure 5). The above values of discharges were considered as an input to the Maillet karstic model and by using the trial and error technic it was possible to determine the extent karstic area that contributes to the spring discharges in the area of Koiliaris River basin. Based on this approach and after the calibration of the Maillet karstic model, the extent karstic area that contributes to the above water volume was estimated to be 138 km^2. Due to the lack of previous data a geological study was performed in order to determine the karstic area. Based on this study it was found that in the area of interest two main karstic faults exist. These faults are connected as it is shown in Figure 5. Accordingly to this, the estimated extended karstic area that contributes to the discharge of the above springs appears in Figure 5.

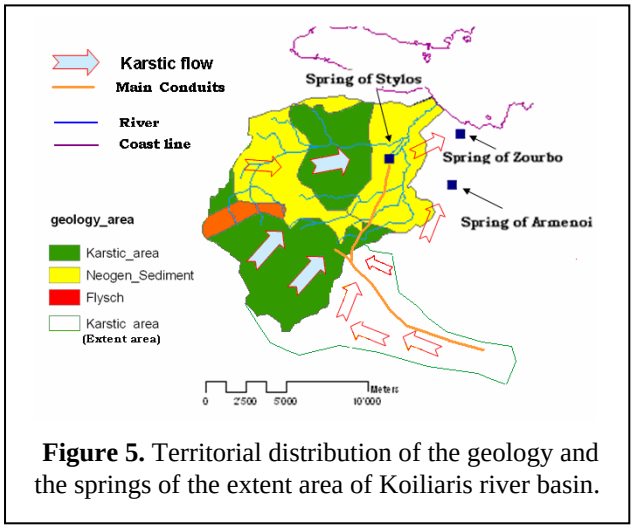


Figure 5. Territorial distribution of the geology and the springs of the extent area of Koiliaris river basin.

The karstic model of Maillet incorporates the equations of the snow melt rate. These equations take into account the quantity of rainfall, and the air temperature, (Yang et al., 1997) but they do not consider the topography of the area and the intensity the solar

radiation. In order to take these two terms into consideration GIS was used. The territorial mountainous area was divided into 5 different zones that were characterized by 5 different snow melt rates (very low, low, medium, high, and very high), (Shaban et al., 2006). The final map of territorial snow melt rate was created based on the topographic characteristics and the intensity of solar radiation

The areas and the average altitudes of the 5 different snow melt zones were computed with the use of GIS. Using the Weighted Moving Average of each area with their corresponding altitudes, the average altitude of the total karstic area was determined to be 1400.41 m. The computed daily rainfall and temperature data were imported to the karstic model. Based on this, the discharge from Stylos spring was computed and used as input to HSPF. The HSPF model has the ability to simulate the surface flow from the basin to the river. The final objective was a more accurate simulation of the Koiliaris River basin flow (surface and groundwater).

4. THE HSPF MODEL

The Hydrological Simulation Program - FORTRAN (HSPF) is a set of computer codes that simulate the hydrological and associated water quality processes on pervious and impervious land surfaces, in streams and well-mixed impoundments. The model simulates the time response of the watershed based on the hydrological and geochemical mass balance and the wet and dry depositions to surface water. The HSPF model is a physically based model that incorporates GIS data (Bicknell et al., 2001).

4.1 The Hydrology of the HSPF model

The HSPF model considers that in the river a complete mixture takes place and that the flow is uniform. It takes into consideration the equation of continuity:

$$VOLE = VOLS + sumIVOL - sumOVOL + PR\text{-}EVAP \quad (4)$$

Where:
VOLE= volume of water at the end of the time step, VOLS= volume of water at the beginning of the time step, IVOL= inflow volume, OVOL= outflow volume, PR= rainfall volume, EVAP= evaporation volume. The OVOL is computed as:

$$OVOL = \Delta t\,(KS*OS + (1-KS)*OE) \quad (5)$$

Where:
KS= gravity coefficient with values between 0.0 – 0.5, OS= outflow at the beginning of the time step and OE= out flow at the end of the time step

Substituting Eq. [4] into Eq. [5] yields:

$$VOLE = VOLS + sumIVOL + PR\text{-}EVAP - [\Delta t\,(KS*OS + (1-KS)*OE)] \quad (6)$$

In Equation [6] the only unknown parameter is the OE, which can be calculated using the FTABLE table that connects the geometry of watercourse of the sector with the flow (Bicknell et al., 2001).

4.2 An application of the hydrologic model HSPF to the River basin for the time period 09/01/2006 to 01/05/2007

The HSPF model operates in the frame of the BASINS model. The new edition (2007) of the BASINS 4 model was used. The BASINS 4 model has the potential to divide the basin

into sub-basins, so that the hydrology of the entire basin to be simulated with more accurate precision. The essential geographic information required by BASINS 4 model was the watershed, the sub-basins, the hydro network, the uses of land and the elevations. This information was supplied to the model using GIS. In the present work the basin of the Koiliaris River was divided in 6 sub-basins, (SWS), (Figure 6).



Figure 6. Segregation of the basin into 6 sub-basins in the environment of the BASINS 4.

The main criteria of selecting these 6 sub-basins were the uniform hydrological, topographical, geological and land-planning characteristics. The above sub-basin information becomes the input for the BASINS-4 model which creates data files in the form that HSPF model requires. Also, by using the Watershed Data Management (WDM) program all the meteorological time series are logged on and used for running HSPF. For each sub-basin different time series were used regarding the temperature and the rainfall. These time series are related to the annual rainfall gradient and monthly temperature gradient for the region of interest. Specifically, the average altitude of each sub-basin was calculated and then the daily rainfall and temperature were determined using the hypsometric method. The evaporation was calculated by the Penman method using the WDM program that considers the data of the daily rainfall, relative sunlight, relative humidity, and temperature.

Regarding the solar radiation the only available data were from the meteorological station of Chania, which started to operate on 09/01/2006. Based on this, the hydrologic simulation was performed for the time period of 09/01/2006 to 01/05/2007. During this time period hourly data for the flow were also available from the hydrometric station of Agios Georgios (in the SWS 1), which assisted the calibration of the HSPF model.

All the input requirements for all sub-basins were provided to HSPF model using the Basins 4 model. The time-series of the karstic discharge from the Stylos spring were entered in the code of the User Control Input (UCI) file of HSPF, in order to calculate both the surface and underground discharges.

Following the simulation the process of calibration of the parameters was performed. The simulation results were introduced to the BASINS model for analysis using the tool Scenario Generator (GenScn).

The following parameters were used for the calibration (Kim et al., 2007):
a) The LZSN: is the lower zone storage nominal, b) The UZSN: is the upper zone storage nominal, c) The INFLT: is the soil infiltration rate, d) The LZETP: is the lower zone of evapotranspiration, e) The AGWRC: is the groundwater recession rate, it is rate of the groundwater discharge.

4.3 The Results of simulation of the HSPF model

In order to compare the field data with the simulated ones the Root Mean Squared error (RMSE) was used, defined as:

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} R_i^2} \quad (7)$$

Values of RMSE close to zero indicate perfect fit and the model calibration is considered satisfactory only when RMSE < 3.0. For the above case the RMSE = 2.58. The Figure 7 shows the comparison between the simulated results (HSPF model) and the observed data for sub-basin 1.

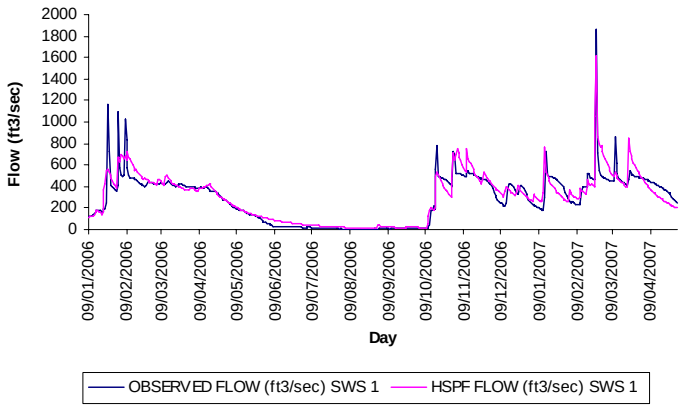


Figure 7. Comparison of simulated and observed flow for sub-basin 1.

4.4 The Sensitivity Analysis of the HSPF Hydrologic Model

After the calibration of the HSPF model was completed, the sensitivity analysis was performed to study the sensitivity of the calibrated results in comparison to field data (Al-Abed et al., 2002). In order to accomplish this, the main hydrological parameters of the calibration were disturbed by ±10%, in order to calculate the sensitivity of each parameter. The sensitivity analysis was only performed for the 'Coastal sub-basin' SWS 1, since it was the only sub-basin with field measurements. The sensitivity index S used in this part of the study is defined as:

$$|S| = (Q_{new} - Q_{calibrated}) / (K_{new} - K_{calibrated}) \quad (8)$$

Where:
$K_{new,}$ $K_{calibrated}$ is the new and calibrated value of each respective hydrological parameter (K) of the HSPF model and $Q_{new,}$ $Q_{calibrated}$, are the simulated values for $K_{new,}$ and $K_{calibrated}$ respectively. The lowest the value of S the less the sensitivity of the parameter is.

The results of the sensitivity analysis for the most important parameters of calibration are summarized in Table 2. The highest sensitivity appears for the parameter INFILT followed by UZSN, and LZSN.

Table 2. Sensitivity Analysis of main parameters of model calibration

Sensitivity Value of INFILT	Sensitivity Value of UZSN	Sensitivity Value of LZSN
0.59	0.09	0.02

5. CONCLUSIONS

The aim of this study was the modeling of the hydrologic activities of the Koiliaris River basin (Prefecture of Chania, Crete). The preliminary results of the simulation, using the HSPF model, have shown that there was not a good agreement between the simulated and observed data. This is due to the karstic geomorphology of the system which has a significant contribution to the Koiliaris River flow in the form of groundwater flow. In order to improve the agreement between simulated result and field data the HSPF model was combined with the 'two part' Maillet Karstic model which also considers the snow melt rate. As it was expected the snow melt rate has an important role to the final form of hydrograph. The final results of the above simulation showed a very good agreement with the observed field data. Since in the area of study flood phenomena occurred from time to time the present model could become a useful tool for the prediction of a flood event and for the better organization of a flood management plan.

6. REFERENCES

Al-Abed N.A., H.R. Whiteley, Calibration of the Hydrological Simulation Program Fortran (HSPF) model using automatic calibration and geographical information systems, *Hydrological Processes, 16*, 3169-3188, 2002.

Bicknell, B.R., J.S Imhoff, J.L. Kittle, T.H. Jobes, and A.S. Donigian, Hydrological Simulation Program - FORTRAN (HSPF): User's Manual –Version 12. National Exposure Research Laboratory, Office of Research And Development, U.S. Environmental Protection Agency, Athens, Georgia, U.S.A, 2001.

Jaquet, O., P. Siegel, G. Klubertanz, H. Benabderrhamane, Stochastic discrete model of karstic networks, *Advances in Water Resources, 27*, 751-760, 2004.

Kim, S.M., B.L. Benham, K.M. Brannan, R.W. Zeckoski, and J. Doherty, Comparison of hydrologic calibration of HSPF using automatic and manual methods, *Water Resources Research*, Vol 43, W01402, 2007.

Nikolaidis, N.P., H. Rajaram, J.L. Schnoor, K.P. Georgakakos, A Generalized Soft Water Acidification Model, *Water Research, 24*(12), (1983-1996), 1988.

Shaban, A., M. Khawlie, Ch. Abdallah, Use of remote sensing and GIS to determine recharge potential zones: the case of Occidental Lebanon, *Hydrogeology Journal, 14*:433 -443, 2006.

Stamati, F., N. Nikolaidis, K. Bozinakis, D. Papamastorakis, M. Kritsotakis, Stochastic Modeling of the Karstic System of western Apokoronas in Crete, In VIII International Conference Protection and restoration of the environment VIII, CChania Greece, 184-192, 2006.

Tzoraki O. and N.P. Nikolaidis, A generalized framework for modeling the hydrologic and biogeochemical response of a Mediterranean temporary river basin, *Journal of Hydrology*, vol 346, pp.112-121, 2007.

U.S. Army Corps of Engineers, Snow Hydrology, Washington DC, Department of Commerce, Office of Technical Services, PB 151660, 1956.

Yang Z.-L., R.E. Dickinson, Validation of the Snow Submodel of the Biosphere–Atmosphere Transfer Scheme with Russian Snow Cover and Meteorological Observational Data, *Journal of Climate*, *10*: 353-373, 1997.

iEMSs 2008: International Congress on Environmental Modelling and Software
Monitoring strategies and spatial distribution of processes in models for natural resources management on different scales - vision and reality
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.S5
Martin Volk and Ulrike Bende-Michl (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Application of the HSPF Model for Flood Simulation with Analysis of the Results in Terms of Monitoring Uncertainties /Case Study of the Lesnovska River, Bulgaria/

P.Ninov[a], I.Ribarova[b], P.Kalinkov[b], G. Dimova[b]
[a] *National Institute of Meteorology and Hydrology – Bulgarian Academy Sciences (plamen.ninov@meteo.bg)*
[b] *University of Architecture, Civil Engineering and Geodesy , (ribarova_fhe@uasg.bg)*

Abstract: Floods are amongst the most damaging natural disasters worldwide. The new EU Floods Directive and the implementation of the EU Adaptation Green Paper will require improving of the flood-related research. In this study, the HSPF model was applied to simulate and analyze a very significant and unexpected flood occurred in 2005 along the Lesnovska River, a right tributary of the Iskar mainstream in the western part of Bulgaria. The difference between the simulated and the measured flood volumes was found very high and equal to 130% for the flood of 2005, while percent differences of 25-33% were observed for the seasonal, annual and low flow distribution. These results were compared with an application of the same model for another part of the river system and another flood event in order to find explanation of the discrepancies. The monitoring uncertainties are discussed to be the main reason of the unsatisfactory model performance for the flood occurred on August 2005. On the basis of the obtained results, the flow was recalculated and the probable values of the peak flow were estimated. The needs of implementation of new monitoring practices and technoquies for better flood event knowledge and modeling are outlined.

Keywords: HSPF Model; Modeling; Flood; Monitoring Uncertainties.

1. INTRODUCTION

Floods are amongst the most damaging natural disasters worldwide. The new EU Floods Directive and the implementation of the EU Adaptation Green Paper will require improving of the flood-related research. In the scope of an EU project, Floodmed, eight models (*HEC-HMS, Vflo, MISD, Mike-drift, TRIMR2D, IMGW, Vidra and HSPF*) were applied in seven European catchments (Greece, Poland, Romania, Germany, Italy, Bulgaria, Serbia) to test the model applicability for flood identification [Ribarova, 2008c]. The models showed different capabilities. Although the simulations could be assessed as successful, some imperfections were noticed. The HEC-HMS could not offer a sufficient simulation of the second peak event in hydrographs with two successive flow peaks proving that antecedent moisture conditions play an important role in hydrograph simulation efficiency. TRIMR2D is suitable to simulate flood inundation if high resolution input data are available. IMGV should be used and the results should be interpreted by experienced hydrologists. The selection of the modeling method and the mean precipitation estimation should be passed with relation to conditions within the catchment as well in neighborhood sub – catchments.

The main common problem in modeling of all flood events was reported to be data availability and accuracy [Ribarova, 2008c]. This reflects to modeling uncertainty and is associated with:

- Lack of detailed rainfall data sets. In many catchments the rainfalls were calculated according to the data from the existing stations and sometimes this number was not adequate, especially where temporal distribution of rainfall greatly vary geographically.
- Lack of detailed soil studies, describing the physical properties of soil types such as texture, hydraulic conductivity, etc. Thus, infiltration was estimated from the geological map of the basin, based simply on general knowledge about the permeability of different formations.
- Errors of measured stages coming from problems related to the equipment maintenance as well as enourmous water quantity during the floods.

In this paper, the results of the application of the HSPF model of the US EPA (Environmental Protection Agency) to simulate and analyze a very significant and unexpected flood occurred in 2005 along the Lesnovska River, a right tributary of the Iskar mainstream in the western part of Bulgaria are reported. The problems of data uncertainties are discussed as well. Furthermore, the paper attempts at using the model results for recalculation of the most likely flood values. The HSPF model has been chosen because of its capability to run at time steps of less than a day, since flood events occur over very short periods, but generate large changes in river flow. In addition, the model was applied successfully for flood identification in other studies [Albek, 2004; Karatzas, 2007; Ribarova et al., 2008a].

2. THE FLOODED AREA OF THE LESNOVSKA RIVER

The Lesnovska River is flowing through the Sofia plain and is predominantly a flat river with total area of 1096 km^2. The catchment area is relatively densely populated with numerous small towns and villages gravitating to the capital Sofia [Gerasimov et al., 1986]. Mountainous and hilly areas surround the valley. The climate is moderate continental with temperatures varying between minus 15^oC in the winters and more than 35^oC in the summers. The average annual precipitation is 832 mm. Snowfalls are common in winter and snowmelt contributes to the high flows is in the spring. The high flows in the other seasons are due to the precipitations only.

Extreme flood events were registered in the Lesnovska river watershed, at the Dolni Bogrov Gauge station (North Latitude 4733833.102; East Longitude 212859.780), on 5th of August and 6th of August, 2005 (Figure 1). The average daily flows were reported to be 449.8 m^3/s and 403.0 m^3/s respectively.

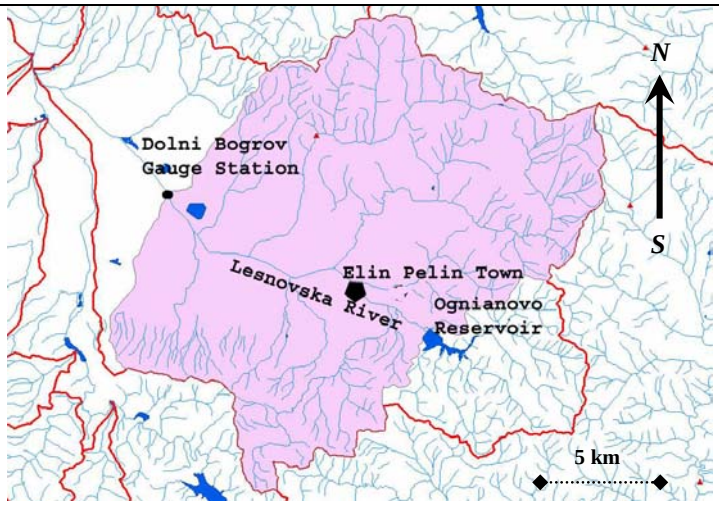


Figure 1. The Lesnovska River watershed with the modelled area between the Ognianovo Reservoir and the Dolni Bogrov Village

Figure 2 visualises the average daily flow for Lesnovska river in August, 2005, compared with the average long term yearly flow of 3.290 m^3s^{-1} for the period 1961-2006 [National Institute of Meteorology and Hydrology, 2005]. The monthly average flow in August, 2005 was more than 100 times higher than the typical average August flow of previous years. This flood event destroyed the life of local people and caused significant infrastructural and agricultural damages.

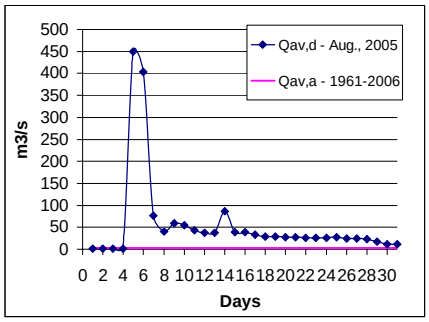


Figure 2. Average daily flows in August 2005 for the Lesnovska river at Dolni Bogrov compared to the average long term yearly flow [Ribarova et al, 2008b]

3. APPLICATION OF THE HSPF MODEL

The HSPF model simulates the time response of the watershed based on hydrological mass balances integrating point and non-point sources [Duda, et al., 2001]. The HSPF model requires two types of data: GIS based data describing the studied watershed and daily hydrological and meteorological time series. The HSPF model was created for the BASINS system [Russel et al., 2003]. It was successfully applied by the research group for modeling of the river run off and quantity at normal climate conditions as well as for simulation of the first flood event after dry summer period [Ninov et al., 2004; Ribarova et al., 2005; Ribarova et al., 2008a, b].

The HSPF model was used to simulate the daily flows at the equipped river section of Dolni Bogrov for the period 2003-2005. The calibration of the model for the Lesnovska River was based on the collected meteorological, hydrological and GIS data for the period 2003-2004. The model was verified for year 2005 when also an extreme flood event occurred in August. The hydrology calibration process follows three subsequent steps: adjustment of the annual trends and water balances, seasonal discrepancies and high and low flows distributions (including stormy periods). The water balances were calculated by estimating the volume of the water passing through the reach according to the hydrometric station (observed data). They were compared to the output volume simulated by the HSPF model. The losses in the watershed were accounted assessing the flow paths, the evapotranspiration and the infiltration. The model uses numerous calibration parameters: functions of the soil; climate; topography; vegetation cover; land use; and geology conditions [Ribarova et al., 2008a, b]. Table 1 shows the values selected for the main hydrological parameters for the Lesnovska river basin.

Table 1. Lesnovska River basin: calibration parameters values for the HSPF model

Parameter	Definition	Units	Values
LZSN	Lower zone nominal storage	inch	9
INFILT	Index to the infiltration capacity of the soil	inch/hr	0.01-0.20
SLSUR	Slope of the assumed overland flow plane	-	0.001
KVARY	Variable groundwater recession	1/inch	0
AGWRC	Base groundwater recession	-	0.995
DEEPFR	Fraction of GW inflow to deep recharge	-	0.35
BASETP	Fraction of remaining ET from baseflow	-	0.20
AGWETP	Fraction of remaining ET from active GW	-	0
CEPSC	Interception storage capacity	inches	0.12
UZSN	Upper zone nominal soil moisture storage	inches	2

During the calibration phase, different scenarios were examined with the aim of reducing discrepancies in the annual water balance and in the seasonal and high/low flow distribution adjustment. The simulation results are shown in Figure 3. Because of the very high peak of the flow on 5th and 6th August, 2005 the curves at the bottom are "crushed". Therefore, the lower flow range (less than 50 m^3/s) is visualised in Figure 4 for better explicitness.

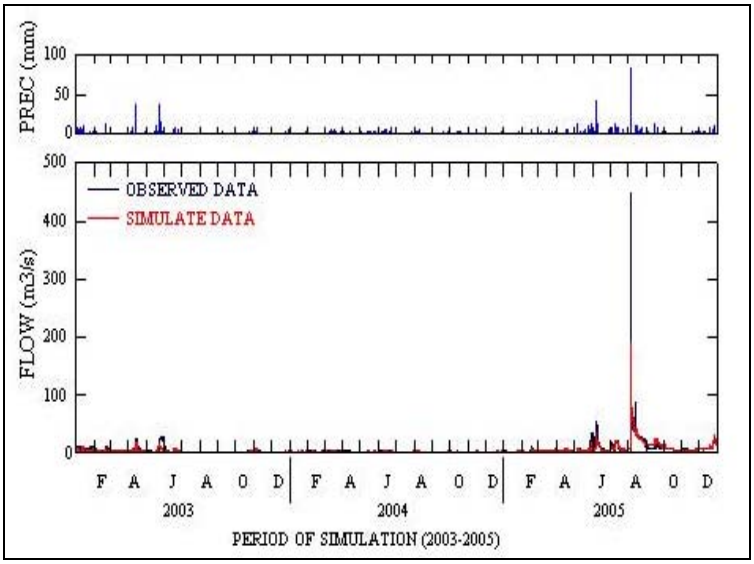


Figure 3. Observed and simulated daily flows and precipitation at Dolni Bogrov Gauge Station – Lesnovska River

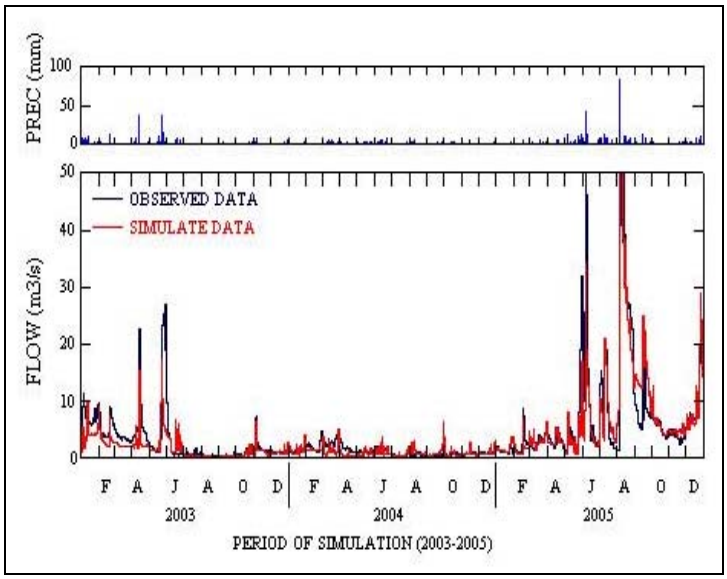


Figure 4. Low-range (less than 50 m^3/s): observed and simulated daily flows and precipitation at Dolni Bogrov Gauge Station – Lesnovska River

According to the calculations (Table 2) the results for the annual balance, seasonal discrepancies and the low flows distribution for the Lesnovska River can be assessed with the absolute volume errors approximately between 1/4 and 1/3 (or 25% to 33%, respectively). Similar are the results for the calibration and the verification periods assessing them as satisfactory with a good fit for the two periods (Table 2). However, the results for the flood peak simulations are not satisfactory – the simulated flow is 130% less than the observed one (Table 2).

Table 2. Annual water balance, seasonal water balance, highest and lowest water balance – Lesnovska River (2003-2005)

Period	Observed flow (m^3)	Simulated flow (m^3)	Absolute volume error (%)
Total period	444 956 588	353 381 547	-25.9
Period of calibration	131 935 095	96 749 975	- 28.9
Period of verification	313 021 464	256 631 572	- 22.0
Winter seasons	53 954 197	40 492 222	-33.2
Summer seasons	213 612 211	159 207 228	-34.2
Sum of lowest flows (Aug-Sept 2003)	1 778 969	1 431 006	-24.3
Sum of highest flows (5/8/05-8/8/05)	88 923 980	38 564 720	-130.0

There are two possible explanations for the high discrepancy between the measured and the simulated flood data: 1) modelling problems – HSPF model is not capable to account for high flood peaks; 2) monitoring problems - objective difficulties of the hydrological and precipitation monitoring for the flood period and uncertainty of the information. We consider that in the case of Lesnovska river flood event, the second explanation is more likely. On one hand, the past experience with the HSPF model shows good results for high flows simulations dealing with reliable hydrological information [Ribarova et al., 2008a]. On the other hand, during the studied flood event, a large area around and within the village of Dolni Bogrov was covered with slowly moving and possibly even stagnant water, which was most likely added to the measured flow. For this reason, some regional experts doubt the reliability of the existing hydrological information.

The research group had applied previously HSPF model successfully for the Upper Iskar River, located upstream of the Lesnovska tributary. Both watersheds have similar physical-geographic and basic hydrological characteristics. The used calibration parameters have the same or close values [Ninov et al., 2004]. Figure 5 presents the simulated and observed flows of Upper Iskar River for the first flood event after summer period, happened in November, 20004.

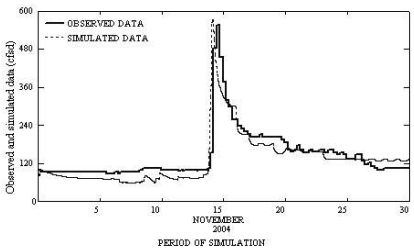


Figure 5. Observed and simulated hourly flows for the flood event in Upper Iskar River in November, 2004 [Ribarova et al., 2008a]

This figure shows clearly the HSPF model simulated quite satisfactory another studied flood event. It gives us confidence to state that the HSPF modeling revealed the uncertainty of the observed flood flows during the studied period, in August 2005 and arouses the question for identification of the flood monitoring problems and the needs of implementation of new practices and technoquies [Ribarova, et al., 2008b].

4. IDENTIFICATION OF FLOOD MONITORING PROBLEMS

The most serious problem concerning the flood monitoring is of hydrometric nature. During the floods the river flows go out of the river beds covering parts of the terraces and valleys around the rivers. The usual measurements of water levels sometimes become impossible due to inundated or inapproachable stations. The effect of floods may augment if dikes and dams are broken. The hydrological characteristics of the floods have a temporal and spatial changeability linked to the changeability of the originating factors as: precipitation intensity

and continuity, diversity of their spatial distribution, geographic and topographic conditions in the affected watersheds, level of urbanization, etc. [Gerasimov et al., 1992]. Some indirect approaches are applied for flood events to determine the hydrological elements, such as: the usage of the traces on the banks, trees and buildings showing the highest water level during the floods; determination of river bed profiles before and after the flood; usage of hydraulic formulas to assess flow velocities and etc. All these indirect approaches impact on the quality of hydrologic information during the floods and sometimes increase significantly the uncertainty of the information.

Another problem concerning flood monitoring is the frequency of the measurements. The usage of automatic stations can not guarantee the regular and precise information in short intervals. Serious damages and failure (stopping) of the automated stations, located nearby the river beds, are often before the highest pick of the flood.

The data from the State monitoring hydrological network of the National Institute of Meteorology and Hydrology, even correct, reflect the situation in a narrow river sector and could not be representative for the whole damaged zone when the disaster is enormous and spreads over a big area. Uncertainties related to stage-discharge relationships for high stage values are also a common problem. In this case the network information has to be obligatory enriched with expedition observation and measurements.

The uncertain observations for hydrological data and the unreliable flood informational input in the model (as is the case of Lesnovska River flood, August 2005) show that there is a need for implementation of new monitoring practices and techniques for better flood event knowledge and modeling. There are not commonly accepted opinions or proposals how to cope with these problems bringing up uncertain hydrological information, as well as what practices can be applied to guarantee the reliable monitoring information. The modern remote methods give the most complete picture for the real flood distribution of the affected area – photography in the optic, infra-red and supersonic part of the electromagnetic specter of frequencies using planes, helicopters or cosmic satellites. Unfortunately these methods are expensive and are still not widely spread. Using a dense automatic stations network with frequent, in-time information is technically feasible and very useful. The cheapest way remains the expedition observations during the flood with all risks to add a stagnate water to the mainstream increasing the real running volume during the flood, as perhaps was the case of the Lesnovska River, in August 2005.

5. RECALCULATION OF THE PROBABLE FLOWS DURING AUGUST 2005 FLOOD

We have discussed above that most likely the measured values of the flood event in August 2005 are not correct. The exact uncertainty of observed flows due to measurement errors is very difficult to be assessed. The measurements were done in extraordinary conditions for a relatively short period. The river flows overflowed the bed and banks, covering the nearby fields and running trough the streets of the urbanized territories, shaping areas with stagnant water or slowly moving currents. Some places were dangerous to be approached. Years after disaster we could only rely on other methods, as hydrological models usage, to assess quality of the observed flood information.

We suggest using the results of the HSPF modelling for rough assessment and recalculation of the probable flood flows of Lesnovska River. The model calculates the annual, seasonal and low flows balances approximately 25% to 33% less than the observed ones, except for the period of the flood (5-6 August, 2005), when this value is 130% (Table 2). If we accept that in principal the HSPF model calculates any flood values correctly, we may expect that for the case of Lesnovska river the most probable values of the flood flow would be also 25% to 33% higher than the simulated ones (as for all other data during the 3 years calibration and verification period). Of course, this conclusion, based on the HSPF model results interpretation, can not be strictly proven especially several years after the disaster. Table 3 shows the observed and simulated daily highest flows and those increased with

25% and 33%, outlining the probable range of the real running daily flood pick. The most probable values of the daily highest peaks are calculated to be between 237.9 and 253.1 m^3/s.

Table 3. Observed, simulated and probable highest daily peaks during the flood 5 – 6 August 2005 in m^3/s – Lesnovska River at Dolni Bogrov Station

Flow	Values
Observed daily highest peak	449.78
Simulated daily highest peak	190.3
Simulated daily highest peak increased with 25%	237.9
Simulated daily highest peak increased with 33%	253.1
Most likely daily flood peaks	237.9 – 253.1

Figure 6 visualizes the range where the real daily highest peaks are most likely to be disposed during the flood occurred on August 2005. The presented boundaries are based only on the absolute volume error of the annual, seasonal and extreme balances between simulated and observed data and have a tentative character. However, the study is a step towards the clarification of the real picture of the flood when there is uncertain, probably strongly exaggerated, monitoring hydrological information.

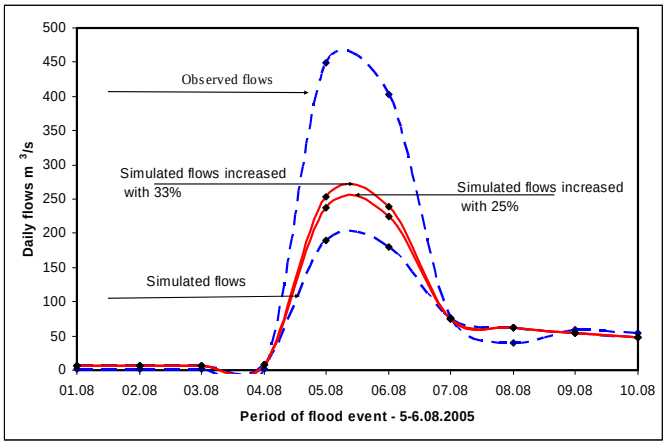


Figure 6. Range of the probable daily highest peak during the flood in 5-6 August 2005 at the Dolni Bogrov Gauge Station - Lesnovska River

6. CONCLUSIONS

The presented investigation reveals the uncertainty of the hydrological information during the extraordinary flood in August 2005 for the Lesnovska River case study. Although being unique one, this event represents some common monitoring problems concerning the flood disasters. A possible solution for overcoming of flood data uncertainty is a combination of the classical methods with modern remote approaches. Through improvement of the quality of the monitoring information during the floods we can enhance our knowledge of the events in order to better simulate and forecast them. Introducing the modern remote approaches as operational practices in the case of extreme flood events would improve the hydrological information and would help the modeling and better forecasting of the disasters. The main problem of their application spreading is the increased financial expense for techniques, staff and organization, as well as the necessary inter-institutional cooperation and agreements.

The application of comprehensive hydrological models, as the HSPF model, helps us not only to identify periods with uncertain information but to try to restore the probable real flows running during the floods.

ACKNOWLEDGEMENT

This work was financially supported by the European Commission under INTERREG III B CADSES project Floodmed. The contents of this paper are the sole responsibility of the authors and can under no circumstances be regarded as reflecting the position of the European Union. The authors thank the Ministry of the Environment and Waters and the National Institute of Meteorology and Hydrology - Bulgarian Academy of Sciences for sharing their data.

REFERENCES

Albek, M., U. Ogutveren, and E. Albek, Hydrological modeling of Seydi Suyu watershed with HSPF, *Journal of Hydrology*, 285 (1-4): 260-271, 2004.

Duda, P., J. Kettle, Jr. Gray, P. Hummel, and R. Dusenbery, WinHSPF, U*nited State Environmental Protection Agency. DC 20460, 2001*.

Gerasimov, J., and J. Galabov, Geography of Bulgaria, Vol.1 Physical Geography, *Pub. House Bulgarian Academy Sciences*, Sofia, 1986.

Gerasimov, S., Specific problems concerning flood investigations, *Scientific conference for people's protection during the disasters*, Conference Proceedings, vol.V, pp 111-118, Sofia, Bulgaria, 1992.

Karatzas, G., and N. Kourgialas, Hydrological Simulation Program-Fortran-HSPF of the Koiliaris River basin, *Floodmed Project Report*, Greece, 2007.

Ninov, P., I. Ribarova, N. Nikolaidis, R. Tsoraki, P. Kalinkov, K. Kukurin, and J. Topalova, Hydrologic simulation of Iskar river subwatershed using the HSPF model, *Conference on Progress in surface and subsurface water studies at the plot and small basin scale*, pp. 167-170, Turin, Italy, 2004.

Ribarova, I., J. Topalova, P. Ninov, K. Kukurin, and P. Kalinkov, Assessment of the flush event significance for a river with typical seasonal run off, 10 International Specialist *Conference on Watershed and River Basin Management*, Conference Proceedings, CD, Alberta, Canada, 2005.

Ribarova I., Pl. Ninov, and D. Cooper, Modeling nutrient pollution during a first flood event using HSPF software: Iskar River case study, Bulgaria, *Ecological modeling*, vol.211, pp. 241-246, 2008a.

Ribarova, I., P. Ninov, K. Daniell, N. Ferrand, and M. Hare, Technical and non-technical approaches for flood identification, *Conference Water Down Under 2008*, Australia, 2008b.

Ribarova I., Summary on modeling software robustness and suitability for the CADSES region, Part 2: Model application, *Floodmed Project Report*, Bulgaria, 2008c.

Russel, K., P. Cocca, E. Partington,, M.Wellman, and D. Wells, Better Assessment Science Intagrating Point and Nonpoint Sources., U*nited State Environmental Protection Agency*, 2003.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Role of Models for Building an Efficient Monitoring Design

M.G. Erechtchoukova[a] and P.A. Khaiter[a]

[a]Atkinson Faculty of Liberal and Professional Studies, York University, Toronto, Ontario, Canada (marina/pkhaiter@yorku.ca)

Abstract: Efficient assessment of water quality requires an integrated approach incorporating such tools as sampling programs supplying data for statistically valid assessment along with watershed and water quality models which are used to transform the collected data into information. It is important to quantify and possibly to reduce the uncertainty of the information supplied by a monitoring system. The paper describes the algorithm for an efficient temporal monitoring design for data collection at a given site of a fixed station tiered monitoring system. The algorithm takes into account features of models used to obtain estimates of investigated water quality indicators. Efficiency of a monitoring design is considered under the assumption that the cost of a sampling program is a monotonously increasing function of a number of observations. Application of the proposed algorithm to various water quality ingredients reveals a wide range of sampling frequencies needed for estimation of water quality ingredient load with a desired level of uncertainty. Monitoring design which is common for all observed water ingredients at a given site and supporting evaluation of their annual loads with the same level of uncertainty can hardly be attained due to financial and technical constraints. For tiered monitoring systems, different allowable levels of uncertainty can be recommended for water ingredients according to their importance for a particular site.

Keywords: Monitoring system; Water quality indicator; Chemical load; Monitoring design; Uncertainty

1. INTRODUCTION

Sustainable decision making requires adequate information about the present quality of the environment and its possible changes in the future. Environmental quality is usually determined via comparison of a set of values of selected indicators with existing standards and is aimed at verifying the suitability of an environmental resource for a designated use (e.g., recreation or drinking water supply). As a rule, environmental indicators are quantifiable variables reflecting physical, chemical or biological characteristics of natural ecosystems at a given moment in time and a certain point in space. With respect to the aquatic environment, its quality refers to physical conditions including temperature and presence of particulate matter and to chemical conditions described by concentrations of dissolved ingredients. Values of environmental indicators must be obtained by direct observation and measurement implemented under a certain program. Long-term standardized measurement, observation, evaluation and reporting of the aquatic environment in order to define status and trends of a water body are called monitoring.

Monitoring systems comprise of such components as collection and analysis of physical, chemical and biological data and quality assurance and control programs to ensure that the data are scientifically valid. Canada-wide framework for water quality monitoring [WQTG, 2006] identifies the following aspects which must be taken into account: monitoring program objectives, monitoring program design, field sampling program, laboratory

analysis and procedures, data analysis and interpretation, reporting and follow-up. While all these aspects are important for monitoring system functioning, the current study deals only with those which require mathematical tools. For efficient assessment of water quality, EPA recommends an integrated approach incorporating several techniques [USEPA, 2003]. These techniques include sampling programs supplying data for statistically valid assessment along with watershed and water quality models which are used to transform collected data into information.

One of the essential characteristics of information supplied by a monitoring system is uncertainty [Harmel et al., 2006]. According to Quality Assurance Plan [USEPA, 2003], it is important to understand and quantify the uncertainty and incorporate its estimates into environmental assessment. Uncertainty of an estimator utilized depends on its mathematical properties, the variability of an investigated environmental indicator and an available data set [Erechtchoukova, 2005]. The larger the set, the lesser the uncertainty of values calculated based on the set. However, for the majority of important water quality indicators and many sampling sites, extensive observations are not possible due to the logistic and financial constraints. Models can improve significantly the reliability of obtained information by reducing uncertainty of model outcomes if available data sets are sufficient for model application. Strictly speaking, the quality of generated information depends on the model used [Erechtchoukova and Khaiter, 2007]. Investigation of model properties can give an insight about possible ways to reduce the uncertainty of the information and to suggest monitoring designs improving estimates of environmental indicators [Strobol et al, 2006; Urban, 2000].

Selection of a model depends to a great extent on investigated indicators. Shrestha et al. [2008] pointed out that for effective water quality management estimates of ingredient loads are more important than concentrations. Chemical loads can be used as an objective for effective design of a monitoring program [Hooper et al., 2001].

The paper investigates the role of models in water quality monitoring. It describes the algorithm for an efficient temporal monitoring design for data collection at a given site which takes into account features of models used for data analysis and uncertainty associated with the collected data. Chemical load is chosen as a water quality indicator of interest.

2. MONITORING SYSTEMS

Monitoring systems provide broad sets of data collected in accord with a program designed for a specific set of scientific, environmental or managerial objectives. Canada-wide framework for water quality monitoring includes three main phases: (1) planning which determines objectives and scope of the program; (2) collection/analysis incorporating field sampling, laboratory analysis, data management and processing, data analysis and interpretation, data reporting and quality assurance/quality control; and (3) information utilization for decision making, education and policy development and enforcement [WQTG, 2006]. Although monitoring objectives differ for various monitoring systems, in general, they include such common tasks as determination of water quality standards to be attained, attainment of the standards, identification of impaired waters, as well as causes and sources of water quality impairments and detection of long-term trends [USEPA, 2003]. Data collection must be conducted according to a proposed monitoring design. The monitoring design is to reflect these objectives as well as to provide reliable data for decision making. The extent, to which collected data represent real state of the aquatic environment, depends on a chosen spatial and temporal monitoring design. That is why its selection is important for many tasks of environmental assessment.

There are several approaches to a monitoring design. Fixed station approach assumes that the same sites are repeatedly sampled at regular time intervals over a long period of time. Short-term monitoring is a specific study which investigates particular water quality

problems and creates a 'snapshot' of the conditions in a given area. Rotating-basin approach is based on intensive short-term surveys which are conducted periodically. It may identify changes in water quality conditions over time. In a probability-based approach, sites are selected randomly from the total set of sites on water bodies in a selected area. An exhaustive approach requires sampling or surveying of all water bodies in the area. A tiered approach [USEPA, 2003] is adopted in many monitoring systems. The approach requires identification of a core set of water quality indicators which reflect designated uses and can be monitored routinely to assess attainment of applicable water quality standards. In addition to the core set of indicators of the aquatic environment, it is also necessary to identify supplemental indicators dictated by the site or project specific needs. In general, the selection of variables to be measured depends on such factors as monitoring objectives, site specific water quality issues and designated uses of interest which may result in significant variations of core sets for different sites across an investigated region. A consensus on core sets of indicators is strongly desirable since it creates compatible and sharable data sets for large scale analysis and data generalization [WQTG, 2006]. Water quality indicators from core and supplemental sets are observed with different frequencies at the same site. Moreover, core indicators can be observed with different frequencies at different cross-sections of the same section of the natural stream because of the importance of a particular location.

No single monitoring approach is sufficient to provide the data for all information needs. To meet the objectives, monitoring systems integrate several designs or programs of observations. Thus, fixed station approach along with tiered monitoring design coupled with sampling programs reflecting environmental heterogeneity is useful for long-term trend detection, for assessment of critical reaches of large streams, and at the same time provides site-specific water-quality data.

3. MODELS IN MONITORING DESIGN

Observation data supplied by monitoring systems are a mandatory component of any environmental decision making process, but this component can be useful only if the data are synthetized into information. Ideally, information has to be comprehensive and complete to meet multiple relevant needs. In many cases of environmental decision making, understanding of interactions of key environmental processes is vital. It cannot be achieved by observations along. This is the moment when models come into play. Environmental models form a diversified set of techniques based on different mathematical and computational methods [Jørgensen and Bendoricchio, 2001; Straškraba and Gnauck, 1985]. Models transform observation data into information by extracting aggregate values from raw data, projecting values of selected environmental indicators and detecting trends to track changes in water quality. Models, in order to be applied, impose additional requirements on the way the data are acquired. A data set available for analysis must satisfy the assumptions that underlie mathematical techniques employed for data analysis (models from group 2 in Figure 1). These assumptions must accord with models from group 1. The latter can be used to determine frequencies of observations sufficient for deriving statistically meaningful results.

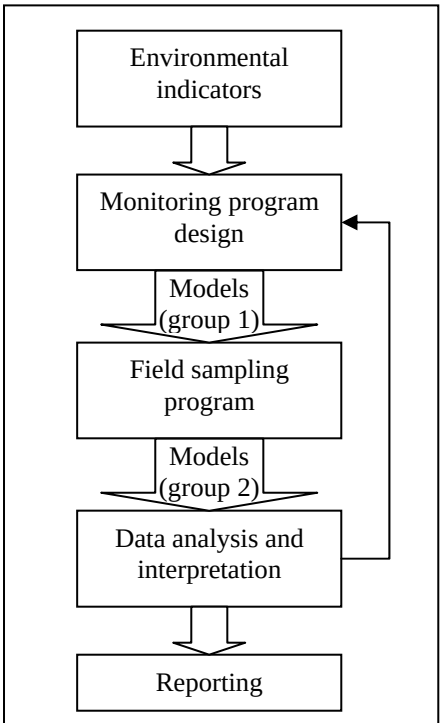


Figure 1. Model application to monitoring design.

Strictly speaking, frequencies of observations conditioned by models from group 1 reflect variability of the model results rather than the natural environmental heterogeneity. They are derived from the model properties and seem to contain an error. Being a simplified representation of reality, no model can fully duplicate real system behaviour and, thus, introduces an error which is also referred to as model uncertainty. Model uncertainty, in its turn, transforms into the errors in recommended frequencies. These errors can be minimized by selecting a model which describes an investigated system better than others under given assumptions. Information extraction from observation data heavily depends on chosen models. At the same time, model selection is significantly restricted by available observation data and, hence, data collection must fit entire modelling process [Richardson and Berish, 2003]. This interdependency of data and models calls for the necessity to design sampling programs based on specific statistical or mathematical assumptions which must be consistent with the ways collected raw data are analysed and with the type of models to be used for this purpose.

4. UNCERTAINTY IN WATER QUALITY MONITORING

Conclusions drawn from monitoring data always contain uncertainty which is introduced by all components of a monitoring system. One of the sources of uncertainty, particularly model uncertainty, has been already mentioned. It is necessary to add observational artifact and the uncertainty introduced by selecting sampling sites. The very idea of the monitoring to describe continuous fields of environmental indicators by discrete samples collected from time to time implies the uncertainty since it is based on the assumption that values of observed indicators remain steady in a neighbourhood of a sampling site for some period of time which is not always valid. This type of uncertainty can be reduced by optimizing spatial and temporal monitoring design and most likely by introducing additional sampling sites with higher frequencies of observations, but can hardly be eliminated. Observational artifact is caused by measurement tools and analytical methods used in the laboratories in order to obtain values of environmental indicators of interest. Although some improvements of the results are possible, this type of errors is always present in monitoring data [Harmel et al., 2006]. Usually, monitoring guidelines recommend keeping both types of uncertainty under a 10% level.

Model uncertainty plays an important role in the analysis and interpretation of monitoring data. Models receive the uncertainty from previous monitoring phases and transform it into information required for decision making. The resulting uncertainty must be understood, quantified, and limited to a reasonable extent with respect to the cost of possible consequences of decision errors.

Commonly accepted definition of model uncertainty describes it as deviations of simulated system variables from their known or observed values [Campolongo et al., 2000]. There are various sources of model uncertainty. First of all, the uncertainty is caused by the model intrinsic feature as an abstraction of reality. Other sources of model uncertainty depend on model structure, mathematical formulae and approaches employed in model components. Model uncertainty undoubtedly influences the process of conversion of monitoring data into information.

It is necessary to evaluate the extent to which the model outputs are uncertain and to attempt to reduce this uncertainty in the results. Model uncertainty reduction is mainly achieved by selecting a particular model structure or parameter values. Although these two sources of model uncertainty are extensively investigated [e.g., Snowling and Kramer, 2001; van Nes and Scheffer, 2005], they create confounding effect which impedes identification of uncertainty caused solely by the model structure. That is why the following aspect must be taken into account during model development. The uncertainty of a chosen model under a given set of observation data must be less than the uncertainty delivered by any other model on the same data set. Therefore, the properties of the chosen model can be used for planning an efficient monitoring design. Strictly speaking, an

efficiency of a sampling program can only be determined against certain criteria. It is possible to relate these criteria with some kind of financial feasibility study or particular technical aspects that also suggest monetary estimates. Under the assumption that the cost of a sampling program is a monotonously increasing function of a number of observations, the criterion of efficiency can be rephrased as the problem of minimizing a required number of observations sufficient to keep the resulting uncertainty at a desired level.

5. TEMPORAL MONITORING DESIGN

Selection of a model for data analysis and interpretation is stipulated by monitoring objectives and an approach to a monitoring design. Chemical loads are important indicators of water quality. Using annual chemical load of water ingredients as selected water quality indicator is an effective way for organizing a large-scale monitoring system [Hooper et al., 2001]. For management purposes, Shrestha et al. [2008] also give preference to ingredient loads over their concentrations.

There are several approaches to load estimation. The overview of these approaches and their classification can be found, for example, in Aulenbach and Hooper [2006]. Different formulae for load calculation are presented in Preston et al. [1989]. Chemical load estimates usually use values of water discharge and concentrations of an ingredient determined from the samples collected at a particular cross-section of a water body over a period of time.

Uncertainty of a load estimate denotes possible deviations of values calculated based on available data sets and a selected formula from the actual value. Statistically, load uncertainty can be interpreted as the variance of the formula chosen for approximation. Thus, resulting uncertainty can be minimized by selecting an estimator (i.e. model) that outperforms others on the same set of observation data. At the same time, scarce observation data can magnify the resulting uncertainty of the selected estimator. Therefore, the chosen model must be used for planning an efficient sampling program.

Improvements in the performance of an estimator are usually achieved in two ways: additional sampling or modifications of the estimator bringing additional knowledge about natural phenomena. Majority of load estimates use water discharges which are described by more detailed series of values than concentrations of water ingredients. With this respect, regression and ratio estimates have to be considered. Regression models are probably more popular. They are used independently or as a part of the composite method [Aulenbach and Hooper, 2006]. A regression model describes the relationships between water discharge and concentrations of a water ingredient and is actually used to restore missing or predicted values of concentrations. It is assumed that the relationship is steady and the model is suitable for the forecasting. The main restriction in application of regression analysis is the number of available samples. If the relationship between concentrations and water discharge is linear and concentrations are normally distributed, the rule of thumb requires at least 50 samples to evaluate the reliability of regression coefficients. This number increases significantly for other types of regression equations and water quality indicators with high variability.

Ratio estimates are based on calculation of average water discharge using all available data and instantaneous load values only when concentrations are measured. Strictly speaking, the ratio estimator has a bias which must be corrected. Although ratio estimators have minimal uncertainty if the relationship between two investigated variables is linear, there is no need to make such a rigorous assumption for entire investigated period. Stratification of the investigated period can significantly improve the performance of load ratio estimators. The stratified estimator of an ingredient load dominates non-stratified one, if the variance of the instantaneous load within the strata is less than its variance between the strata. The efficiency of a stratified estimator certainly depends on a stratification scheme applied. It is worth to note that stratification reducing load uncertainty can be achieved for the majority

of natural streams with distinct hydrological seasons [Erechtchoukova and Tsirkunov, 1989]. Then the following formula provides an efficient estimation of the load:

$$L = T\sum_{j=1}^{k}\frac{N_j}{N}Q_j\left(\frac{l_j}{q_j}+B_j\right), \tag{1}$$

where Q_j is the average water discharge in j-th stratum calculated from the most frequent measurements, q_j is the average water discharge in j-th stratum calculated using values corresponding to the observed concentrations, l_j is the average instantaneous ingredient load calculated from the observed concentrations, N is the total number of water discharge observations over the period, N_j is the number of water discharge observations in j-th stratum, T is the duration of the investigated period, B_j is the bias of the ratio estimate. The formula for the bias estimation can be found in Cochran [1963].

The required number of concentrations which keeps the uncertainty of load estimator (1) under a given level can be obtained from the problem of mathematical programming [Bodo and Unny, 1983; Erechtchoukova and Tsirkunov, 1989]:

$$\min\sum_{j=1}^{k} n_j, \text{ subject to } D(L)\le V,$$

$$\text{where} \quad D(L) = T^2\sum_{j=1}^{k}\frac{N_j^2}{N^2}S_j^2 \ . \tag{2}$$

Here n_j is the required number of concentrations per j-th stratum, V is the given level of uncertainty and $D(L)$ is the ingredient load variance, S_j^2 is the variance of average instantaneous load estimator in j-th stratum which depends on the number of observations, n_j, conducted during this stratum, k is the total number of strata. Such formulation is possible due to consistency of estimator (1). The given level of uncertainty V is determined using a desired accuracy and the calculated value of the ingredient load under the assumption of normal distribution of statistic (1).

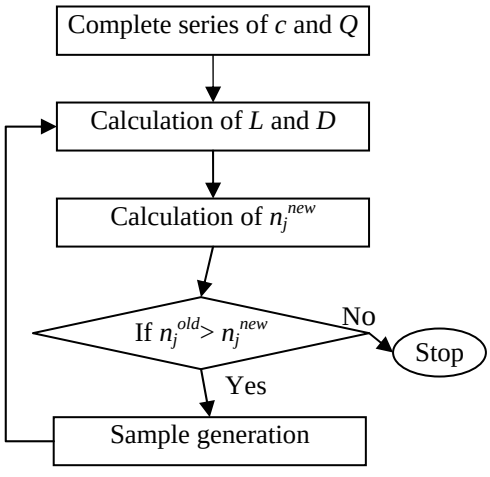


Figure 2. Determination of sampling frequency.

Problem (2) can be solved numerically according to an algorithm utilizing the Lagrange multiplier method and presented in Figure 2. Initially, the algorithm requires the most complete series of concentrations of an investigated water quality ingredient. Such a series can be obtained via pilot sampling or interpolation in time between observed values. To improve robustness of the algorithm, at each iteration one hundred random samples were made and the average number of observations per each stratum was calculated.

The proposed algorithm was used in a case study to determine the number of observations per year in order to estimate annual load of major ions at the cross-section Vyatskiye Polyany of the Vyatka River. The Vyatka River is a large Eastern-European river with a length of 1,370 km and a watershed area of 129,000 km^2. The

selected cross-section is characterized by annual water discharge of about 22.6 km^3. The year 1949 with unimodal type of hydrograph and sharp rising and falling limbs for spring-summer high flow events was selected. The annual loads of chloride ions, hydrocarbonate ions and total dissolved solids (TDS) were determined as 63,907 t, 2,465,922 t and 3,703,000 t, respectively.

Table 1. Frequency of observations guaranteeing a 5% uncertainty

Ingredient	Number of samples per stratum				Total per year
	I	II	III	IV	
TDS	3	17	2	4	26
Cl	10	57	28	12	107
HCO_3	3	21	2	7	33

The stratification of annual series of concentrations and water discharges was implemented according to main hydrological seasons. Stratum I represented winter low water events. Stratum II combined low parts of rising and falling limbs of the hydrograph. Stratum III included pick discharges and upper parts of rising and falling limbs of the hydrograph. Stratum IV corresponded to summer-fall low water events. Sampling frequencies for estimation of the annual loads of the selected water quality indicators with a 5% uncertainty are presented in Table 1.

6. DISCUSSION AND CONCLUSIONS

The investigated case study showed that the efficiency of formula (1) depends not only on the total number of concentrations available, but on their distribution among the temporal strata. Stratification is an important factor for the reduction of the required number of observations. In the presented case study, the stratification was implemented according to hydrological regime, not taking into account the variance of selected water quality indicators that presumably makes easier sampling recommendations. Such stratification gives lower values of the water discharge variance, but dates for each stratum vary from year to year making recommendations for sample collection still hard to follow. Temporal stratification common for several subsequent years is preferable, but it results in higher numbers of required observations.

Suggested monitoring designs are obviously model-dependent. If ingredient loads were estimated based on another formula, neither of the designs would guarantee the desired level of uncertainty. The determined number of observations not only conforms with the estimator (1), but was obtained by utilizing some additional information, particularly, about the shape of the hydrograph and boundaries of the main hydrological seasons. The even distribution of observations over a year would significantly increase the number of observations that are necessary to achieve the desired level of uncertainty. The proposed algorithm assumes that the accurate values of at least daily water discharges are available. In reality, the series of water discharges are not always accurate and, moreover, are not maintained for many cross-sections. In the latter case, the values can be obtained via hydrodynamic simulations using more sophisticated models.

As a rule, concentrations of the investigated indicators are obtained from the same sample, but recommended frequencies of their observations vary up to four times for different ingredients. A temporal monitoring design which is common for all observed water ingredients at a given site and supporting evaluation of their annual load with the same level of uncertainty is hardly attainable due to financial and technical constraints. For tiered monitoring systems, different allowable levels of uncertainty can be recommended for water quality indicators according to their importance for a particular site.

ACKNOWLEDGEMENTS

The authors are grateful to anonymous reviewers for their thoughtful suggestions and helpful comments on the manuscript. Part of the research was done based on the data sets prepared in Hydrochemical Institute, the Russian Federation.

REFERENCES

Aulenbach, B.T., and R.P. Hooper, The composite method: An improved method for stream-water solute load estimation, *Hydrological Processes*, 20, 3029-3047, 2006.

Bodo, B., and T.E. Unny, Sampling strategies for mass-discharge estimation, *Journal of Environmental Engineering*, 109 (4), 812 – 829, 1983.

Campolongo, F., A. Saltelli, T. Sørensen and S. Tarantola, Hitchhiker's guide to sensitivity analysis. In A. Saltelli, K. Chan and E.M. Scott (eds.), *Sensitivity Analysis*, Wiley, Chichester, 15-47, 2000.

Cochran, W.G., *Sampling Techniques*, 2nd ed., Wiley, New York, 1963.

Erechtchoukova, M.G., and P.A. Khaiter, Data-model issues in environmental impact assessment, *The International Journal of Environmental, Cultural, Economic and Social Sustainability*, 3 (6), 149-156, 2007.

Erechtchoukova, M. G., Uncertainty transformation in ecological simulation models. In A. Zerger and R.M. Argent (eds.), *MODSIM 2005 International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand, 2477-2483, 2005.

Erechtchoukova, M.G., and V.V. Tsirkunov, Determination of sampling strategy for calculation of mass-discharge in streams with a given accuracy. In *Proc. of the Conf. Problems of Surface Hydrology, February, 1987*, Gidrometeoizdat, Leningrad, 171-189, 1989.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney and J.G. Arnold, Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Transactions of the ASABE*, 49 (3), 689-701, 2006.

Hooper, R.P, B.T. Aulenbach and V.J. Kelly, The National Stream Quality Accounting Network: A flux-based approach to monitoring the water quality of large rivers, *Hydrological Processes*, 15, 1089-1106, 2001.

Jørgensen, S.E., and G. Bendoricchio, *Fundamentals of ecological modelling*, Elsevier, Amsterdam, 2001.

Preston, S.D., V.J. Bierman, Jr. and S.E. Silliman, An evaluation of methods for the estimation of tributary mass loads, *Water Resources Research*, 25(6), 1379-1389, 1989.

Richardson, J. R., and C. W. Berish, Data and information issues in modelling for resource management decision making: communication is the key. In V.H. Dale (ed.), *Ecological modelling for resource management*, Springer, New York, 167-179, 2003.

Shrestha, S, F. Kazama and L.T.H. Newman, A framework for estimating pollutant export coefficients from long-term in-stream water quality monitoring data, *Environmental Modelling and Software*, 23, 182-194, 2008.

Snowling, S.D., and J.R. Kramer, Evaluating modelling uncertainty for model selection, *Ecological Modelling*, 138, 17-30, 2001.

Straškraba, M., and A.H. Gnauck, *Freshwater ecosystems: modelling and simulation*, Elsevier, Amsterdam, 1985.

Strobol, R.O., P.D. Robillard, R.D. Shannon, R.L. Day and A.J. McDonnell, A water quality monitoring network design methodology for the selection of critical sampling points: Part I, *Environmental Monitoring and Assessment*, 112, 137-158, 2006.

Urban, D., Using model analysis to design monitoring programs for landscape management and impact assessment, *Ecological Applications*, 10(6), 1820-1832, 2000.

US Environmental Protection Agency (USEPA), Elements of a state water monitoring and assessment program (EPA 841-B-03-003), online URL http://www.epa.gov/owow/monitoring/elements/index.html, 2003

van Nes, E. H., and M. Scheffer, A strategy to improve contribution of complex simulation models to ecological theory, *Ecological Modelling*, 185, 153-164, 2005.

Water Quality Task Group (WQTG), A Canada-wide framework for water quality monitoring. PN 1369, online URL http://www.ccme.ca/assets/pdf/wqm_framework_1.0_e_web.pdf, 2006.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Multilinear Diffusion Analogy Model for Real-Time Streamflow Routing

Muthiah Perumal[a] **, Tommaso Moramarco**[b]**, Bhabagrahi Sahoo**[c] **and Silvia Barbetta**[b]

[a] *Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee-247 667, India (p_erumal@yahoo.com)*

[b] *National Research Council, IRPI, Via Madonna Alta 126, 06128 Perugia, Italy (t.moramarco@irpi.cnr.it, s.barbetta@irpi.cnr.it)*

[c] *National Academy of Agricultural Research Management (ICAR), Hyderabad, India; Formerly at Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee-247 667, India (bsahoo2003@yahoo.com)*

Abstract: A multilinear routing method using diffusion analogy model as a sub-model is proposed in this study. The diffusion analogy model can be considered as the next level of approximation to the full Saint-Venant equations. The applicability of the method is first tested by simulating the outflow hydrographs generated by routing the hypothetical inflow hydrographs in hypothetical channel reaches using the Saint-Venant equations and, subsequently, by studying the flood wave propagation in a 23 km reach of Tiber River in Central Italy. The results demonstrate the suitability of the diffusion analogy model as a sub-model for its use in the multilinear routing method for real-time applications.

Keywords: Diffusion; Multilinear; Routing; Streamflow; Modelling.

1. INTRODUCTION

Flood routing problems are solved using the hydraulic and hydrological methods. However, the hydraulic based methods, which generally use the full Saint-Venant equations, are of limited use primarily due to non-availability of topographical and hydrological inputs required at smaller spatial scales. Further, the computational limitations of the numerical schemes adopted in the solution procedure also restrict the use of the full Saint-Venant equations for many practical problems such as flood forecasting. An alternative way of overcoming these data and computational problems is by using the simplified hydraulic routing methods, which are derived from the Saint-Venant equations, but at the same time, they are not data intensive. While many researchers [Cunge et al., 1980; Lai, 1986] favour the use of the full Saint-Venant equations in flood routing studies, various other researchers [NERC, 1975; Ferrick, 1985] argued for using the simplified routing methods. Two categories of the simplified routing methods are used in practice: 1) the linear simplified routing methods which use constant parameters for routing a given flood wave, and 2) the variable parameter simplified routing methods which use the model structure of the simplified routing methods, but the parameters varying at every routing time step. The variable parameter Muskingum-Cunge (VPMC) method and its variants [Ponce and Yevjevich, 1978; Ponce and Chaganti, 1994], the multilinear Muskingum method [Perumal, 1992] and the multilinear discrete cascade model [Perumal, 1994] are some of the available variable parameter simplified methods. While assessing the future developments of flood routing methods Fread [1981] and Lai [1986] opined that the simplified methods would continue to be used in practice, especially as components of precipitation-runoff basin models for routing overland and channel flows associated with the network of headwater streams.

The current study focuses on the development of an improved multilinear routing method using diffusion analogy model [Hayami, 1951] as the submodel. The diffusion analogy model can be considered as the next level of approximation to full Saint-Venant equations [Fread, 1981]. This method is more appropriate for its use in real-time flood forecasting schemes. The applicability of the method is first tested by simulating the outflow hydrographs generated by routing the hypothetical inflow hydrographs in hypothetical channel reaches using the Saint-Venant equations and, subsequently, testing the field applicability of the method by studying the flood wave propagation in a 23 km reach of Tiber River in Central Italy.

2. MULTILINEAR ROUTING MODELS

The linear flood routing methods such as the Muskingum and Kalinin-Milyukov methods are widely used as the basic models in real-time flood forecasting schemes. However, flood waves are inherently nonlinear in nature and, therefore, it is generally desirable to use nonlinear models for studying flood wave movement in channels. But the nonlinear models are more difficult to apply in the field than the linear models. The convenience of linear analysis can still be used for modelling the nonlinear hydrological processes by working within the limitation imposed by its assumption. One simple method by which the nonlinearity of the flood routing process could be taken care off is by using the multilinear modelling approach advocated by Keefer and McQuivey [1974] and followed by Becker [1976], and Becker and Kundzewicz [1987]. This approach was further extended by employing the Muskingum method [Perumal, 1992] and the discrete cascade model [Perumal, 1994] as the sub-models of the multilinear routing methods. All these multilinear routing methods amply demonstrate the appropriateness of this approach by closely reproducing the solutions of the full Saint-Venant equations in off-line mode, which were considered as the benchmark solutions.

Moramarco and Singh [1999] investigated the use of diffusion analogy model as a sub-model of two multilinear modelling approaches. The principle employed in these approaches is to distinguish on the input hydrograph, different input components, each of which is subsequently routed through the diffusion analogy sub-model. These two modelling approaches differ from each other by inflow portioning approach, depending on whether the inflow is divided up either horizontally or vertically, i.e., the horizontal distinctions represent the different zones of discharge, whereas the vertical distinctions are introduced at fixed times. The former approach has been used by the early investigators of the multilinear routing methods [Keefer and McQuivey, 1974; Becker, 1976; Becker and Kundzewicz, 1987]. Becker and Kundzewicz [1987] introduced the vertical portioning of the inflow hydrograph into different time zones with each zone characterized by a unique instantaneous response function defined by the average flow characteristics of the respective zones. This scheme of portioning is alternatively known as the time distribution scheme. Perumal [1992, 1994] and Moramarco and Singh [1999] restricted the time distribution scheme of the inflow hydrograph to that of the routing time interval, thus, accounting for the nonlinear feature of the flood wave characteristics in a more efficient manner. Moramarco and Singh [1999] directly employed the response function of the diffusion analogy submodel for convoluting with the inflow hydrograph. However, the theoretically correct way of obtaining the Δt-hour response function to be used for convolution with the given inflow hydrograph ordinate is the one obtained by convoluting a unit inflow ordinate with the instantaneous unit hydrograph (IUH) of the diffusion analogy model. However, when the routing interval Δt is small, say an hour, the difference between the IUH and the corresponding Δt-hour response function may not be significant, implying that one may directly employ the IUH for convolution with the given inflow hydrograph ordinates.

3. MULTILINEAR DIFFUSION ANALOGY ROUTING MODEL

The routed outflow at any time in response to the given inflow hydrograph observed till the same time may be expressed by the discrete convolution approach as:

$$Q(i\Delta t) = I_b + \sum_{i=1}^{j} h(i\Delta t)\left(I\left(\left(j-i-1\right)\Delta t\right) - I_b\right) \qquad (1)$$

where I_b is the initial flow in the considered river reach; $j\Delta t$ denotes the time t in discrete intervals with j=1 corresponding to the first ordinate; and $h(i\Delta t)$ denotes the ordinate of the Δt-hour response function of the diffusion analogy sub-model at time $i\Delta t$. The Δt-hour response function of the diffusion analogy sub-model corresponding to the unit ordinate input is expressed using the convolution approach as:

$$h(t) = \int_{t-\Delta t}^{t} u(\tau) d\tau \tag{2}$$

where $u(t)$ is the IUH of the diffusion analogy model [Hayami, 1951] expressed as:

$$u(t) = \frac{x}{\sqrt{4\pi D t^3}} \exp\left[\frac{-(x-ct)^2}{4Dt}\right] \tag{3}$$

where x is the routing reach length, c is the wave celerity in m/s, and D is the hydraulic diffusivity in m^2/s expressed as:

$$D = \frac{Q_0}{2S_0 B} \tag{4}$$

where Q_0 is the reference discharge at time level t, S_0 is the channel bed slope, and B is the water surface width corresponding to Q_0.

The wave celerity c is expressed as:

$$c = \left.\frac{dQ}{dA}\right|_{Q=Q_0} = \frac{1}{B}\left.\frac{dQ}{dy}\right|_{Q=Q_0} \tag{5}$$

The reference discharge used in the estimation of the parameters c and D of the diffusion analogy model is estimated as [Perumal, 1992, 1994]:

$$Q_O = Q_b + a(I(t) - Q_b) \tag{6}$$

where a is an empirical coefficient with 0< a <1.

Though the diffusion analogy sub-model used in the proposed method has a larger range of applicability limit than the Muskingum method used as the sub-model in the multilinear Muskingum (MM) method [Perumal, 1992] and the discrete cascade sub-model [Perumal, 1994] used in the multilinear discrete cascade routing method, its application to nonlinear modelling on the same lines as these multilinear models is difficult to achieve. The difficulty lies with the establishment of the pulse response, required for convolution with the inflow hydrograph, as the closed form integration of the IUH given by equation (2) cannot be obtained. Although there are a number of built-in functions available in the modern day personal computers for a close estimation of the pulse response function, the use of such a solution approach is not mathematically elegant.

To overcome this problem, the use of the Adam-Moulton multi-step numerical integration method [Atkinson, 2003] is used for the numerical integration of equation (2) and the expression of pulse response for $t \geq 2\Delta t$ is expressed as:

$$h(i\Delta t) = \frac{\Delta t}{12}\left(8u_{t-\Delta t} + 5u_t - u_{t-2\Delta t}\right) \tag{7}$$

where Δt is the duration of the pulse response or routing time interval; and $h(t)$ is the ordinate of the IUH at time t.

The pulse response at time Δt is expressed as:

$$h(\Delta t) = \frac{u(0) + u(\Delta t)}{2} = \frac{u(\Delta t)}{2} \tag{8}$$

where $u(0) = 0$.

4. APPLICATION

The proposed multilinear diffusion analogy routing method was applied first for simulating a number of hypothetical flood hydrographs obtained by routing hypothetical inflow hydrographs in rectangular and trapezoidal channel reaches, with no lateral inflow within the routing reach. The inflow hydrograph, defined by a mathematical function, is routed in a given channel reach for a specified distance using the proposed method and is compared with the corresponding solutions obtained using the Saint-Venant equations. The form of the inflow hydrograph used in the numerical experiments of the present study is expressed as:

$$I(t) = I_b + \left(I_p - I_b\right)\left(\frac{t}{t_p}\right)^{\frac{1}{(\gamma-1)}} \exp\left[\frac{\left(1-\frac{t}{t_p}\right)}{(\gamma-1)}\right] \tag{9}$$

where I_b is the initial steady flow in the reach (100 m^3/s), I_p is the peak inflow, t_p is the time to peak, and γ is the skewness factor.

Different inflow hydrographs were routed in the considered rectangular channel with a width of 100 m and in the symmetrical trapezoidal channel with a side slope of *z*: horizontal to 1: vertical and with a bottom width of 100 m. The details of channel configurations and inflow hydrographs used in the routing experiments of the proposed study are given in Table 1. A total of 3360 numerical experiments each with unique combination of inflow hydrograph and channel configuration formed by the parameters given in Table 1 were made for each of the considered rectangular and trapezoidal channels.

For all the routing experiments, the value of coefficient a used in the equation (6) for estimating the reference discharge was taken as 0.30 for both the rectangular and trapezoidal channels. This best value of a was arrived at by a trial and error approach using the values between zero and unity for simulating few cases of benchmark solutions, i.e., the Saint-Venant solutions. The Benchmark solutions were obtained using explicit numerical methods.

The method was also tested for field applications by studying the recorded flood events in a 23 km stretch of the reach of Tiber River in central Italy between Ponte Felcino and Torgiano sections. The cross-section of the reach is approximated as a rectangular channel section with a width of 45 km, and the reach is characterized by a Manning's roughness coefficient $n = 0.04$ and the bed slope, S_0 of the reach is estimated as 0.0014. Out of the five events simulated, three events (1985, 1986 and 1992) may be considered as having negligible lateral flow $(<1\%)$ and the other two events (1991 and 1982) have a small loss and gain of volume as -5.15% - and 12.28%, respectively. The best value of the coefficient a for simulating all these five events was arrived at by trial and error approach starting with a value of $a = 0.30$, and it was found that all the five events could be closely reproduced for a value of $a = 0.24$.

5. PERFORMANCE CRITERIA

The performance of the method in reproducing the benchmark solutions is evaluated by comparing its solution with the corresponding solution of the Saint-Venant equations based on the following criteria:

1) Accuracy of reproduction of the hydrograph shape and size given by the measure of variance explained [Nash and Sutcliffe, 1970] expressed as:

$$VAREX\ (\text{in}\ \%) = \left(1 - \frac{\sum_{i=1}^{N} (Q_{oi} - Q_{ci})^2}{\sum_{i=1}^{N} (Q_{oi} - \overline{Q}_{oi})^2} \right) \times 100 \qquad (10)$$

where Q_{oi} is the ith ordinate of the benchmark discharge hydrograph ordinate at the outlet of the reach, Q_{ci} is the ith ordinate of the routed or computed discharge hydrograph by the proposed model and N is the total number of discharge hydrograph ordinates to be simulated.

2) The accuracy of reproduction of the peak discharge of the benchmark solution is estimated by the following measure expressed as:

$$q_{per}(in\%) = \left(\frac{q_{pc}}{q_{po}} - 1 \right) \times 100 \qquad (11)$$

where q_{pc} is the routed or computed peak of the discharge hydrograph at the outlet by the proposed method, and q_{po} is the corresponding benchmark hydrograph. The positive value of q_{per} indicates overestimation of the benchmark peak and the negative value indicates its underestimation.

Table 1. Combination of channel and flow characteristics used in the routing experiments by the multilinear diffusion analogy routing method.

Characteristics		Values
Skewness factor, γ	:	1.05, 1.15, 1.25, 1.50
Channel bed slope, S_o	:	0.0002, 0.0005, 0.0008, 0.001, 0.002, 0.003, 0.005, 0.01
Manning's roughness, n	:	0.01, 0.02, 0.03, 0.04, 0.05
Initial discharge, I_b (m^3/s)	:	100.0
Peak discharge, I_p (m^3/s)	:	1000; 2500; 5000; 7500; 10,000; 12,500; 15,000
Time to peak discharge, t_p (h)	:	10.0, 15.0, 20.0
Channel bottom width, b_m (m)	:	100.0
z		0.0; 1.0

3) The accuracy of reproduction of the time to peak of the benchmark solution is estimated by the following measure expressed as:

$$t_{per}(in\%) = \left(\frac{t_{pc}}{t_{po}} - 1 \right) \times 100 \qquad (12)$$

where t_{pc} is the time corresponding to routed or computed peak of the discharge hydrograph at the outlet, and t_{po} is the time corresponding to peak of the benchmark discharge hydrograph at the outlet.

As the solution of the method is obtained by convolution approach, the volume of the routed hydrograph is always conserved provided the duration of convolution exceeds the memory of the system.

6. DISCUSSION OF RESULTS

A reach length of 40 km was used in all the experimental runs to arrive at the outflow hydrographs using the proposed multilinear diffusion analogy routing method and these solutions are compared with the corresponding benchmark solutions of the Saint-Venant equations. Figures 1 and 2, respectively, show the typical routing results obtained using the proposed method when a given inflow hydrograph was routed in the considered rectangular and trapezoidal channels for a reach length of 40 km and the corresponding benchmarks solutions are also shown therein. Figures 3 and 4 show the variation of various measures of the accuracy criteria used in this study with reference to the variable of maximum of $(1/S_0)\partial y/\partial x$ estimated for each run. The variable $[(1/S_0)\partial y/\partial x]_{\max}$ has been considered to represent the order of diffusivity associated with each run, and it is expected that higher its magnitude the lower would be the accuracy of the proposed method in reproducing the respective benchmark solution closely.

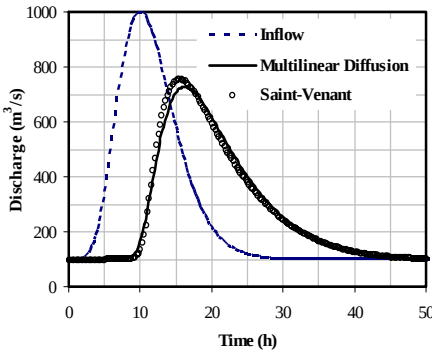


Figure 1. Hydrograph reproductions by the proposed multilinear diffusion and the Saint-Venant methods in rectangular channel (S_o=0.0002, n=0.04, I_b=100 m^3/s, I_p=1000 m^3/s, γ = 1.15)

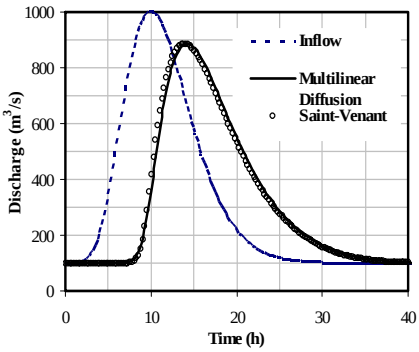


Figure 2. Hydrograph reproductions by the proposed multilinear diffusion and the Saint-Venant methods in a trapezoidal channel (S_o=0.0002, n=0.02, z=1.0, I_b=100 m^3/s, I_p=1000 m^3/s, γ =1.15)

It is inferred from the typical routing results presented in Figures 1 and 2 that the proposed method has the capability in reproducing the benchmark solutions closely. This inference is further confirmed from the results shown in Figures 3 and 4 depicting the variability of the three measures of the performance criteria, *viz.*, the *VAREX*, $q_{per}(in\%)$, and $t_{per}(in\%)$ estimated for the two sets of 3360 runs, each made for the considered rectangular and trapezoidal channel reaches, against the diffusivity measure $[(1/S_0)\partial y/\partial x]_{\max}$ estimated for each run. It is seen from Figures 3a and 4a that the variance explained for almost all the runs made in these channels is greater than 98.5 % with few runs deviating away from this general trend when $[(1/S_0)\partial y/\partial x]_{\max}$ is nearer to unity, i.e., when $[(1/S_0)\partial y/\partial x]_{\max}$ is maximum. There are about ten runs amongst the 3360 runs made for each of the considered rectangular and trapezoidal channels that could not successfully compute the benchmark solutions due to numerical problems, and they have not been considered in Figures 3 and 4. Figures 3b and 4b bring out the variability of the measures of the errors of peak discharge estimates within the range of ±5% for almost all the 3360 runs in each of these channels. It is seen that there exists a trend in the variability of this error with positive values associated with lower order diffusivity and the negative values associated with higher order diffusivity. Figures 3c and 4c depicts the measure of errors in the time to peak which almost lie within the band of ±2.5% and these errors are almost evenly distributed about the zero error line. Based on the inferences made from this study, one may consider that the proposed multilinear diffusion analogy routing method may be used as the improved basic model in flood forecasting models. The proposed model structure maintains the simplicity of the linear routing scheme which is desirable for adopting in a flood forecasting scheme and, yet, is capable of accounting the nonlinearity of the routing process in an efficient manner.

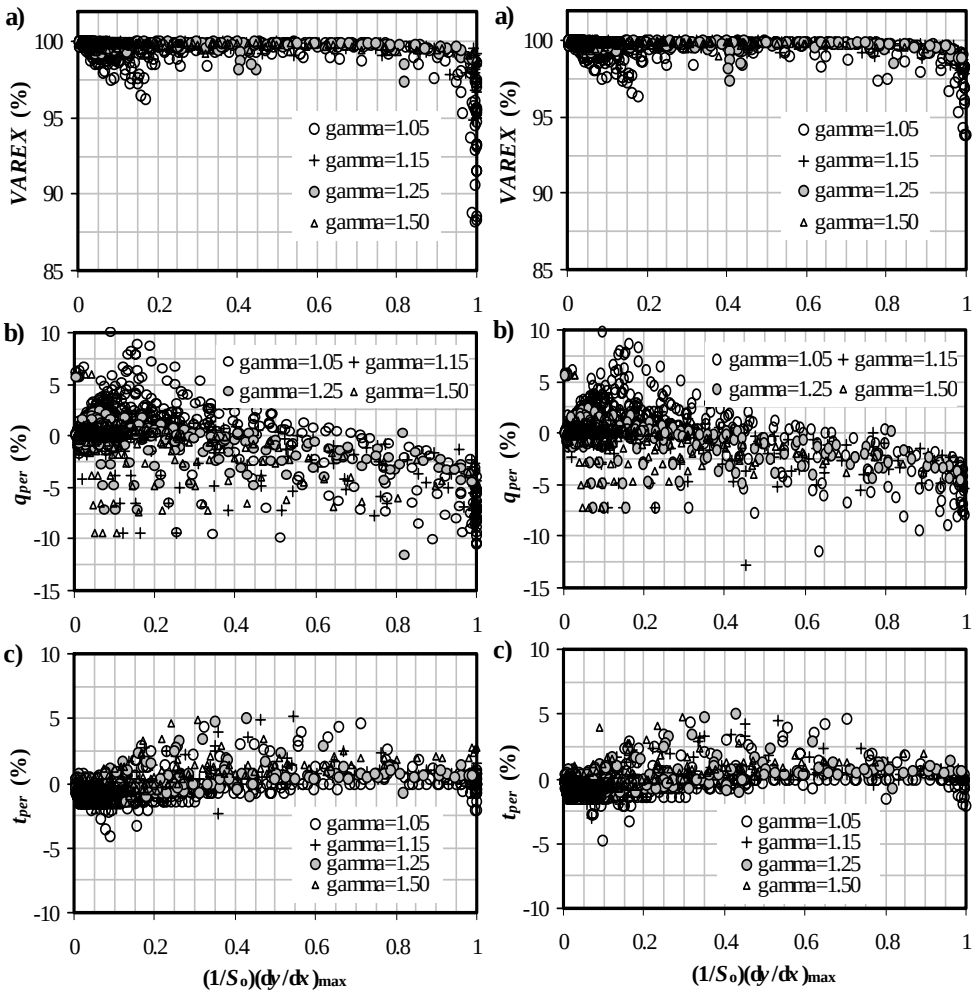


Figure 3. Error estimates of the multilinear diffusion model in comparison with the Saint-Venant's solutions for rectangular channel.

Figure 4. Error estimates of the multilinear diffusion model in comparison with the Saint-Venant's solutions for trapezoidal channel.

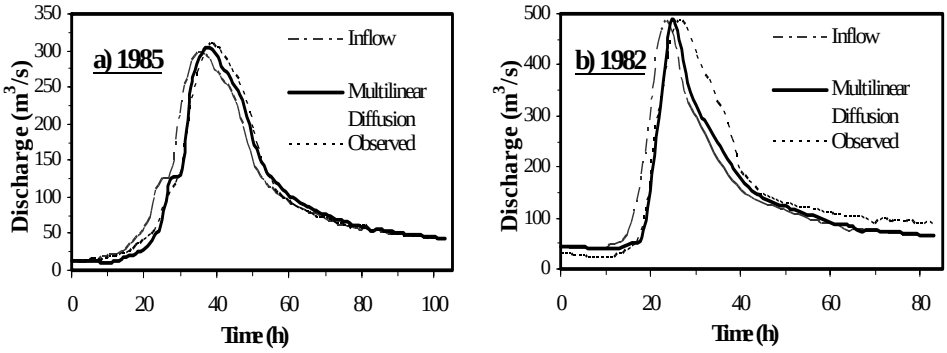


Figure 5. Typical flood routing simulations using the proposed method (Between Ponte Felcino and Torgiano sections of the Tiber River in Central Italy).

It is inferred from the simulation of the five flood events in the reach between the Ponte Felcino and Torgiano sections of the Tiber River in Central Italy that three events (1992, 1985 and 1986) with negligible lateral flow could be reproduced very closely with variance explained criteria $> 97.8\%$. The other two events (1991 and 1982) could be reproduced in best possible manner with variance explained being 94.0% and 92.0% , respectively and these events were subjected to -5.0% and 12.0% lateral flow. Two of the typical reproductions of the field events are shown in Figure 5. It may be seen that the method was able to reproduce the rising limb of the observed hydrograph of the 1982 event quite closely, although there was significant lateral flow contribution within the reach.

CONCLUSIONS

The study demonstrates that the proposed multilinear routing method using the diffusion analogy response function as the sub-model is capable of simulating the benchmark solutions (i.e., the Saint-Venant solutions) closely for a wide range of diffusivity exhibited in the routing process. The method was also tested for simulating few field events of the Tiber River in the reach between Ponte Felcino and Torgiano, and the results obtained indicates the suitability of the method for field application. This method maintains the simplicity of the linear routing scheme which is desirable for adoption in a flood forecasting algorithm and, yet is capable of accounting the nonlinearity of the routing process in an efficient manner. A further investigation of this method is necessary for studying a number of field events for assessing the suitability of this method before recommending it for field application.

ACKNOWLEDGEMENT

The authors wish to thank Mr.Ravindra Kale, Research Scholar, Department of Hydrology, Indian Institute of Technology Roorkee for word processing this paper.

REFERENCES

Atkinson K., Elementary numerical analysis, John Wiley & Sons, Second edition, 2003.

Becker A., Simulation of nonlinear flow systems by combining linear models. IAHS Publication 116, IAHS, pp.135-142, 1976.

Becker A and Kundzewicz Z.W., Nonlinear flood routing with multilinear models. Water Resour. Res., 23: 1043-1048, 1987.

Cunge, J. A., Holly, Jr. F. M., and Verwey, A., *Practical Aspects of Computational River Hydraulics*, Pitman Publishing Limited, London, ISBN: 0273084429, 1980.

Ferrick, M. G., Analysis of river wave types, *Water Resour. Res.*, *21*(2): 209–220, 1985.

Fread, D. L., Flood routing: a synopsis of past, present, and future capability, in *Proc. Int. Symp. Rainfall-Runoff Modeling,* V. P. Singh (Ed.), Mississippi, USA,521−541, 1981.

Hayami S., On the propagation of flood waves, Disaster Prev.Res. Inst.Bull., 1, 39-67, 1951.

Keefer T.N., and McQuivey R.S., Multiple linearization flow routing model. J.Hydraul.Div., ASCE, 100(HY7): 1031-1046, 1974.

Lai, C., Numerical modeling of unsteady open channel flow, *Advances in Hydroscience, 14*, Academic Press, USA, 161–333, 1986.

Moramarco T., Singh V.P., Linear routing in channel networks, In: V.P. Singh, I.W. Seo and J.H. Sonu (Eds.), Hydrologic Modeling, Water Resources Publications, 153-160, 1999.

Nash J.E., and Sutcliffe J. V., River flow-forecasting through conceptual models Part I- A discussion of principles, J. Hydrol., 10: 82-290, 1970.

Natural Environment Research Council (NERC), Flood Routing Studies, in *Flood Studies Report*, Vol. III, Institute of Hydrology, Wallingford, UK, 1975.

Perumal M., Multilinear Muskingum flood routing method, J. Hydrol., 133: 259-272, 1992.

Perumal M., Multilinear discrete cascade model for channel routing, J. Hydrol., 158: 135-150, 1994.

Ponce, V. M., and V. Yevjevich, Muskingum-Cunge method with variable parameters, *J. Hydraul. Div., ASCE, 104*(HY12): 1663–1667, 1978.

Ponce, V. M., and Chaganti, P. V., Variable parameter Muskingum–Cunge method revisited, *J. Hydrol., 162*(3–4): 433–439, 1994.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Influence of the uncertainties of monitoring data on model calibration and evaluation

A. Ullrich[a], M. Volk[a], and G. Schmidt[b]

[a] *Helmholtz Centre for Environmental Research - UFZ Department of Landscape Ecology, Permoserstraße 15, D-04318 Leipzig, Germany (antje.ullrich@yahoo.com; martin.volk@ufz.de)*

[b] *Martin-Luther-Universitity, Halle-Wittenberg, Department of Geosciences, Von-Seckendorff-Platz 3-4, D-06120 Halle (Saale), Germany (gerd.schmidt@geo.uni-halle.de)*

Abstract: The model-based prediction of the impact of different land management on nutrient loading requires measured nutrient flux data. Thereby the accurate calibration and evaluation of the models need an adequate data base in form of monitoring data. Uncertainties in the monitoring data influence the calibration and thus the parameter settings which affect the modelling results. Hence, we compared three different time-based sampling strategies and four different load estimation methods for model calibration and compared the results. For our study we used the river basin model SWAT (**S**oil and **W**ater **A**ssessment **T**ool). Study area is the intensively used loess-dominated Parthe watershed (315 km²) in Central Germany.
Nitrate-N load estimation results differ considerable depending on sampling strategy, used load estimation method and period of interest. For study period the load estimation results for the daily composite data set have the lowest ranges (14% and 2% maximum deviation related to the mean value of all applied methods). In contrast estimation results for the sub-monthly and the monthly data set vary in greater ranges (between 25% and 52%). To show differences between sampling strategies we calculated the percentage deviation of mean load estimations of sub-monthly and monthly data sets related to the mean estimation value of composite data set. The maximum deviation of 82% occurs for the sub-monthly data set in 2000. This affects the model and leads to different parameter settings in model calibration and evaluation. Therefore we recommend both the implementation of optimised monitoring programs and the use of more than one load estimation method to describe the water quality situation in a better way and to establish a good calibration base for simulation models.

***Keywords**:* SWAT; modelling; water quality sampling; load estimation; model calibration.

1. INTRODUCTION

With this paper, we postulate the implementation of optimised and effective monitoring programs to support a sustainable water quality protection and to establish a good calibration base for simulation models. Simulation models are powerful tools to evaluate the impact of land management scenarios on water quantity and quality at the watershed scale [Chaplot et al., 2004; Behera and Panda, 2006; Bracmort et al., 2006; Pandey et al., 2005]. The results of such scenarios can be used for the development of efficient water quality management plans in river basin management. However several problems still exist when using models for spatially explicit simulation of the environmental impact of land management options and environmental measures. Accurate calibration and evaluation of the models need an adequate data base in form of monitoring data. Uncertainties in the monitoring data influence the calibration [Harmel et al., 2006a] and thus the parameter settings which affect the modelling results.

There are two main aspects of the influence of monitoring data uncertainty on model calibration and evaluation: i) sampling frequency of water quality data and ii) load estimation method. For water quality data sampling the point of time and frequency at

which discrete water samples are collected are important to reflect temporal changes [Harmel et al., 2006b; Tate et al., 1999]. Therefore, sampling strategies should include frequent samples taken for the entire range of observed flow to characterise water quality. However, due to financial and personnel constraints, the number of samples that can be collected is often limited [Harmel et al., 2002, 2003; Harmel and King, 2005; King and Harmel, 2003; Robertson and Roerish, 1999; Tate et al., 1999]. In Germany the basic water quality monitoring is organised by the Federal States. The Regional Authorities for Environment are responsible for water sampling and water quality monitoring. Usually, discrete samples are collected 13 to 24 times per year on a regular time basis [SMUL, 2005]. This is a commonly used sampling strategy for large rivers also in other countries [e.g. Robertson and Roerish, 1999; Robertson, 2003]. However, during storm water events, water levels and pollutant concentrations can change very rapidly especially in small streams and, however, periodic sampling does not adequately describe the rapid changes in water quality [Robertson and Roerish, 1999]. Therefore, small streams (such as the study area) need more intensive sampling strategies to achieve precise and accurate load estimations [Harmel et al., 2003; Harmel and King, 2005]. The accuracy of load estimates depends on the sampling method, sampling frequency, load estimation methodology, and the duration and period of the estimation [Littlewood, 1995; Littlewood and Marsh, 2005]. Inaccuracy or imprecision of load estimates limits its use in environmental assessment and management, trend detection, and watershed simulation [Littlewood and Marsh, 2005]. Therefore, the accurate load estimation and water quality characterization are important to accomplish the objectives of alternative management plans. Diverse studies have dealt with comparative analysis concerning either the use of different sampling strategies or/and the use of different methods for load estimation for different constitutes [e.g. Ferguson, 1987; Izuno et al., 1998, Keller et al., 1997; Littlewood, 1995; Robertson and Roerish; 1999, Stone et al., 2000; Swistock et al., 1997; Walling and Webb, 1981; Webb et al., 1997].

The overall goal of our study was to investigate the influence of uncertainty of monitoring data on model calibration, model parameter settings and model evaluation. Hence, we compared different load estimation methods based on three different time-based sampling strategies to estimate load data for model calibration. For our study we used the continuous-time river basin model SWAT (**S**oil and **W**ater **A**ssessment **T**ool). SWAT has been developed to predict the long-term impacts of land management measures on water, sediment and agricultural chemical yield in large complex watersheds with varying soils, land use and management conditions [Arnold and Fohrer, 2005; Behera and Panda, 2006]. Study area is the intensively used loess-dominated Parthe watershed (315 km²) in Central Germany. The investigated water quality parameter is nitrate-N.

2. MATERIAL AND METHODS

2.1 Study area

The Parthe watershed (study area) is located in the State of Saxony in Central Germany and drains an area of about 315 km² (Figure 1). It is a subbasin of the Weisse Elster catchment in the Elbe River system. The topography of the area is flat with altitudes between 106 m and 230 m above sea level. The mean annual precipitation ranges about 590 mm to 640 mm (1981-2000). The Parthe is a typical lowland river. The runoff dynamics are characterized by high flows in spring due to snow melt and rainfall and long periods of low flows in summer with occasional storm flow events. Mean long-term flow rate is 0.9 m³/s.

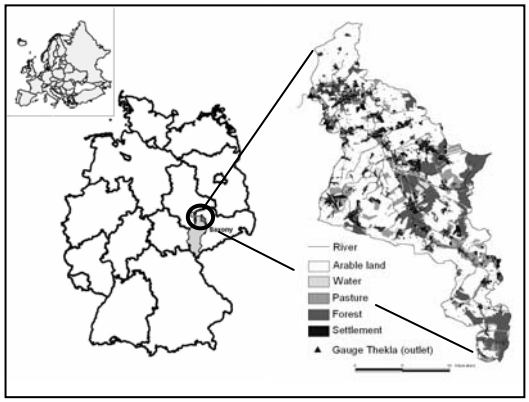


Figure 1. Location of the study area in Germany

2.2 Model description

SWAT is considered as one of the most suitable models to predict the long-term impacts of land management measures on water, sediment and agricultural chemical yield (nutrient loss) in large complex watersheds with varying soils, land use and management conditions [Arnold and Fohrer, 2005; Behera and Panda, 2006; Gassmann et al., 2007]. SWAT is a physically based, conceptual, continuous-time river basin model with spatial distributed parameters operating on a daily time step. It is not designed to simulate detailed, single-event flood routing [Neitsch et al., 2002]. The SWAT model integrates all relevant eco-hydrological processes including water flow, nutrient transport and turn-over, vegetation growth, land use and water management at the subbasin scale. Subbasins are further disaggregated into classes of Hydrological Response Units (HRU), whereby each unique combination of the underlying geographical maps (soils, land use, etc.) forms one class. The water balance for each HRU is represented by the four storages snow, soil profile, shallow aquifer and deep aquifer. The soil profile can be sub-divided up to ten soil layers. Soil water processes include evaporation, surface runoff, infiltration, plant uptake, lateral flow and percolation to lower layers [Arnold and Allen, 1996; Neitsch et al., 2002]. The surface runoff from daily rainfall is estimated with a modification of SCS curve number method [Arnold and Allen, 1996; Neitsch et al., 2002].
In Swat, nitrogen movement and transformation are simulated as a function of nitrogen cycle. SWAT simulates five different pools of nitrogen in the soils; two inorganic and three organic. Nitrogen is added to the soil by fertilizer, manure or residue application, fixation by bacteria, and rain. Nitrogen losses occur by plant uptake surface runoff in the solution and the eroded sediment [Neitsch et al., 2002].

2.3 Input data and model calibration (water cycle)

The applied input data sets are listed following: digital elevation model (30x30 m); several precipitation stations (daily sums, Environmental Operating Company (UBG)); one climate station (daily values, UBG), land use (habitat cartography), 1:10,000, Statistical Office of the Free State of Saxony, aggregated to five classes [arable land, pasture, forest, water, settlement]); waste water treatment plants, State Agency for Environment (StUFA); soil mapping (1968) (1:25,000, M. Thomas-Lauckner, (unpublished)); several crop rotations (conventional managed), including applied tillage operations and fertiliser applications.

A period of three years was used either for model calibration (1994 to 1996) and evaluation (1998 to 2000). First the rates of surface flow, lateral flow and baseflow were adjusted. Basically the following model parameters were adjusted. The curve number (CN2 - lowered), soil evaporation compensation factor (ESCO – 0.85 to 0.9; spatially adjusted), plant uptake compensation factor (EPCO – 0.8), effective hydraulic conductivity in channel alluvium (CH_K1/2 – spatially adjusted), Manning´s roughness coefficient for overland flow (OV_N – increased for agricultural land and pasture), Manning´s roughness coefficient for main and tributary channel (CH_N1/2 – 0.35), surface runoff lag coefficient (SURLAG – 1.0), maximum canopy storage (CANMX – 4.0 for arable land, 5.0 for pasture and 9.5 for forest), groundwater delay times (GW_DELAY – 150 to 350; spatially adjusted), baseflow alpha factor (ALPHA_BF – 0.065), threshold depth of water in the shallow aquifer required for return flow to occur (GWQMN – 0.0), groundwater “revap” coefficient (GWREVAP – 0.03), threshold depth of water in the shallow aquifer for “revap” or percolation to the deep aquifer to occur (REVAPMN – 0.05).

To evaluate the model predictions the following goodness-of-fit parameters were used: mean discharge, standard deviation (STD), coefficient of determination (R^2; indicates the quality of relationship between observed and predicted results); Nash-Sutcliff efficiency (NSE; indicates the model efficiency [Nash and Suttcliffe, 1970]) and prediction efficiency (PE; indicate the model’s ability to describe the probability distribution of the observed results) (see Table 1). The measured and predicted monthly discharge values matches quite well at gauge Thekla. The model efficiency dropped only in very dry and in very wet years of prediction.

Table 1. Goodness-of-fit parameters for calibration of monthly predicted discharge values (watershed outlet-gauge Thekla)

	1994-1996		**1998-2000**	
	observed	**predicted**	**observed**	**predicted**
Mean discharge m³/s	1.19	1.16	1.00	1.35
STD	0.80		0.80	
R^2	0.75		0.80	
NSE	0.72		0.56	
PE	0.88		0.72	

The Parthe system is impaired by several small scale human activities which are difficult to simulate. Furthermore, most sections of the river channel are heavily regulated. Therefore, we consider the results as satisfactory.

2.4 Water quality monitoring

The samples for water quality investigation were collected using three different time-based sampling strategies: periodic grab samples taken on regular time intervals corresponding to biweekly or monthly intervals and composite samples. The time span of daily composite sampling strategy was limited from 2000 to 2001. The periodic grab samples were taken at random times within the day and not adjusted to represent selected flow rates. The composite samples are isochronous (40 ml/hour), not flow-weighted, and stored in a single collection bottle that represents the daily mean concentration. None of the three sampling strategies include flow-stratified sampling.

2.5 Load estimation

Most of the commonly applied load estimation methods differ with respect to the sampling strategy and the hydrological characteristics of the investigated streams and constituents. For the presented study, we chose four equations, which were applied to the data sets of the different sampling strategies:

$$L_{(1)} = F\frac{\overline{Q}_{sd}}{Q_{sd}}\left(\frac{1}{N}\sum_{N=1}^{N} cQ_{sd}\right) \quad (1)$$

$$L_{(2)} = F\frac{1}{N}\sum_{i=1}^{N} cQ \quad (2)$$

$$L_{(3)} = L_{(2)} + F\frac{1}{N}\sum_{i=1}^{N} r_{Qc}\sigma_Q\sigma_c \quad (3)$$

$$L_{(4)} = F\frac{1}{N}\sum_{i=1}^{N} cQ\,;\, T(Q) = aQ^b \quad (4)$$

with: L- load (t), F- factor to take account period of record, c- sample concentration (mg/l), Q- discharge at sample time (m³/s), N- number of samples, Q_{sd}- mean discharge for sampling day (m³/s), $\overline{Q}_{sd}$- mean flow (m³/s), r_{Qc}- correlation coefficient of Q and c, σ_Q- standard deviation of Q, σ_c- standard deviation of c, T- mean daily transport (t)

The first three equations are interpolation methods while the fourth equation is an extrapolation method. Equations (1) and (2) are the first and second choice of the OSPAR (Oslo-Paris) Convention [Littlewood, 1995, OSPAR, 2004]. Webb et al., [1997] found the methods (1) and (2) are quite accurate, but suffer from a high degree of imprecision. For

solute fluxes, they state that these methods produce generally more reliable estimates. For the equation (2), substantial systematic errors only occur with large variability of discharge and concentration within sampling period. Especially if there is no correlative relation between concentration and discharge, the method should lead to good load estimation results [Keller et al., 1997]. Equation (3) uses statistical values to quantify the variability of concentration and discharge. Equation (4) assumes a relationship between transport and discharge at the time of sampling [Webb et al., 1997]. It uses a correlative relation (rating curve) between these variables.

3. RESULTS AND DISCUSSION

3.1 Load estimation

The annual nitrate-N load estimation results have a wide range with respect to i) the water quality sampling strategy, ii) the applied load estimation methods and iii) the period of interest and its hydrological characteristics (see Table 2). For the study period, the load estimation results for the daily composite data set have the lowest ranges (14% and 2% maximum deviation related to the mean value of all applied methods). In contrast the estimation results for the sub-monthly and the monthly data set vary in greater ranges (between 25% and 52%). Therefore, the daily composite data set is assumed to give best results even if this sampling strategy has also its uncertainty. To show the differences between the different sampling strategies we also calculated the percentage deviation of mean load estimations of sub-monthly and monthly data sets related to the mean estimation value of composite data set. The deviation ranges from -15% for the monthly data set in 2001 up to a maximum deviation of 82% for the sub- monthly data set in 2000. That points out again the importance of the choice of the used sampling strategy as well as the randomness of measurements, especially within time-based discrete sampling strategies.

Table 2. Results of nitrate-N load estimation in tons (t) using different equations (**Max; **Min, ***percentage deviation of mean load estimation values of sampling strategies related to mean load estimation value of daily composite data set)*

	(1) [t]	(2) [t]	(3) [t]	(4) [t]	Mean [t]	Max range [t] (percentage deviation related to the mean value)	Percentage deviation***
Year	**2000**						
Composite		141.5	148.8*	128.8**	**139.7**	20.0 (14%)	
Sub-monthly	246.5	243.2	307.9*	219.4**	**254.3**	88.5 (35%)	82%
Monthly	196.2	152.1**	175.5	242.5*	**191.6**	90.4 (47%)	37%
Year	**2001**						
Composite		91.9*	92.2	90.8**	**91.6**	1.4 (2%)	
Sub-monthly	124.1	142.7	167.8*	98.0**	**133.2**	69.8 (52%)	45%
Monthly	89.7*	71.0	70.0**	80.2	**77.7**	19.7 (25%)	-15%

3.2 Nutrient calibration

On the example of annual nitrate-N loads, we executed three separate calibration procedures based on mean estimated load values of each sampling strategy (Table 2). First we calibrated the model based on the result from daily composite data set because it is

assumed to give best results. In Table 3 basic adjustments for nutrient (nitrate) output calibration are listed.

Table 3. Basic adjustments for nutrient (nitrate) output calibration

Parameter		**Setting; description**
FRT_KG	amount of fertilizer applied to HRU	based on farming scale modelling results
BIOMIX	biological mixing efficiency	0.1-0.4; depending on landuse and applied management practices
SOL_NO31	initial nitrate concentration in the upper soil layer	5.0 mg/kg soil; based on average Nmin of 20 kg/ha
SOL_ORGN1	initial organic nitrogen concentration in the upper soil layer	cropland: 100mg/kg; grazing land: 450 mg/kg; urban area: 180 mg/kg; pasture: 80 mg/kg; mixed forest: 150 mg/kg
GWNO3	nitrate in groundwater	25 mg/l
RCN	nitrogen in rainfall	10.0 mg/l; based on nitrogen deposition of 60 kg/ha/a and average, annual precipitation
RSDCO	residue decomposition coefficient	0.05

The different calibration bases resulted in different parameter settings (Table 4) to describe the same processes within the watershed.

Table 4. Selected nitrate output calibration parameter settings based on: mean estimated loads of: composite (S_com), sub-monthly (S_sm) and monthly (S_mo) data set

		Value range	**S_com**	**S_sm**	**S_mo**
NPERCO	nitrogen percolation coefficient	0.01-1.0	0.9	0.9	0.5
CMN	rate factor for humus mineralization of active organic nutrients	0.0001-0.003	0.0001	0.0017	0.0005
N_UPDIS	nitrogen uptake distribution parameter	0.0-100	100.0	1.0	50.0
FRT_SUF	fraction of fertilizer applied to top 10 mm of soil	0.0-1.0	0.4	0.2	0.2

Simulation results for annual nitrate-N load predictions as well as maximum, minimum and mean calculated load estimation results for all sampling strategies are illustrated in Figure 3. Based on different calibration data sets results met the ranges of maximum and minimum estimation except for 2001 with composite data set. In 2001 for the composite data set the nitrate-N load is under predicted. The minimal range of load estimation results for this dataset and year makes it quite difficult to achieve the absolute value. Furthermore, we assume the unsatisfactory water balance simulation for this year led to this result. But we should keep in mind that it is difficult to appropriately evaluate model performance if the uncertainty in model validation data is high. In that case, we do not really know how well the model is reproducing actual conditions.

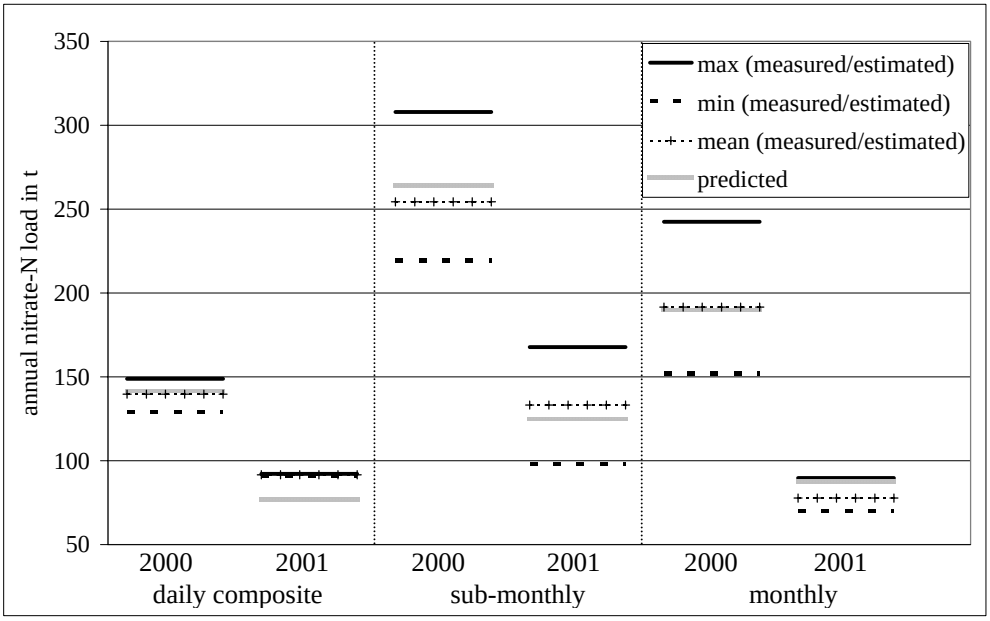


Figure 3. Results of annual nitrate-N load predictions

So load estimation qualities differs every sampling year depending on discharge conditions and randomness of sampling event related to discharge and concentration. This affects simulations of nutrient balance and makes the calibration process difficult.

CONCLUSIONS

For a satisfactory simulation of nutrient processes, a good calibration data base is required. In our case mean load estimation results using different sampling strategies and load estimation methods differ up to 82%, which can affect model evaluation conclusions. Hence, we postulate for the Parthe watershed the implementation of an intensive (daily) monitoring program to reduce measurement uncertainty and allow a more realistic judgement of model performance. Furthermore, we postulate the use of more than one load estimation method because of the fact that load estimations that uses only grab sampling methods are highly-uncertain. Therefore, if only grab sampled data bases are available we recommend the use of value ranges for model simulations. These value ranges should be used for discussion of the results and for suggestion of implementing best management practices.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the colleagues from the Saxony State Agency for Environment and Geology for providing the mean daily discharge data as well as the monthly sampled water quality data at the gauge Thekla of the Parthe watershed.

REFERENCES

Arnold, J.G., and P. Allen, Estimating hydrologic budgets for three Illinois watersheds, *J. Hydrol.* 176, 57-77, 1996.

Arnold, J.G., and N. Fohrer, SWAT2000: current capabilities and research opportunities in applied watershed modelling, *Hydrol. Process.* 19, 563-572, 2005.

Behera. S., and R.K. Panda, Evaluation of management alternatives for an agricultural watershed in a sub-humid subtropical region using a physical process based model, *Agri., Ecosys., Environ.* 113, 62-72, 2006.

Bracmort, K.S., M. Arabi, J.R. Frankenberger, B.A. Engel, and J.G. Arnold, Modeling long-term water quality impact of structural BMPs, *Trans. ASABE* 49(2), 367-374, 2006.

Chaplot, V., A. Saleh, D.B. Jaynes, and J. Arnold, Predicting water, sediment and NO_3-N loads under scenarios of land-use and management practices in a flat watershed, *Water, Air and Soil Poll.* 154, 271-293, 2004.

Ferguson R.I., Accuracy and precision of methods for estimating river loads, *Earth Surf. Proc. Landf.* 12, 95-104, 1987.

Gassman, P.W., M.R Reyes, C.H. Green, and J.G. Arnold, The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions, *Trans. ASABE* 50(4), 1211-150, 2007.

Harmel, R.D., K.W. King, J.E. Wolfe, and H.A. Torbert, Minimum flow considerations for automated storm sampling on small watersheds, *Texas J. Sci*. 54(2), 177-188, 2002.

Harmel R.D., K.W. King, and R.M. Slade, Automated storm water sampling on small watersheds, *Applied Eng. Agric*, 19(6), 667-674, 2003.

Harmel, R.D., and K.W. King, Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds, *Trans. ASAE* 48(5), 1713-1721, 2005.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, and J.G. Arnold. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Trans. ASABE* 49(3), 689-701, 2006a.

Harmel, R.D., K.W. King, B.E. Haggard, D.G. Wren, and J.M. Sheridan. Practical guidance for discharge and water quality data collection on small watersheds. *Trans. ASABE* 49(4), 937-948, 2006b.

Izuno, F.T., R.W. Rice, R.M. Garcia, L.T. Capone, and D. Downey, Time versus flow composite water sampling for regulatory purposes in the everglades agricultural area, *Applied Eng. Agric*. 14(3), 257-266, 1998.

Keller, M., M. Hilden, and M. Joost, Vergleich von Schätzmethoden für jährliche Stofffrachten am Beispiel des IKSR-Messprogrammes 1995 (BfG-1078), BMU-Maßnahme Nr. 4: Weitergehende Bewertung von Gewässergütedaten, JAP-BfG(1997): Nr.413, 1997.

King, K. W., and R.D. Harmel, Considerations in Selecting a Water Quality Sampling Strategy, *Trans. ASAE* 46(1), 63-73, 2003.

Littlewood, I.G., Hydrological regimes, sampling strategies, and assessment of errors in mass load estimates for United Kingdom rivers, *Environ. Int*. 21(2), 211-220, 1995.

Littlewood, I.G., and T.J. Marsh, Annual freshwater river mass loads from Great Britain, 1975-1994: estimation algorithm, database and monitoring network issues, *J. Hydrol*. 305, 221-237, 2005.

Nash, J.E., and J.V. Suttcliffe, River flow forecasting through conceptual models, Part I. A Discusion of Principles, *J. Hydrol. 10(3), 282-290*, 1970

Neitsch SL, Arnold JG, Kiniry JR, Srinivasan R, Williams JR., Soil and Water Assessment Tool, *User´s Manual. Version 2000. GSWRL Report 02-02*, BRC Report 2-06, Temple, Texas, USA, 2002.

OSPAR, OSPAR Guidelines for Harmonised Quantification and Reporting Procedures for Nutrients (HARP-NUT): Guideline 7: Quantification and reporting of the monitored riverine load of nitrogen and phosphorus, including flow normalisation procedures. Reference number: 2004-2-E, 2004.

Pandey, V.K., Panda, S.N., and S. Sudhakar, Modelling of an Agricultural Watershed using Remote Sensing and a Geographic Information System, *Biosys. Engin*. 90(3), 331-347, 2005.

Robertson D.M., and E.D. Roerish, Influence of various water quality sampling strategies on load estimates for small streams, *Water Resourc. Res*. 35(12), 3747-3759, 1999.

Robertson D.M., Influence of different temporal sampling strategies on estimating total phosphorus and suspended sediment concentration and transport in small streams, *JAWRA* 39(5), 1281-1308, 2003.

Stone, K.C., P.G. Hunt, J.M. Novak, M.H. Johnson, and D.W. Watts, Flow-proportional, Time-composited, and grab sample estimation of nitrogen export from an Eastern Coastal Plain Wastershed, *Trans. ASAE* 43(2), 281-290, 2000.

Swistock, B.R., P.J. Edwards, F. Wood, and D.R. Dewalle, Comparison of methods for calculating annual solute exports from six forested Appalachian watersheds, *Hydrol. Process*. 11:655-669, 1997.

Tate, K.W., R.A. Dahlgren, M.J. Singer, and B. Allen-Diaz, Timing, frequency of sampling affect accuracy of water-quality monitoring, *California Agriculture* 53(6), 44-48, 1999.

SMUL: www.smul.sachsen.de/de/wu/umwelt/wasser/inhalt_re_822_823.html on the 18th/July/2005.

Walling, D.E., and B.W. Webb, The reliability of suspended sediment load data, (Proceedings of the Florence Symposium, June1981), *IAHS Publications* 133, 177-194, 1981.

Webb, B.W., J.M. Phillips, D.E. Walling, I.G. Littlewood, C.D. Watts, and G.J.L. Leeks, Load estimation methodologies for British rivers and their relevance to the LOIS RACS(R) programme, *Sci. Total Environ*. 194-195, 379-389, 1997.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marré, J. Béjar, J. Comas, A. Rizzoli and G. Guarisco (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Effects of „ENSO-events" and rainforest conversion on river discharge in Central Sulawesi (Indonesia) – problems and solutions with coarse spatial parameter distribution for water balance simulation

G. Gerold[a] and C. Leemhuis[b]

[a] *Landscape Ecology Unit, Institute of Geography, University of Göttingen, Goldschmidtstr. 5, 37077 Göttingen, Germany (ggerold@gwdg.de)*
[b] *Center for Development Research, University of Bonn, Walter-Flex-Str. 3, 53113 Bonn, Germany (leemhuis@uni-bonn.de)*

Abstract: Natural interannual climate fluctuation as ENSO-events and forest conversion play an important role for the water balance of tropical catchments. Distributed hydrological modelling that relate land cover changes and climate changes with river discharge changes for humid tropical catchment areas at the mesoscale level are rare. This article applies a hydrological modelling approach to analyse the impact of land cover changes and ENSO-events on the water resources of a mesoscale humid tropical catchment. Because of a lack of detailed spatial soil information, a topographic based nested scale approach to parameterise soil physical properties was used for the hydrological model WASIM-ETH (Schulla and Jasper, 1999). The PHA (potential homogeneous soil texture areas) approach produces reasonable results for the water balance components for the whole catchment, but is too coarse for flow component differentiation within the catchment.
Thereafter the distributed hydrological model WASIM-ETH ("Water balance Simulation Model") was calibrated and validated for the current land use (2001/2002, Landsat/ETM+ scene). Model results were generally consistent with the observed discharge data and reproduced the seasonal discharge dynamics well. The implications of possible future climate and land use conditions on the water balance of the Gumbasa River sample catchment were assessed by a scenario analysis. The results of the scenario simulations clearly demonstrate a strong relationship between deforestation rates and increasing discharge variability. Especially a significant increase of high water discharges was simulated for the applied land use scenarios. The main results of the scenario analysis are:
(1) that ENSO anomalies of precipitation (monthly decrease for an average ENSO -21 until -80% from June to October) lead to an increase of discharge variability.
(2) that strong ENSO-events (El Nino) lead to 30% reduction of the water yield and annual crop scenarios up to 1,200 m a.s.l. showed a 42% and the perennial crop scenario (cacao) a 23% increase in river discharge with high increase in overland flow and flooding risks.
(3) that ENSO-events (El Nino) decrease the potential (flooding) area of paddy rice cultivation to one third in the second half of the year.
With regard to the high deforestation rates of the research catchment and increase of El Nino frequency with climate change in the future, one can expect further negative changes for the water resources in Central Sulawesi.

Keywords: *Hydrological Modelling; ENSO; Tropical Deforestation; Spatial Parameter Generation.*

1. INTRODUCTION

Tropical catchment areas are mainly located within developing countries and are predominantly ungauged or poorly gauged. The vast majority of tropical catchment studies have been conducted at the lower meso spatial (<10 km²) and time (<5 years) scale (Bruijnzeel, 1996). Costa et al. (2003) confirm that studies relating changes in land cover with changes in river discharge of tropical catchments are abundant especially on an experimental scale (< 1 km²). The global and local scale, which is applied in the vast majority of tropical hydrological studies (e.g. Güntner, 2002; Döll et al., 2003; Godsey et al., 2004; Fleischbein et al., 2005) is far from the regional reality where decisions on water resource management are developed and implemented (Falkenmark and Rockström, 2004). Data availability is one of the main difficulties for the analysis of tropical catchment processes on a mesoscale level. An appropriate database, which is urgently needed to enable water resources to be developed and managed, is lacking (IAHS 2003). Therefore long term records, which are required to study hydrological trends, are not widely available (Manley and Askew, 1993). Considering the increasing stress on water resources in humid tropical developing countries due to rapid deforestation rates and climate change there is an urgent global research need in humid tropical hydrology and its associated mesoscale catchment processes. After Abbot and Reefsgard (1996) the spatially-distributed and time-dependent hydrological model becomes the “conditio sine qua non” for investigations in this area.

The El Niño/ Southern Oscillation (ENSO) phenomenon is the strongest known natural interannual climate fluctuation. The two most recent extreme ENSO events during 1982/83 and 1997/98 had severe impacts on major parts of Indonesia’s socio-economy. Understanding the hydrology of humid tropical catchments is an essential prerequisite to investigate the impact of climate variability. Additionally, land use changes by forest conversion alter the components of the water cycle in catchments of Indonesia. Rainforest conversion predominately into annual cultures and cacao-systems was intensified in Central Sulawesi during the last decade. Within STORMA (DFG-project) and IMPENSO (BMBF-project) project work we investigate and quantify the impact of ENSO induced climate variability and forest conversion effects on the water balance of a mesoscalic tropical catchment. Main goals are: how ENSO-caused precipitation anomalies and forest conversion effects the main water balance components and river discharge in a mesoscale tropical catchment.

2. STUDY AREA – GUMBASA RIVER CATCHMENT

The Gumbasa River catchment (Fig. 1) is a sub-basin of the Palu River catchment (2,694 km²), which is located in Central Sulawesi, Indonesia (1°S, 120°E). The Gumbasa River catchment has a total drainage area of 1,275 km² and is characterized by a gross annual areal precipitation of 2,000 mm/yr and a mean annual discharge of 22 m^3s^{-1}. The seasonal hydrological regime of the catchment is described by two peaks and strongly corresponds with the bimodal monsoonal pattern of the yearly precipitation distribution of Central Sulawesi, Indonesia. Like other river basins that are located in monsoon regions also the Gumbasa River shows a great variation of seasonal and annual flow (Oyebande and Balek, 1987). ENSO years are characterized by reduced precipitation from July to October, which represents the dry period of the monsoonal setting of Central Sulawesi. The elevation ranges from 99-2,491 m.a.s.l with a mean elevation of 1,119 m.a.s.l. and a mean slope of 10.37°. About two-thirds of the catchments area is classified as slope with moderate and strong inclination (69 %), which emphasizes the mountainous character of the watershed. The irrigated paddy fields and other agricultural land use systems are situated mainly in the valleys. Mountainous tropical forest with a total percentage of 82 % forms the main land cover type within the watershed.

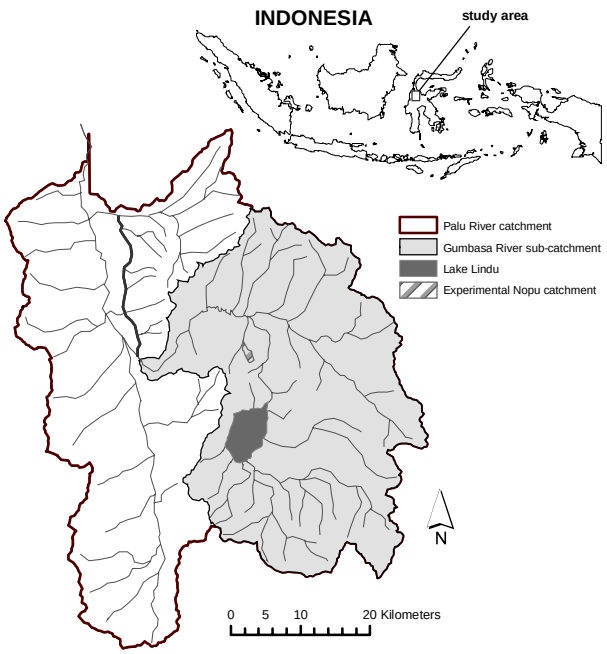


Figure 1. The Palu River watershed, the Gumbasa River modelling catchment and the experimental Nopu catchment, Central Sulawesi, Indonesia (1°S, 120°E)

3. MATERIAL AND METHODS

3. 1 Temporal and spatial data base

To study the impact of forest conversion and ENSO effects on the water balance of the Gumbasa River catchment we applied the Water Flow and Balance Simulation Model WASIM-ETH (Schulla and Jasper, 1999). WASIM-ETH is a process-based fully distributed catchment model. The spatial resolution is given by a grid and the time resolution can vary from minutes to days. The main processes of water flux, storage and the phase transition of water are simulated by physically-based simplified process descriptions (Schulla, 1997). The meteorological input data of the model are interpolated for each grid cell and are followed by the simulation of the main hydrological processes like evapotranspiration, interception, infiltration and the separation of discharge into direct flow, interflow and base flow. These calculations are modularly built and can be adapted to the physical characteristics of the catchments area. For the set up of the Gumbasa River catchment a 500 m x 500 m raster spatial, and a daily temporal resolution was chosen. As spatial input data WASIM-ETH requires a Digital Terrain model, a land cover map and a soil map. All spatial data is required in a raster data format (grid) with unique spatial resolution and extension. Starting from the primary grid of the topography secondary grids were generated with a topography analysis program.

To obtain a feasible temporal data set of meteorological forcing and hydrological calibration data for the performance of a distributed hydrological model the catchment was equipped with four automatic stage recorders and eight automatic climate stations monitoring precipitation, temperature and humidity. Additionally five automatic climate stations measure the complete meteorological parameters including wind speed and global radiation (Fig. 2). We measured hydrological and meteorological data for the period Sept. 2002 – Sept. 2004. Yearly areal precipitation of Gumbasa catchment (2,150 mm) is comparable with yearly precipitation averages from climate stations (e.g. Kulawi 1916-1941: 2,446 mm). For the calibration (01/09/2002–31/08/2003) and validation (01/09/2003–31/08/2004) of the hydrological model WASIM-ETH we chose the land cover classification of the Landsat/ETM+ scene from 2002. The land cover class concept is based

on a global hierarchical class definition system (LCCS) and has been adapted to the project region in Central Sulawesi. Each land cover type refers to corresponding physical vegetation parameters like stomata resistance, Leaf Area Index (LAI), albedo, vegetation height, rooting depth and vegetation cover. The parameterization of the vegetation physical properties was mainly derived from a global vegetation guide by Matthews (1999), single studies (Mo et al., 2004; Dijk and Bruijnzeel, 2001; Körner, 1994) and local studies (Bohmann, 2004; Falk, 2004) which were conducted within the research area.

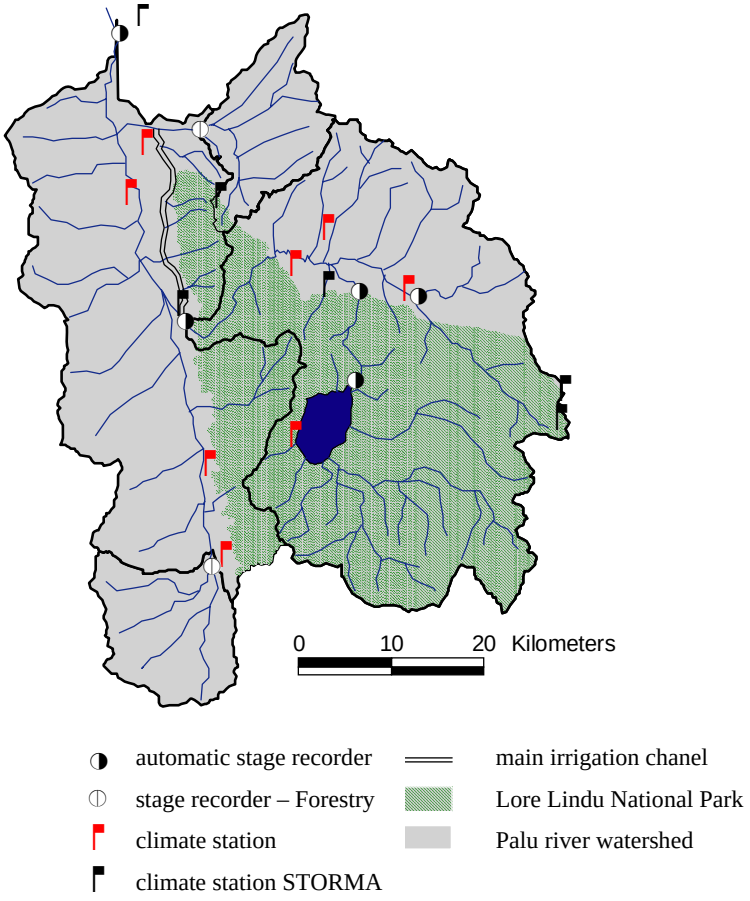


Figure 2. Location of the climate and hydrological stations within the Palu river watershed, Central Sulawesi, Indonesia

3. 2 Application of the topographic based soil parameterisation approach

The only available soil map of the research area is the soil map of Indonesia (1 : 1,000,000), which is based on available exploration soil maps that were published in various years. Therefore the “Pedotransfer Function” approach (Bohne, 1993) for the parameterisation of soil hydraulic properties could not be applied at all. PTFs are the most widely used method to estimate soil hydraulic properties for larger areas (Hodenett and Tomasella, 2002). Sobieraj et al. (2001) evaluated the performance of nine published PTFs for estimating the soil hydraulic properties of a rainforest catchment for modelling stormflow generation. They conclude that the published PDFs are inadequate to model stormflow generation, because runoff was overestimated for all events. The main differences between van Genuchten soil water-retention parameters for temperate and

tropical soils were investigated by Hodenett & Tomasella (2002). Their survey showed that most of the PTFs have been developed using databases for the soils of temperate regions, which are non-transferable for tropical soils.

Based on an experimental small headwater catchment of the Gumbasa River (Nopu 2.3 km^2, Fig. 1), studied for the effects of forest conversion on water balance and nutrient output since 2001 within the STORMA project (Kleinhans, 2004), a topographic based nested scale approach was developed. The topographic based nested scale approach to parameterise soil physical properties for mesoscale hydrological modelling consists of an experimental hillslope soil analysis and a classification of the mesoscale catchment area in potential homogeneous soil texture areas (PHA) according to their topographic characteristics. The surveyed catena of the experimental catchment located within the mesoscale catchment area serves as a leitcatena assuming a similar soil hydraulic behaviour for equal PHAs. The topographic based nested approach assumes that texture classes and bulk density mainly affect the soil physical properties and depend on the degree of slope and its geomorphologic location (basin, valley, slope, summit area). For the classification of PHA`s complex morphometric terrain factors are calculated on the basis of the Digital Terrain model (DTM). The relevant classification factors are (a) the relative altitude above channel line (a.a.c.l.), (b) catchment area and (c) slope angle. Table 1 lists the morphometric terrain description of a PHA classification.

Table 1: PHA classification and its corresponding morphometric terrain factors

	Morphometric description	mean values		
		m.a.s.l. (m)	**a.a.c.l. (m)**	**slope (°)**
I	Intramontane basin (low altitude above channel line)	591.46	2.68	1.60
II	Intramontane basin (moderate altitude above channel line)	453.80	8.00	2.50
III	Valleys outside the intramontane basin catchment (large catchment area)	802.05	5.17	9.31
IV	Valleys outside the intramontane basin catchment (moderate catchment area)	1154.31	3.71	12.14
V	Slope with moderate inclination	1191.37	88.13	17.43
VI	Slope with strong inclination	1143.57	121.84	30.77
VII	Summit area with low inclination	1337.96	87.88	5.31
VIII	Summit area with moderate inclination	1340.72	168.00	12.74

Table 2: PHA classes and its associated soil physical parameters

PHA	Pot. texture	θ_s	θ_r	α	n	ψ	k_s	k_{rec}	l_t	l_n
I	sand	0.45	0	7.36	1.23	385	9.0E-4	0.1	1.0	10
II	sand	0.45	0	7.36	1.23	385	5.0E-4	0.1	1.0	10
III	loamy sand	0.41	0	1.86	1.26	375	4.0E-4	0.1	0.9	10
IV	loamy sand	0.41	0	1.86	1.26	375	5.0E-4	0.1	0.9	10
V	sandy loam	0.45	0	4.01	1.2	345	1.5E-3	0.1	0.9	10
VI	loam	0.49	0	4.01	1.2	350	5.3E-3	0.1	0.7	10
VII	sandy clay loam	0.51	0	2.0	1.13	290	5.0E-4	0.1	0.3	10
VIII	sandy clay loam	0.51	0	2.0	1.13	290	4.0E-4	0.1	0.9	10

The surveyed catena of the experimental catchment served as a leitcatena for the whole mesoscale catchment area. Hence a similar soil hydraulic behaviour was assumed for equal PHAs. Table 2 lists the required soil hydraulic parameters for the soil model of WASIM-

ETH according to the PHAs classes. Instead of using the measured values of the saturated hydraulic conductivity k_s we increased the value consistently. With the hydrological model application of the experimental catchment study using the initial measured saturated hydraulic conductivity values no satisfactorily calibration modelling results were achieved. This effect is related to the modelling concept of the soil model using the Richards Equation. Here the soil is represented by a homogeneous matrix and therefore the flux of water through the matrix is characterised by a homogeneous matrix flow. But in reality the soil has an inhomogeneous structure, which implies a heterogeneous flux of water with preferential flow through macropores. Thus if we assume a flux of water through a homogeneous soil matrix the saturated soil hydraulic conductivity is relatively higher than the measured value. The recession constant k_{rec} represents an adjustable calibration parameter that determines the varying saturated hydraulic conductivity with increasing soil depth. Initially this parameter is set to 0.1. The parameter values of the suction ψ were transferred from the study of Niehoff (2001). The parameters describing the van Genuchten equation were computed from measured values. Empirical studies (Woesten et al., 2001) showed a good parameter adjustment if the residual water content θ_r was set to zero. Accordingly θ_r was set to zero for all PHAs classes. In general it should be reminded that a calibration parameter no longer describes the measured point value but rather an effective parameter that represents the average value for the element (Grayson, 2000). Hence they reproduce the bulk behaviour of a finite volume and cannot directly be related to point measurements at all.

3. 3 Calibration and validation

The calibration period was set for the period 01/09/2002-31/08/2003, which corresponds to the first year of the measurement period. To achieve realistic soil moisture conditions for the initial conditions of the calibration run, the start of the whole model run was predated to the 01/09/2001. The required temporal data for this period of model initialisation was copied from the first year of measurements. The validation period was split into the validation period (01/09/2003-31/08/2004), and the total period of simulation (01/09/2002-31/08/2004). Table 3 list the complete calibration and validation results for the model efficiency of the Danau Lindu, Sopu and Takkelemo gauging sites simulation results on a daily and weekly resolution. Further the graphical evaluation of observed and simulated discharge for the Lake Lindu, Takkelemo and Sopu sub-catchment (Fig. 3, Leemhuis, 2005) shows a very good agreement. The baseflow is simulated within realistic values and also the discharge fluctuation during the rainy season from March till May is well reproduced. The flat hydrograph of observed discharge for the Lake Lindu sub-catchment, which occurs end of April, is due to a faulty design of the hydrological station.

Table 3: Coefficients of efficiency R^2 for the sub-catchments Lake Lindu, Sopu and Takkelemo on a daily and weekly resolution for the calibration, validation-split sample and validation whole period

R^2	Lake Lindu		Sopu		Takkelemo	
	daily	weekly	daily	weekly	daily	Weekly
calibration period	0.83	0.86	0.79	0.77	0.58	0.77
validation period - split sample	0.61	0.62	0.58	0.44	0.41	0.50
validation period - whole	0.74	0.76	0.69	0.68	0.45	0.61

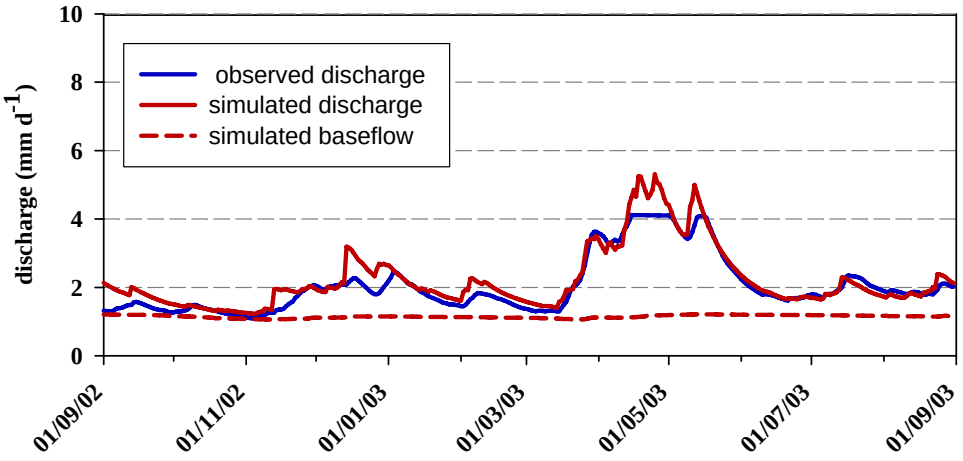


Figure 3. Calibration of the Lake Lindu sub-catchment (daily resolution) : comparison between observed and simulated discharge (R^2 = 0.83)

3. 4 ENSO scenario generation

A statistical scenario approach was applied to generate scenarios of spatially and temporally variable ENSO-caused rainfall anomalies as input data for a hydrological scenario run of the Gumbasa River watershed. Five climate stations with a long term historical record (1981-1999) were located within the area around the climate stations, which were used for the calculation of areal precipitation. The interdisciplinary climate research group of the IMPENSO project calculated the monthly precipitation anomaly of ENSO years in percent related to the mean monthly precipitation of the observed years. Therefore two graded ENSO scenarios which reflect the mean anomaly of all observed ENSO events (1987, 1991, 1994, and 1997) and on the other hand the extreme ENSO event of 1997, were generated. Thereafter the climate stations used for the interpolation of areal precipitation were allocated to one of the long term stations according to their elevation and topographic position. The allocation is mainly basin related although each of the long-term climate stations is located within one of the main basins,. The study of Aldrian (2003) has shown that ENSO events have a significant influence on the climate of Central Sulawesi between the dry period from June till October. The ENSO scenario was applied for this time frame which is affected by ENSO related precipitation anomalies. For the generation of the two graded ENSO scenarios the modelling year 2003 was taken as a base year.

4. DISCUSSION AND RESULTS

4. 1 Land use change

The application of the hydrological model WASIM-ETH on the Gumbasa River catchment in Central Sulawesi, Indonesia is the model's first simulation of the water balance of a mesoscale humid tropical catchment. The achieved model efficiencies can be classified among other WASIM-ETH model applications in various catchment studies of different climatic regions (Schulla, 1997; Jasper et al., 2002; Piepho, 2003). The modularly based character of WASIM-ETH allows a general adaptation of the hydrological model to different climatic regions and catchment data bases in space and time. If the overall model uncertainty is assessed with regard to the available spatial data set the land use classification and the soil classification represent in general uncertainty sources. With regard to land use classification derived from satellite images like Landsat ETM+ it should

be further analysed if different land use classification methods result in different modelling results.

Table 4: Total yearly simulated water balance for the Gumbasa River catchment and the Lake Danau and Takkelemo sub-catchment for 2003, compared to an annual crop scenario, perennial crop scenario (cacao) and ENSO-scenarios with average ENSO-event and 1997 ENSO-event; Δ represents the changes of precipitation, evapotranspiration and total discharge in percent to the control run (current climate and land use 2003)

Scenario	**catchment forest area (%)**	**precipitation**		**evapo-tranpiration**		**total discharge**	
		P (mm)	**Δ P** (%)	**aET** (mm)	**Δ aET** (%)	**Q** (mm)	**Δ Q** (%)
current climate a. land use 2003	Gumbasa (82) L.Danau (95) Takkelemo (79)	2150 2192 2259	- - -	1543 1295 1428	- - -	590 896 804	
current climate 2003 a. annual crop[1]	Gumbasa (82) L.Danau (95) Takkelemo (79)	2150 2192 2259	- - -	1265 1112 1124	- 20.0 - 14.1 - 21.3	838 1218 1113	42.0 35.9 38.4
current climate 2003 a. cacao plantation[2]	Gumbasa (82) L.Danau (95) Takkelemo (79)	2150 2192 2259	- - -	1361 1144 1222	- 11.8 - 11.7 - 14.4	724 1097 1000	22.7 22.4 24.4
average ENSO event a. land use 2003	Gumbasa (82) L.Danau (95) Takkelemo (79)	1614 1710 1724	- 24.9 - 22.0 - 23.7	1496 1281 1423	- 3.0 - 1.1 - 0.4	401 684 561	- 32.0 - 23.6 - 30.2
ENSO event 1997 a. land use 2003	Gumbasa (82) L.Danau (95) Takkelemo (79)	1586 1699 1709	- 26.2 - 22.5 - 24.4	1487 1279 1421	- 3.1 - 1,2 - 0.5	397 684 557	- 32.7 - 23.6 - 30.7

[1] forest conversion with annual crop until 1200 m a.s.l.
[2] forest conversion with cacao plantation until 1200 m a.s.l.

We simulated a land-use history scenario with the Landsat/ETM land cover classification of 2001 as input land cover grid for the analysis of the impact of observed land cover changes from 2001 till 2003 within the Gumbasa River catchment (details on land cover changes see Leemhuis et al., 2007). Furthermore we applied a total change scenario, which assumed that all forest up to 1,200 m a.s.l. is converted in a first conversion phase into annual crops such as maize or dry rice and within a second step into perennial crops (table 4).

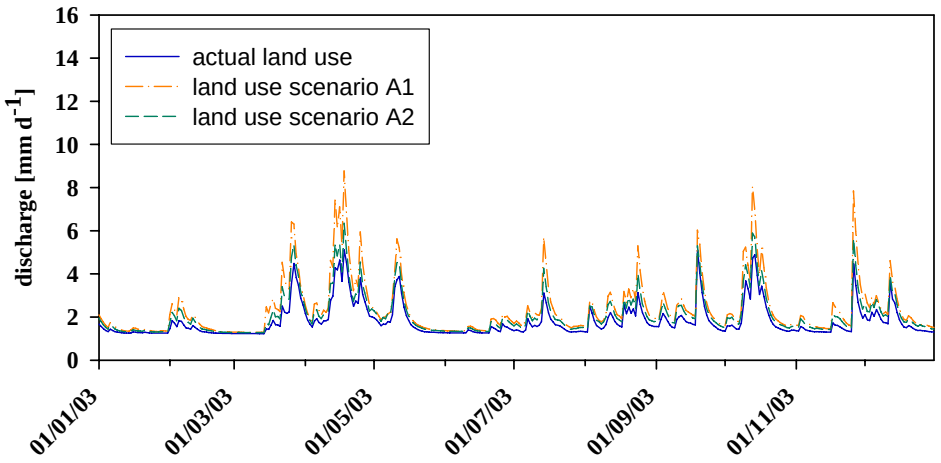


Figure 4. Scenario results for daily discharge (Gumbasa catchment) with two different land use developments (A1 = conversion of rainforest ≤ 1,200 m a.s.l. in annual crops; A2 = coversion of rainforest ≤ 1,200 m a.s.l. into cacao plantation)

Cacao is a cash crop in Central Sulawesi and therefore is most likely to be planted as perennial crop. In Central Sulawesi the cacao plant can be cultivated up to an elevation of 1,200 m a.s.l., since up to this elevation the temperature does not drop beneath 16°C at night (Rehm, 1989). For both scenarios the physical vegetation parameters were altered according to the applied land cover type. The soil physical properties were kept constant, because the changes of soil physical parameters induced by land cover changes were not investigated for the Gumbasa River catchment.

The comparison of the simulated yearly water balance and the applied land cover scenarios gives a general first overview about the impact of land cover change on the hydrological performance of the observed watershed. Table 4 compares the changes induced by land cover change for the total Gumbasa River catchment and the sub-catchments of Danau Lindu (95% forest cover) and Takkelemo (79% forest cover) respectively of the following water balance components: areal precipitation (P), actual evapotranspiration (aET) and total discharge (Q). With respect to the actual land cover changes from 2001 to 2002 the effect on the yearly simulated water balance is rather insignificant. The hydrological model calculated a 0.2% of the total water balance decrease of total yearly discharge for the Gumbasa River catchment (Leemhuis et al., 2007).
The annual crop scenario with an increase of 42% of the yearly total run-off for the Gumbasa River catchment and the perennial crop scenario with a rise of 23% of the yearly total discharge indicate a great increase of the yearly total discharge with cumulative deforestation rates (Fig. 4). All changes of the total discharge were induced by an alteration of the yearly evapotranspiration rate, due to a decline of LAI, root depth, vegetation height and stomata conductance. Less water is evaporated by interception and hence throughfall increases and more water can infiltrate into the soil. Plot research results within STORMA shows a reduction in aET for annual crops (maize) with an order of 10-20% and for perennial crops (cacao) with 5-12 % (Falk 2004). The reduction of the evapotranspiration causes the soil to be wetter and therefore more responsive to rainfall (Bruijnzeel, 2004). The graduated impact of annual and perennial crop scenario is related to the more similar physical vegetation parameters of the cacao tree in comparison to the broadleaf evergreen forest of the catchment, comparing relationship between forest cover change and changes of the yearly total run-off. Regarding the crop scenario for the Danau Lindu sub-catchment the lower increase of discharge is due to the less decrease of evapotranspiration.

4.2 ENSO effects

For the observed ENSO-events two scenarios which reflect the mean anomaly of all observed ENSO-events (1987, 1991, 1994 , 1997; = average ENSO event) and the extreme ENSO event of 1997 were generated. The current conditions of 2003 serve as a control run for the evaluation of the proportional changes. During ENSO scenario A the total yearly areal precipitation of the Gumbasa River catchment is reduced by 25 %, whereas the total yearly evapotranspiration rate is only slightly minimised by 3 %. The main difference of the yearly water balance were observed for the total yearly discharge, which is reduced by 32 % (Fig. 5).

With a reduction of the total discharge by 24 %, the Danau Lindu sub-catchment is least affected by the precipitation anomaly, whereas the small Takkelemo catchment reaches similar reduction values like the Gumbasa River catchment. Even though the strong ENSO event of 1997 is applied by ENSO scenario B, total yearly changes of the water balance in comparison with ENSO scenario A are minimal. For example the decrease of the yearly total discharge only is reduced by 0.7 % for the Gumbasa River catchment. The comparison of the yearly water balance of the three different catchment types shows that the ENSO-caused precipitation anomaly leads to a slight variance of the yearly precipitation decline among the catchments of 2.9 % for ENSO scenario A and 3.7 % for ENSO scenario B. For both ENSO scenarios the Danau Lindu subcatchment indicates with 22.0 % (Scenario A) and 22.5 % (Scenario B) the lowest yearly precipitation decline, whereas the Gumbasa River catchment has a yearly average precipitation decline of 24.9 % (Scenario A) and 26.2

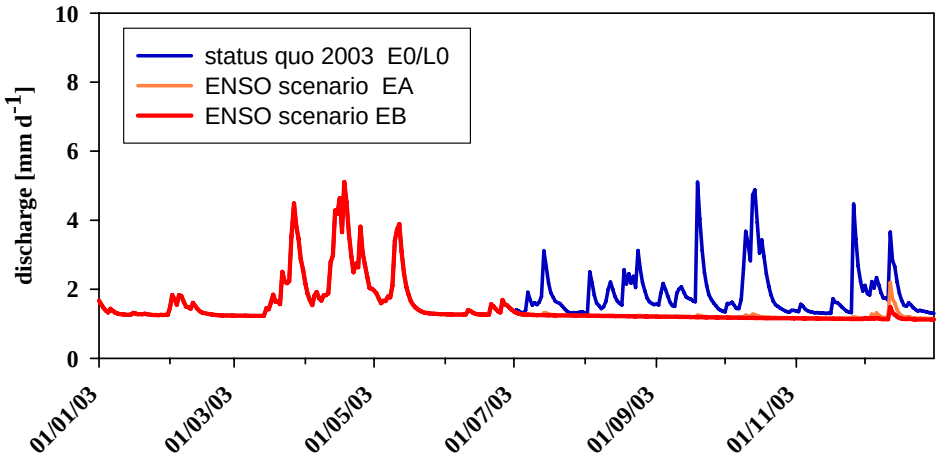


Figure 5. Scenario results for daily discharge (Gumbasa catchment) with two different ENSO-events (EA = average year with rainfall decrease from 1.07. – 30.11.2003 with - 60%; EB = average year with rainfall decrease 90%)

% (Scenario B). The simulated yearly precipitation decline is not linear with the simulated yearly discharge decline, which indicates a far wider variance of the different catchment types with 8.4 % for ENSO scenario A and 9.1 % for ENSO scenario B. Again for both scenarios the Danau Lindu catchment has the lowest discharge decline with 23.6 % for both ENSO scenarios and the Gumbasa River catchment the highest discharge decline among the observed catchments with 32.0 % (ENSO scenario A) and 32.7 % (ENSO scenario B). The simulation for the Takkelemo catchment reaches almost similar precipitation and discharge decline values as for the Gumbasa River catchment. The Danau Lindu catchment is characterized by a low discharge variability due to its retention capability. In general the discharge variability during ENSO years is significantly higher.

5. Conclusions

Common problems in mesoscalic hydrological modelling in developing countries are insufficient time series of input- and output data and insufficient spatial data resolution. For the Palu river catchment only one Indonesian climate station and two automatic stage recorders from the Forestry Department of Palu for two subcatchments existed at the beginning of the IMPENSO-project. Therefore an additional network of six automatic stage recorders and eight additional climate stations have been used to collect data since September 2002 for the application of the water balance model WASIM-ETH. The main problem was the coarse scale of available soil maps (1 : 1 Mill.) and lumped representation of soil data from STORMA (isolated small areas). Therefore a more general topographic- and knowledge-based approach has been applied, that classifies potential homogenous areas (PHA) with morphometric terrain factors (Böhner et al., 2002, Leemhuis, 2005). The PHA approach produces reasonable results for the water balance components for the whole catchment, but is to coarse for flow component differentiation within the catchment. However, the calculated decrease of the yearly total discharge by about 30% during ENSO events corresponds with a cross-correlation analysis of discharge time series for two catchments in Central Sulawesi (Miu and Wuno River; Leemhuis, 2005).

ENSO caused precipitation anomalies lead to an overall increase of the discharge variability. Moreover the ENSO scenario discharge simulations shows that mainly the low water and mean discharges are effected in the second half of the year (August till December) with a shift of one to two months. Strong ENSO-events like 1997 lead to 32%

reduction in water yield. Catchment characteristics, mainly change in forest area, cause different hydrological response magnitude to the ENSO impact. The relative lower discharge decrease of the Danau Lindu catchment is due to the higher retention capability of the Lake Lindu. The higher portion of forest area is coupled with slower drying out of the soil water reservoir and a reduced evapotranspiration rate for the high altitude rainforest. It can be concluded, that for the “Gumbasa Irrigation Scheme”, based on the water resources of the Gumbasa catchment, the conditions of the mountainous rainforest headwater catchments like the Danau Lindu with less disturbed rainforests are important for the input of irrigation water. With average ENSO events scenario results show for the second half of the year a decrease in potential irrigation area of the Gumbasa River Irrigation Scheme by about -23 until -66 % (Leemhuis, 2005).

The results of the applied land use scenarios agree with the findings of the tropical catchment study by Kleinhans (2004)], who applied the hydrological model WASIM-ETH on a small headwater catchment of the Gumbasa River catchment. Both studies used the same hydrological model, therefore it was likely that similar land use scenario results would be also calculated on a bigger scale. Experimental paired catchment studies confirm the simulated increases of peak flows after forest removal (Bonel and Balek, 1993; Chandler & Walter, 1998), which is explained by the associated reduction in ET that causes the soil to be wetter and therefore more responsive to rainfall (Bruijnzeel, 2004). Furthermore the scenario results show that after the complete conversion of all deforested area into cacao plantations still the high water discharge is characterised by enhanced peak flows, which indicates a continued alteration of the water balance for cacao plantations.

But cacao plantations have a more moderate impact on the water balance of a catchment than annual crops like maize. Interpreting the simulation results of the dry season flow (low water discharges) requires caution. Because numerous studies of smaller tropical catchments produced higher baseflow after forest removal, whereas other studies report a decreased baseflow. Regarding these contradictory catchment study results, Bruijnzeel (2004) raised the discussion about the “low flow problem”. Whether the baseflow is raised or reduced after forest removal is strongly connected to the degree of surface disturbance. Costa (2005) argues that as a result of the decreased hydraulic conductivity due to soil consolidation less water can infiltrate, which again leads to an increased direct discharge and actually leads to a decrease of the low water discharge. The “low flow problem” indicates that information on land cover change (from Landsat/ETM+ satellite images) and data on changes of the soil physical properties after forest removal are necessary to improve simulation correctness for the low water discharges. Further, it would also improve the simulation of the mean and high water discharges. In a recent study in the small headwater Nopu catchment the importance of the change of saturated hydraulic conductivity with age and intensity of traditional smallholder land use is shown (De Vries et al., 2008). There extensive used cacao plantations conserve high rainfall infiltration into the soil and leads to higher groundwater recharge.

We can conclude that green water flow plays a significant role within tropical catchments. Most of the alterations of the water balance due to forest conversion can be explained by changes of the green water flow. Besides monitoring of alterations of physical vegetation parameters, there is urgent need of research within tropical catchment areas to describe the corresponding changes of the soil physical parameters and to have higher spatial resolution for effective physical soil parameters.

ACKNOWLEDGEMENTS

This study was supported by the Federal Ministry of Education and Research (BMBF) and by the German Research Foundation (DFG). Special thanks to the Institut Pertanian Bogor (IBP), Java, Indonesia and the Universitas Universitas Tadulako (UNTAD), Palu, Central Sulawesi, Indonesia for the productive research cooperation.

REFERENCES

Abbott, M.B. and J.P. Refsgaard, Terminology, modelling protocol and classification of hydrological model codes. In: Abbott, M. B. and Refsgaard, J. P. (Eds.): Distributed hydrological modelling. Kluwer Academic Publishers, Dordrecht, Boston, London, 17-37, 1996.

Aldrian, E., Simulations of Indonesian rainfall with a hierarchy of climate models. Diss., Universität Hamburg, 23-45, 2003.

Alves, D. S., Escada, M. I. S., Pereira, J. L. G. and C.A. Linhares, Land use intensification and abandonment in Rondônia, Brazilian Amazônia. *International Journal of Remote Sensing*, 24, 899-903, 2003.

Asdak, C., Jarvis, P. G., van Gardingen, P. and A. Fraser, Rainfall interception loss in unlogged and logged forest areas of Central Kalimantan, Indonesia. *Journal of Hydrology*, 206, 237-244 1998.

Barnes,C. and M. Bonell, How to chose an appropiate catchment model. In: Bonell, M. and Bruijnzeel, L. A. (Eds.): Forests, Water and People in the Humid Tropics. University Press, 717-741, 2005.

Böhner, J., Köthe, R., Conrad, O., Gross, J., Ringeler, A. and T. Selige, Soil regionalisation by means of terrain analysis and process parameterisation.- EU Soil Bureau Research Report No. 7, EUR 20398 EN, 213-222, 2002.

Bohman, K., Functional and morphological diversity of trees in different land use types along a rainforest margin, in Sulawesi, Indonesia. Dissertation, Universität Göttingen, 55-63, 2004.

Bohne, K., Horn, R. and T. Baumgartl, Bereitstellung von van - Genuchten Parametern zur Charakterisierung der hydraulischen Bodeneigenschaften. *Zeitschrift für Pflanzenernährung und Bodenkunde*, 156, 229-233, 1993.

Bonell, M. and J. Balek, Physical processes. In: Bonell, M., Hufschmidt, M., Gladwell, J. (eds.): Hydrology and water management in the humid tropics. Cambridge, 176–177, 1993.

Bonell, M., Runoff generation in tropical forests. In: Bonell, M. and Bruijnzeel, L. A. (Eds.): Forests, Water and People in the Humid Tropics. University Press, Camebridge, 314-406, 2005.

Bruijnzeel, L.A., Hydrology of moist tropical forests and effects of conversion: a state of knowledge review. UNESCO International Hydrological Programme, 20-20, 1996.

Bruijnzeel, L.A., Tropical forests and water yield. Laxenburg, Austria, 2000.

Bruijnzeel, L.A., Hydrological functions of tropical forests: not seeing the soil for the trees? *Agriculture, Ecosystems & Environment*, 104, 185-228, 2004.

Chandler, D. G. and M.F. Walter, Runoff responses among common land uses in the uplands of Matalom, Leyte, Philippines. In: Transactions of the ASAE 41, 1635–1641, 1998.

Costa, M. H., Botta, A. and J.A. Cardille, Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. In: *Journal of Hydrology* 283, 206–217, 2003.

Costa, M. H., Large scale hydrological impacts of tropical forest conversion. In: Bonell, M. and Bruijnzeel, L. A. (Eds.): Forests, Water and People in the Humid Tropics. University Press, Camebridge, 590-598, 2005.

De Vries, K. and G. Gerold, The consequences of rainforest conversion to agricultural land for saturated hydraulic conductivity in Cental Sulawesi, Indonesia. *Catena* (submitted), 2008.

Dijk, A. and L.A. Bruijnzeel, Modelling rainfall interception by vegetation of variable density using an adapted analytical model. Part 2. Model validation for a tropical upland mixed cropping system. *Journal of Hydrology*, 247, 239-262, 2001.

Döll, P., Kaspar, F. and B. Lehner, A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology*, 270, 105-134, 2003.

Elsenbeer, H., Pedotransfer functions in hydrology. *Journal of Hydrology*, 251, 121-122, 2001.

Falk, U. , Turbulent fluxes of CO2, H2O and energy in the atmospheric boundary layer above tropical vegetation investigated by Eddy-Covariance measurements. Diss. Göttingen., 2004.

Falkenmark, M. and J. Rockström, Balancing water for humans and nature. Earthscan, London, 181-199, 2004.

Fleischbein, K., Wilcke, W., Goller, R., Boy, R., Valarezo, C., Zech, W. and K. Knoblich, Rainfall interception in a lower montane forest in Ecuador: effects of canopy properties. *Hydrological Processes*, 19, 1355-1371, 2005.

Godsey, S., Elsenbeer, H. and R. Stallard, Overland flow generation in two lithologically distinct rainforest catchments. *Journal of Hydrology*, 295, 276-290, 2004.

Grayson, R. and G. Blöschl, Spatial modelling of catchment dynamics. In: Grayson, R. and Blöschl, G. (Eds.): Spatial Patterns in Catchment Hydrology. Camebridge University Press, Camebridge, 51-81, 2000.

Güntner, A., Large-Scale hydrological modelling in the Semi-Arid North-East of Brasil. Dissertation, Universität Potsdam, 2002.

Hodnett, M.G. and J. Tomasella, Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water retention pedo-transfer functions developed for tropical soils. *Geoderma*, 108, 155-180, 2002.

IAHS, Prediction in ungauged basins. International Hydrology today, 26-28, 2003.

Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E. and E.D. Schulze, A global analysis of root distributions for terrestrial biomes. *Oecologia*, 108, 389-411, 1996.

Jasper, K., Gurtz, J. and H. Lang, Advances flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. *Journal of Hydrology*, 267, 40-52, 2002.

Kleinhans, A., Einfluss der Waldkonversion auf den Wasserhaushalt eines tropischen Regenwaldeinzugsgebietes in Zentral Sulawesi (Indonesien). Diss. SUB Göttingen 2004.

Körner, C., Leaf diffusive conductances in the major vegetation types of the globe. In: Schulze, E. D. and Caldwell, M. M. (Eds.): Ecophysiology of Photosynthesis. Springer, Berlin, 463-490, 1994.

Leemhuis, C., The impact of El Nino Southern Oscillation Events on water resource availability in Central Sulawesi, Indonesia. EcoRegio 21, 2005.

Leemhuis, C., Erasmi, S., Twele, A., Kreilein, H., Oltchev, A. and G. Gerold, Rainforest conversion in Central Sulawesi, Indonesia: Recent development and consequences for river discharge and water resources. *Erdkunde*, 61, 284-293, 2007.

Matthews, E., Global vegetation guide (1972-1983). http://www.daac.ornl.gov., 1999.

Mo, X., Liu, S., Lin, Z. and W. Zhao, Simulating temporal and spatial variation of evapotranspiration over the Lushi basin. *Journal of Hydrology*, 285, 125-142, 2004.

Piepho, B., Untersuchungen zum hydrologischen Reaktionsverhalten eines kleinen Einzugsgebietes unter Verwendung des Wasserhaushaltsmodells WaSiM-ETH. Dissertation, TU Braunschweig, 2003.

Rehm, S., Spezieller Pflanzenbau in den Tropen und Subtropen. Ulmer, Stuttgart, 437-446, 1989.

Schulla, J., Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen. Züricher Geographische Schriften, Verlag Geographisches Institut ETH Zürich, Verlag Geographisches Institut ETH Zürich, 1997.

Schulla, J., Jasper, K., Model Description WASIM-ETH, 1999.

Scurlock, J. M. O., Assner, G. P. and S.T. Gower, Global Leaf Area Index from Field Measurements, 1932-2000. http://www.daac.ornl.gov., 2001.

Sobieraj, J. A., Elsenbeer, H. and R.A. Vertessy, Pedotransferfunctions for estimating saturated hydraulic conductivity: implications for modeling storm flow generation. *Journal of Hydrology*, 251, 202-220, 2001.

Whitten, T. and G.S.M.M. Henderson, The ecology of Sulawesi.Periplus, Singapore, 2002.

Wösten, J. H. M., Pachepsky, Y. A. and W.J. Rawls, Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, 251, 123-150, 2001.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Uncertainty in Great Barrier Reef Catchment soil nutrient data - implications for land use management

B. S. Sherman[a] and A. M. Read[b]

[a]*CSIRO Land and Water, GPO Box 1666, Canberra ACT 2601 Australia (brad.sherman@csiro.au)*

[b]*CSIRO Land and Water,GPO Box 1666, Canberra ACT 2601 Australia (arthur.read@csiro.au)*

Abstract: The reduction of particulate nutrient loads (nitrogen and phosphorus) has been an important objective for managers of Great Barrier Reef Lagoon (GBRL) catchments. These loads are believed to be dominated by non-point sources, but end-of-catchment load measurements provide little insight into the distribution and relative magnitude of pollutant sources within the catchment. This has led to a reliance on computer modelling to identify probable pollutant sources and to devise and compare management strategies to deal with them. The most common modelling approach used in the GBRL region is to compute spatially-distributed budgets of nutrient sources and sinks. The largest source is usually hillslope erosion which is calculated using a form of the revised universal soil loss equation (RUSLE) and then assigned nitrogen and phosophorus contents to the soil based on values held in the Australian Soil Resource Information System (ASRIS). To this is added subsoil erosion from river banks and gullies where subsoil has traditionally been assumed to have spatially uniform phosphorus and nitrogen contents.

We assessed the uncertainty in the most recently updated ASRIS GBRL soil nutrient data and the implications of using these data for modelling and management of particulate nutrient loads at spatial scales ranging from individual farms to subcatchments with areas > 1000 sq. km. Bias in the assumed subsoil nutrient content suggests that particulate loads of subsoil phosphorus may be underestimated by 14% whereas subsoil nitrogen loads are likely to be overestimated by a factor of three. Variability (expressed as std dev/mean) of surface soil nutrient content at a single sampling site was 34% for nitrogen and 21% for phosphorus. When considering variability within the smallest resolved spatial units (unique mapping area, UMA) in ASRIS encompassing four or more sample sites the results were 38% and 48% for N and P, respectively. This level of intrinsic variability in the observed data in combination with the relatively large size of a UMA makes it very difficult to discriminate differences in nutrient fluxes at scales much less than 25 sq. km with confidence.

Keywords: catchment modelling; uncertainty; erosion; nutrients; ASRIS; Great Barrier Reef

1. INTRODUCTION

The reduction of sediment, nutrient (nitrogen and phosphorus), and pesticide loads is an important objective for managers of Great Barrier Reef Lagoon (GBRL) catchments (Brodie et al. 2003; Cogle et al. 2006; Williams 2001; Environment Australia 2002). Measurements of particulate nutrient loads at the mouths of some of the catchments have

suggested a very large increase in pollutant loads entering the Great Barrier Reef Lagoon following European settlement and the commencement of large scale mining, agricultural and pastoral industries in the area (Furnas 2003; McCulloch et al. 2003). With > 95% of the land area subject to grazing, agricultural, or forestry uses and < 2% of the land area urbanised the pollutant loads are believed to be dominated by non-point sources. Cogle et al. 2006 used the catchment model SedNet/ANNEX to estimate nutrient loads for the GBRL catchments and found point sources to contribute 16% of dissolved nitrogen and 44% of dissolved phosphorus loads but just 4% of and total nitrogen and 3% of total phosphorus loads.

End-of-catchment load measurements have, however, provided little insight into the distribution and relative magnitude of pollutant sources within the catchment. This has led to a reliance on computer modelling to identify probable pollutant sources and to devise and compare management strategies to deal with them. The most common modelling approach used in the GBRL region is to use the model SedNet/ANNEX to compute sediment supply due to river bank, gully and hillslope erosion processes and then assign nitrogen and phosphorus contents to the sediment based on values held in the Australian Soil Resource Information System (ASRIS)(Bormans et al. 2004; McKergow et al. 2005; Cogle et al. 2006). Subsoil sediment (defined here as B-horizon and below) has traditionally been assumed to have a constant (non-spatially varying) nutrient content (due to a lack of sufficient spatial data) whereas hillslope sediments have a spatially varying nutrient content. Sediment supplied by hillslope erosion is calculated using a form of the revised universal soil loss equation (RUSLE) (Wilkinson et al. 2004).

Catchment models require, as input data, spatially distributed estimates of a range of parameters such as rainfall, potential evapotranspiration, slope, vegetation, land use, etc. Much, if not most, of the apparent spatial resolution of these data sets is based either on interpolation or extrapolation of much sparser measurement data sets (e.g. rainfall, gully density, soil nutrient content) and/or may rely on associations between directly observed parameters (e.g. satellite remote sensing of the electromagnetic spectrum) and the parameters of interest (e.g. various landscape attributes)(Jeffrey et al. 2001; Henderson et al. 2001). These input data sets will contain uncertainty arising from intrinsic natural variability of the directly measured data as well as from the inaccuracies in the procedures used to map the source data points into a spatial grid.

Here, we present an assessment of the uncertainty in GBRL soil nutrient data found in the ASRIS database and the implications for modelling and management of particulate nutrient loads at spatial scales ranging from individual farms to subcatchments with areas > 1000 sq. km. This is done by assessing the variability of the original measurement data stored in the Queensland Soil and Land Information (SALI) database system from which the ASRIS values were derived.

2. METHODS

2.1 The ASRIS database

The ASRIS data set contains the up-to-date soil-related information for all of Australia. The structure of the data set is described in detail by (McKenzie et al. 2005). ASRIS is a repository for data and allows for 7 ranges of spatial resolution ranging from 10 m to 30 km. The resolution of any particular data set accessed through ASRIS reflects the resolution of the data supplied by the contributing agency. The resolution of soil nutrient data varies across GBRL catchments depending on the nature of the original surveys (e.g. soil survey, land system survey, land resource area survey) for which the data were collected (Brough et al. 2006). The soil nutrient data described here was originally sourced from the Queensland Soil and Land Information (SALI) database system and contributed by Queensland Department of Natural Resources, Mines and Water to the ASRIS database.

The ASRIS database was revised in late 2007 for Queensland and now provides soil attribute data that are based on direct measurements, where available, and by extrapolation based on a range of landscape attributes where no measurement data are available. ASRIS provides a single value of an attribute (e.g. total phosphorus content) for each unique mapping area (UMA). The size of UMAs varies because of the different survey methods used in different parts of the catchments with map scales ranging typically from 1:25 000 to 1:1 million. UMAs are generally discriminated using attributes such as slope, regolith material, climate, lithology and soil suborder (McKenzie et al. 2005). Within the GBRL catchments, UMAs have a wide range of sizes from 5 hectares to more than 7000 sq. km with an average size of roughly 35 sq km. This imposes a fundamental constraint on our ability to resolve spatially the particulate nutrient fluxes at the paddock and farm scales typically considered for the prioritisation of land use management interventions.

2.2 The SALI database

A detailed description of the SALI data and how it was processed prior to being provided to ASRIS is given by Brough et al. (2006). Total phosphorus values were determined using XRF and total nitrogen content was measured using an automated Kjeldahl method. For most measurement sites a number of samples from various depths in the soil profile are available and were analysed for nitrogen and/or phosphorus depending on landuse and the motivation for a particular soil survey. Soil horizon information is provided as well. A measurement performed on a single soil sample is referred to here as a 'spot' measurement. In addition to spot measurements, some sites also have 'bulk' measurements. Bulk samples are a composite of a number of surface soil samples collected in the vicinity of the sample site coordinates. The collection of samples is mixed together and then subsampled for analysis. The bulk sampling method is typically undertaken as part of soil fertility analyses and represents an average nutrient content for soil within some distance of the site coordinates. Bulk samples are flagged in the SALI database.

In this work, subsoil is defined as belonging to the B horizon or lower and where the top of the sample was located at least 0.3 m below the soil surface. Some sample locations had more than one subsoil sample. For total phosphorus (TP) there were 3298 discrete sample locations with a total of 9661 individual samples. Total nitrogen was less frequently measured with a total of 197 discrete locations and 445 individual samples.

A surface soil sample is defined as any sample having an upper sample depth of 0 m and may be either a spot or a bulk sample. The sample thickness was typically 0.1 m. Total phosphorus was determined for a total of 1413 sites with bulk samples (total of 1545 determinations) and for 3901 sites with spot samples (total of 4286 determinations). Both spot and bulk samples were available for P at 347 sites and for N at 547 sites.

2.3 Calculations

Two methods were used to compute mean subsoil nutrient concentrations: simple averaging of all individual samples; and calculation of a weighted mean subsoil concentration for each discrete location. For the weighted mean concentration, each concentration value was multiplied by the thickness of the sample strata, the results summed, and then divided by the sum of the sample thicknesses.

The surface concentrations were determined by averaging all values at a discrete location (latitude, longitude) where the top of the sample had a depth of 0 m. Bulk and spot samples were considered separately.

Basic statistical characteristics of the data (average, minimum, maximum, standard deviation, standard error) were computed for each site with at least four samples of a particular type.

All data were kept in a mySQL database and all calculations performed using the SQL language.

3. RESULTS

3.1 Subsoil nutrient content

Subsoil (B-horizon and below) nutrient contents averaged 0.028% P and 0.036% N. These measured values compare with the assumed values of 0.025% P and 0.10% N used in all modelling of GBR catchment nutrient loads up to and including 2006. Clearly, in regions with substantial gully and stream bank erosion the particulate nitrogen load is likely to be over-estimated by nearly a factor of 3 whereas the error in subsoil phosphorus load is just 14%.

3.2 Surface soil small scale variability

The small scale natural variability of soil N and P content within a sampling site was assessed by comparing composite bulk samples with spot samples where both methods were used at a site. After linearly regressing spot sample against bulk sample data and eliminating a few conspicuous outliers (Figure 1) the P data fit closely to a 1:1 slope (r^2 = 0.91) whereas the N data exhibited a bias in slope (spot:bulk = 0.93) and were slightly more variable (r^2 = 0.81).

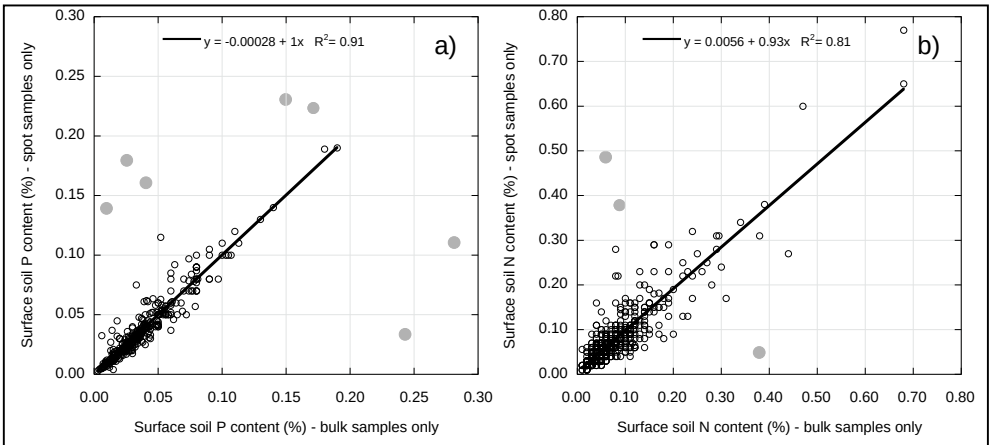


Figure 1. Comparison of surface soil nutrient content measured using spot and bulk sample collection methods for phosphorus a) and nitrogen b). Outliers omitted from curve fit are shown as grey circles.

3.3 Surface soil measurement precision

It is important to understand how accurately a single measurement represents the true soil nutrient concentration at a particular location. An estimate of the precision of the mean value for a specific site and sampling method (bulk or spot) was computed by calculating the means, ranges and standard deviations for each site where there were four or more determinations made using the same technique. Results averaged over the 37 qualifying sites with N measurements and 56 sites with P measurements are given in Table 1. Soil N content was less precisely determined than was soil P with the average standard deviation / mean being 0.34 (n = 37) and 0.21 (n = 56), respectively.

Table 1. Precision of multiple soil nutrient measurements at a single site where each site has at least four values for a specific measurement type (spot or bulk). Average values considering all sample sites for: mean, minimum, maximum P and N concentrations; standard deviation/mean; range and standard error.

	Average mean conc (%)	Average minimum conc (%)	Average maximum conc (%)	Average std dev/ mean	Average Range (%)	Average Std err of mean	n
N	0.078	0.049	0.13	0.34	0.077	0.01	37
P	0.029	0.022	0.038	0.21	0.016	0.003	56

3.4 Variability of soil properties within unique mapping areas

Having established how well we know the soil nutrient content at a single site we now wish to estimate by how much the soil nutrient content varies within a unique mapping area - the smallest scale at which soil nutrient content is typically resolved for catchment modelling in the GBRL catchments. Variability of soil nutrient content within a UMA was assessed by calculating the mean, range and standard deviation of values for all sites within each UMA containing at least 4 sampling sites. For sample sites with more than one value, the site average of the observed values was used in determining the UMA average, i.e. only one value per site was used. No consideration was given to the spatial distribution of sample sites within a UMA - possible biases in true UMA mean concentrations arising from distribution of vegetation, soil types, etc and how well they were represented in the selection of sample sites are not considered here.

Results are given in Table 2 for the 298 qualifying UMAs with N data and 245 UMAs with P data. In contrast to the measurement precision at a single sample site (Table 1), the average UMA soil nutrient content is less precisely determined (i.e. std dev/mean is greater). Not surprisingly, there is a much greater range of concentrations within a UMA and this has relevance for decisions focusing on sub-UMA scale areas. Whereas soil P content is relatively more precisely determined than soil N content at the scale of an individual sample site, P content is more variable than N content at the scale of a UMA.

Table 2. Unique mapping area averages for: mean, minimum, maximum P and N concentrations; standard deviation/mean; range and standard error. Only UMAs with 4 or more sample sites are included. Results are shown for all Queensland data (QLD) and GBRL region only (GBR).

	Average mean conc (%)	Average minimum conc (%)	Average maximum conc (%)	Average std dev/ mean	Average Range (%)	Average Std err	n
N (QLD)	0.13	0.065	0.23	0.38	0.16	0.02	298
N (GBR)	0.13	0.066	0.23	0.39	0.17	0.02	227
P (QLD)	0.055	0.024	0.112	0.48	0.09	0.01	245
P (GBR)	0.048	0.021	0.099	0.51	0.08	0.01	173

Ultimately, the utility of the ASRIS data set will depend on the spatial scale at which it is used and this can be assessed in terms of how well the ASRIS values correlate with individual site measurements and with UMA averages. A comparison between ASRIS and individual SALI site values is shown in Figure 2. A trend line is evident, but there is a significant spread of values about it (r^2 = 0.46 TN, 0.40 TP). The mean concentration values considering all data are for TP (0.056% SALI, 0.054% ASRIS) and for TN (0.12% SALI, 0.12% ASRIS).

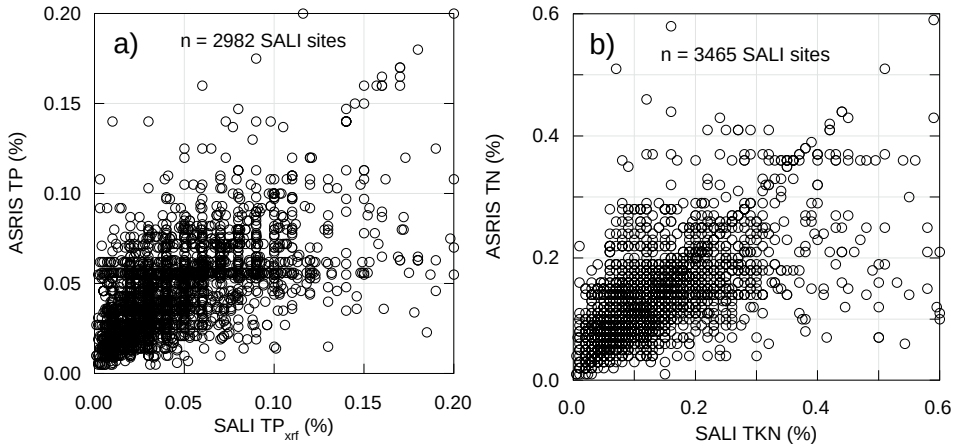


Figure 2. ASRIS soil TP a) and TN b) concentrations compared to SALI measured values at each SALI site in the GBRL catchments.

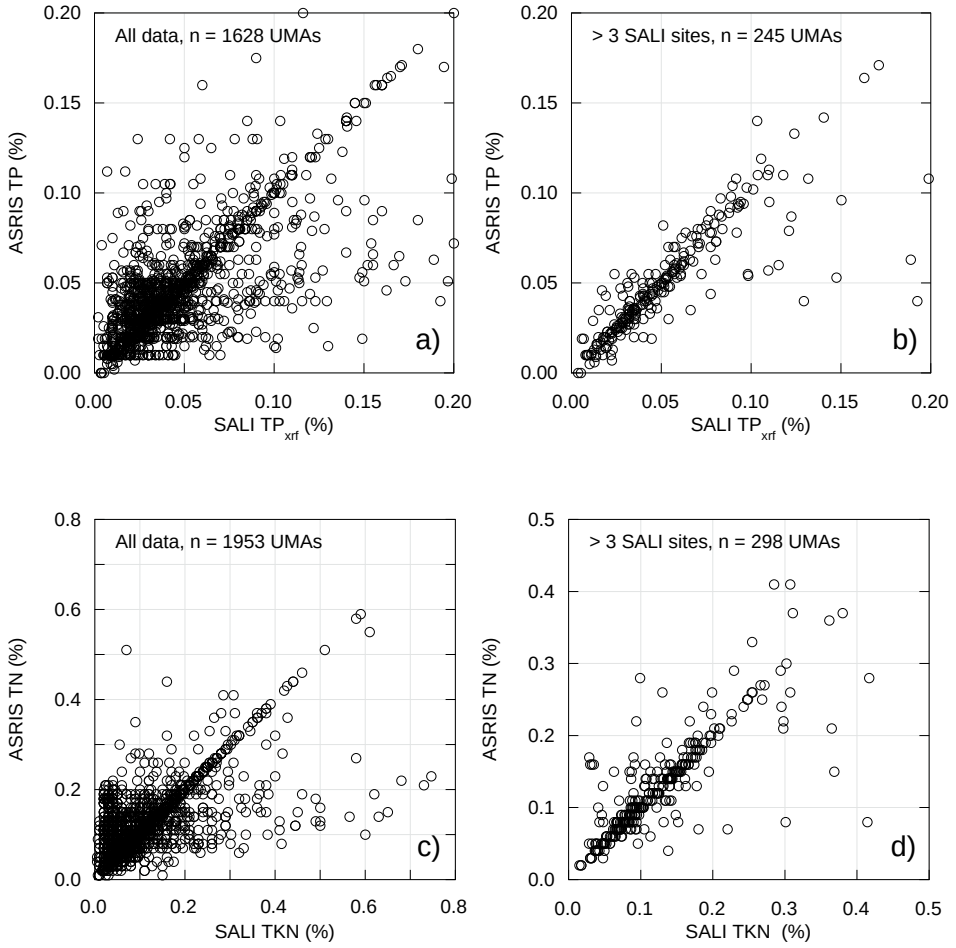


Figure 3. ASRIS surface soil nutrient content compared with corresponding mean SALI nutrient content for each unique mapping area. Soil total phosphorus for all UMAs a); total phosphorus for UMAs containing > 3 SALI sites b); total nitrogen for all UMAs c); total nitrogen for UMAs with > 3 SALI sites d).

The comparison improves at the UMA scale (Figure 3) if one considers the mean SALI concentration within a UMA. For UMAs containing > 3 SALI measurement sites the linear fit to the data is reasonable (r^2 = 0.65 TP, 0.67 TN) but decreases when all UMAs

containing a SALI measurement site are considered (r^2 = 0.41 TN, 0.49 TP). There appears to be a bias towards higher TN values in the ASRIS data (Figure 3c).

3.5 ASRIS area coverage of GBRL catchments

The ASRIS database has 14327 UMAs that are fully or partially within the GBRL catchments covering an area of 465,669 km^2. Roughly 4% of UMAs contain at least one sample site. This includes 867 UMAs with phosphorus data covering a total area of 107371 km^2 (23% of GBRL catchments) and 1015 UMAs with nitrogen data covering 128480 km^2 (28% of GBRL catchments). The mean and median areas of UMAs containing SALI measurement sites are 137 km^2 and 45 km^2, respectively. There is no correlation between UMA area and soil nutrient content and similarly the relative error (defined as (SALI - ASRIS)/SALI) does not correlate with UMA area although the range of relative errors is greater as UMA area decreases.

4. CONCLUSIONS

Surface soil nutrient content is not very sensitive to the sampling method. Both spot and bulk (composite) sample techniques yield very similar results. In other words, the spot measurements represent concentrations within the bulk sampling radius with acceptable accuracy.

The intrinsic natural variability in soil nutrient content coupled with the variability of nutrient content between sites within a UMA constrains our ability to predict differences between UMAs. UMA surface soil nitrogen and phosphorus content in the GBR catchments should be assumed to have relative uncertainties of at least ± 35% and ± 45% (std dev/mean), respectively. When considering individual sites within a UMA the uncertainty increases to approximately ± 75% for both TN and TP.

Both ASRIS and SALI datasets have the same mean values for TN and TP when averaged over all of the available data. This confirms that the mapping of SALI values into the ASRIS data set is sound. Care should be taken when interpreting catchment model predictions of particulate nutrient loads based on the ASRIS data set to ensure the appropriate level of uncertainty is applied.

The model used most frequently in GBRL catchments is Sednet/ANNEX. This model has a variable spatial resolution based on the size of subcatchments that drain into first order streams. As these subcatchments range in size from very small to 45 km^2 or larger the relative uncertainty of the predicted particulate nutrient loads can be expected to decrease from ± 75% to about ± 40%. Sednet/ANNEX typically is used to predict changes in nutrient fluxes arising from different management scenarios and to identify nutrient supply 'hot spots' within a catchment. The scale dependence of soil nutrient content uncertainty means that differences in nutrient loads from subcatchments > 45 km^2 will be more confidently discriminated than those for smaller subcatchments.

REFERENCES

Bormans, M., P. W. Ford, H. Hunter, A. Read, R. Dehayr, And C. Fellows. Catchment nutrients and sediment budgets: identification of knowledge gaps, National River Contaminants Program Project CLW54 Final Report, CSIRO Land and Water, Canberra, 2004.

Brodie, J., L. A. Mckergow, I. P. Prosser, M. J. Furnas, A. O. Hughes, And H. Hunter. Sources of Sediment and Nutrient Exports to the Great Barrier Reef World Heritage Area, Australian Centre for Tropical Freshwater Research Research Report 03/11, 2003.

Brough, D. M., J. Claridge, And M. J. Grundy. Soil and lancscape attributes: A report on the creation of a soil and landscape information system for Queensland, Department of Natural Resources and Mines Report QNRM06186, 2006.

Cogle, A. L., C. Carroll, And B. S. Sherman. The use of SedNet and ANNEX models to guide GBR catchment sediment and nutrient target setting Volume 1, Department of

Natural Resources and Mines Report QNRM06138, 2006. <http://www.wqonline.info/products/short_term_modelling.html>

Environment Australia. The Framework for Marine and Estuarine Water Qualtiy Protection, 2002. <http://www.deh.gov.au/coasts/pollution/cci/framework/part2.html>

Furnas, M. J. Catchment and corals: terrestrial runoff to the Great Barrier Reef, Australian Institute of Marine Science, Townsville, 2003.

Henderson, B., E. Bui, C. Moran, D. Simon, And P. Carlile. ASRIS: Continental-scale soil property predictions from point data, Technical Report 28/01, 2001.

Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick. Using spatial interpolation to construct a comprehensive archive of Australian climate data, *Environ. Modell. Softw.*, 16, 309–30, 2001.

McCulloch, M., S. Fallon, T. Wyndham, E. Hendy, J. Lough, and D. Barnes. Coral record of increased sediment flux to the inner Great Barrier Reef since European settlement, *Nature*, 421, 727–30, 2003.

McKenzie, N. J., D. W. Jacquier, D. J. Maschmedt, E. A. Griffin, and D. M. Brough. ASRIS Australian Soil Resource Information System Technical Specifications Version 1.5, 2005.<http://www.asris.csiro.au/downloads/ASRIS%20Tech%20Specs%201.5.pdf>

McKergow, L. A., I. P. Prosser, A. O. Hughes, and J. Brodie. Regional scale nutrient modelling: exports to the Great Barrier Reef World Heritage Area. *Marine Pollution Bulletin*, 51, 186-199, 2005.

Wilkinson, S., A. Henderson, Y. Chen, and B. Sherman. SedNet User Guide. Client Report, CSIRO Land and Water, Canberra, 2004.

Williams, D. M. Impacts of Terrestrial Run-off on the Great Barrier Reef World Heritage Area, CRC Reef Research Centre, Townsville, 2001.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Data Uncertainty Estimation Tool for Hydrology and Water Quality (DUET-H/WQ): Estimating Measurement Uncertainty for Monitoring and Modelling Applications

R.D. Harmel[a], D.R. Smith[b], K.W. King[c], R.M. Slade[d], and P. Smith[e]

[a] *USDA-ARS, 808 E. Blackland Rd., Temple, TX 76502, USA*
(daren.harmel@ars.usda.gov)
[b] *USDA-ARS, West Lafayette, IN*
[c] *USDA-ARS, Colombus, OH*
[d] *USGS (Retired), Austin, TX*
[e] *Texas A&M University, College Station, TX*

Abstract: In spite of advanced modelling capabilities, hydrologic and water quality data remain vital for scientific assessment, management, decision-making, and modelling. Although uncertainty in measured data affects all of these applications, measurement uncertainty is typically ignored in monitoring projects. To change this, we published an uncertainty estimation framework for measured discharge and water quality data in 2006. From this framework, the Data Uncertainty Estimation Tool for Hydrology and Water Quality (DUET-H/WQ) was designed as a user-friendly tool to facilitate uncertainty estimation. DUET-H/WQ provides published uncertainty estimates for data collection procedures and then estimates the uncertainty within each procedural category as well as the cumulative uncertainty. The software estimates uncertainty for individual measured values as contributed by measurement and data processing and management. It does not account for uncertainties associated with spatial variability or influences of scale. The broad applicability of DUET-H/WQ was established by its application to data collected in five monitoring projects from a variety of watershed conditions. Results indicated that uncertainty in individual values was typically least for discharge, higher for sediment and dissolved N and P, and higher yet for total N and P. The uncertainty inherent in measured data has numerous economic, societal, and environmental implications; therefore, scientists can no longer ignore measurement uncertainty in data collection and reporting. It is our hope that DUET-H/WQ will contribute to making uncertainty estimation a routine component in hydrologic and water quality monitoring projects.

Measurement uncertainty also has important implications in modelling applications. The impact of uncertainty in model calibration and validation data is commonly discussed, but rarely included, in the evaluation of model accuracy. In order to change this oversight, we recently modified several goodness-of-fit indicators to incorporate measurement uncertainty into model calibration and validation. A similar method is currently being tested that incorporates both measurement and model uncertainty into model goodness-of-fit evaluation.

Keywords: *Data collection; discharge; water quality; model calibration; model validation.*

1. INTRODUCTION

In spite of advanced modeling capabilities, natural resource decision-making relies on measured environmental data, for which an understanding of data uncertainty is important in many aspects (Brown et al., 2005). Specifically for water resources, measured

hydrologic and water quality data remain vital for scientific assessment, management, and modeling (Silberstein, 2006). First, optimal water quality monitoring can only be achieved if measurement uncertainty and alternatives to reduce uncertainty are understood and considered in project design and implementation (Beven, 2006a; Harmel et al., 2006b; Rode and Suhr, 2007). Second, enhanced decision-making and stakeholder understanding can only be fully realized if measurement uncertainty is estimated and adequately communicated to other scientists, modelers, public interest groups, regulators, and elected officials (Collins et al., 2000; Bonta and Cleland, 2003; Reckhow, 2003; Nature, 2005; Beven, 2006a; Pappenberger and Beven, 2006). Similarly, analysis of uncertainty in measured data, which drive model calibration and validation, improves model application and enhances decisions based on modeling results (Reckhow, 1994; Kavetski et al., 2002; Pappenberger and Beven, 2006; Beven, 2006b; Shirmohammadi et al., 2006; Harmel and Smith, 2007).

According to Beven (2006a), the first step in advancing hydrologic and water quality science related to measurement uncertainty is determining realistic methods of representing that uncertainty. The value of uncertainty estimates, as well as the scientific integrity of communicating measurement uncertainty, prompted Harmel et al. (2006a) to make this initial step by developing a framework for quantifying the uncertainty in measured discharge and water quality (chemical constituent) data collected at the field and small watershed scale. That framework was then applied to estimate the cumulative uncertainty for a variety of arbitrary "data quality" scenarios. While several researchers (e.g. Gentry et al., 2007; McCarthy et al., 2008; Keener et al., 2008) have accepted and applied this framework, it can be cumbersome to apply to data sets with multiple values for multiple parameters. Therefore, we developed the Data Uncertainty Estimation Tool for Hydrology and Water Quality (DUET-H/WQ) and applied it to estimate the uncertainty in measured discharge and water quality data collected in several monitoring projects.

2. METHODS

2.1 Development of DUET-H/WQ

DUET-H/WQ (Harmel et al., 2008) was developed to be a user-friendly application of an existing uncertainty estimation framework for discharge and constituent flux measurements. The framework (Harmel et al., 2006a) contains two foundational components: 1) procedural categories within which to classify monitoring methods and 2) an established method for estimating cumulative uncertainty in individual measured values resulting from individual steps within procedural categories.

The first component established four procedural categories: Discharge Measurement, Sample Collection, Sample Preservation/Storage, and Laboratory Analysis. An additional category, Data Processing and Management, was later added to include uncertainty introduced by missing values, assumptions made to estimate missing values, and mistakes in data management and reporting.

The second component presented the Topping (1972) root mean square error (RMSE) propagation calculation and applied it to estimate cumulative or "combined" uncertainty in measured hydrology and water quality data. The RMSE method was selected for the framework because it is simple and has been widely applied to estimate cumulative uncertainty in individual discharge measurements (Cooper, 2002; Sauer and Meyer, 1992), in sediment volume estimates (Allmendinger et al., 2007), and in pesticide analytical methods (Cuadros-Rodriquez et al., 2002). Within the uncertainty estimation framework, the uncertainty from each step with each procedural category is propagated to produce a realistic uncertainty estimate, which is best termed "cumulative probable uncertainty" as represented by the probable error (eq. 1),

$$EP = \sqrt{\sum \left(E_Q{}^2 + E_C{}^2 + E_{PS}{}^2 + E_A{}^2 + E_{DPM}{}^2 \right)} \quad [1]$$

where: EP = the probable error (± %), E_Q = uncertainty in discharge measurement (±%), E_C = uncertainty in sample collection (±%), E_{PS} = uncertainty in sample preservation/storage (± %), E_A = uncertainty in laboratory analysis (± %), and E_{DPM} = uncertainty in data processing and management (±%).

The RMSE calculation for individual measured values assumes that uncertainty is symmetric about the value and thus bi-directional with equal likelihood of over- and under-estimation and that errors for each procedural step are independent (Topping, 1972). Thus in the absence of contrary data, uncertainties for procedural steps are assumed to be independent and the covariance is omitted.

2.2 Application of DUET-H/WQ

The uncertainty estimation framework was initially applied to arbitrary best case, worst case, and typical scenarios in Harmel et al. (2006a). While those results can be used to establish reasonable uncertainty estimates in the absence of project-specific information, it was also important to apply the framework to real-world monitoring data to quantify uncertainty in actual field data. Thus, the DUET-H/WQ software tool based on this framework was applied to measured data from small watersheds in Texas (Riesel, Hamilton, and Austin), Indiana (Waterloo), and Ohio (Centerburg) (Harmel et al., 2008). These sites were selected to represent a wide range of monitoring conditions with respect to hydrologic setting, land use, watershed size, and field and laboratory techniques to demonstrate its broad applicability. For each measured data set, DUET-H/WQ was used to estimate the uncertainty in each step within each procedural category and to estimate the cumulative uncertainty for individual measured values. Specifically, uncertainties were estimated for individual discharge, total suspended solids (TSS), NO_3-N, PO_4-P, total N, and total P data.

A Beta version completed in December 2007 was used in the present analyses. This version is available at ftp.brc.tamus.edu/pub/outgoing/gmitchell/DataUncertaintyEstimationTool/. It is important to note that the DUET-H/WQ framework basis was developed for discharge, sediment, and nutrient data collection from small watersheds. Therefore, application to basin scale data and/or additional constituents may require appropriate adjustment.

The first step in applying DUET-H/WQ to estimate the uncertainty for individual measured values is to enter the individual data collection steps utilized to collect that data. To accomplish this, the user selects the appropriate techniques used and/or conditions encountered in the appropriate DUET-H/WQ lookup tables for the discharge measurement, sample collection, sample preservation/storage, and laboratory analysis procedural categories (Figs. 1-4). For these procedural categories, a DUET-H/WQ lookup table lists the common techniques utilized and monitoring conditions encountered along with published uncertainty estimates. These tables are based on Harmel et al. (2006a), which provides a detailed description of methodologies and associated uncertainties within each procedural category. Then, the user selects an appropriate uncertainty estimate from published uncertainty data displayed by the software. The user can adjust these uncertainty estimates based on project-specific information and/or professional experience. DUET-H/WQ then calculates the uncertainty introduced by each procedural category (eq. 1). The software then allows the user to input the uncertainty contributed by project-specific data processing and management issues (Fig. 5). Finally, DUET-H/WQ calculates the cumulative uncertainty for individual discharge, concentration, or load values (eq. 1). The user can choose to either apply that same estimated uncertainty to other data collected with the same procedure and under similar conditions or to repeat the uncertainty estimation procedure for other measured values.

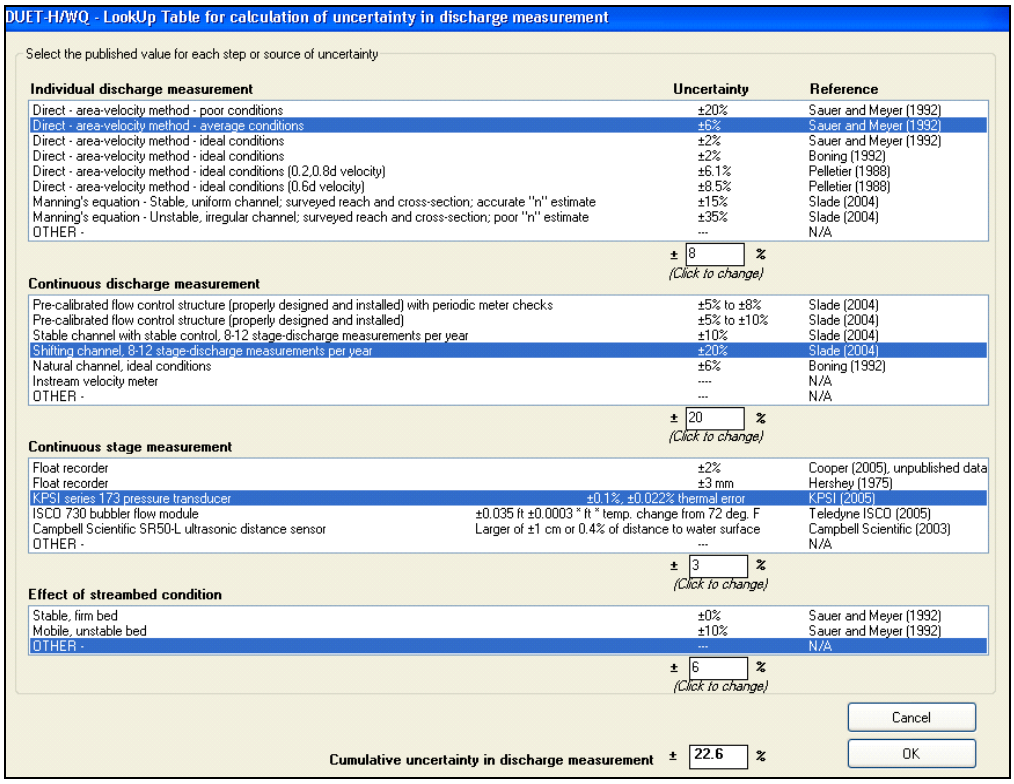


Figure 1. Example DUET-H/WQ Lookup Table for discharge measurement.

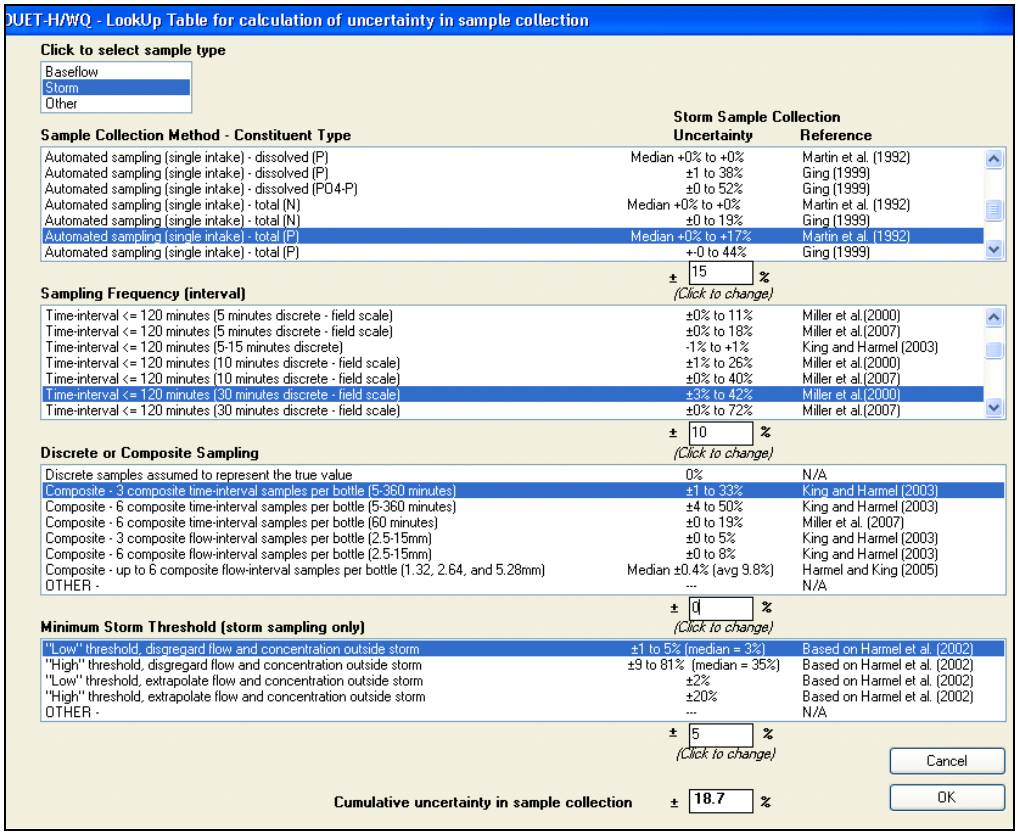


Figure 2. Example DUET-H/WQ Lookup Table for sample collection.

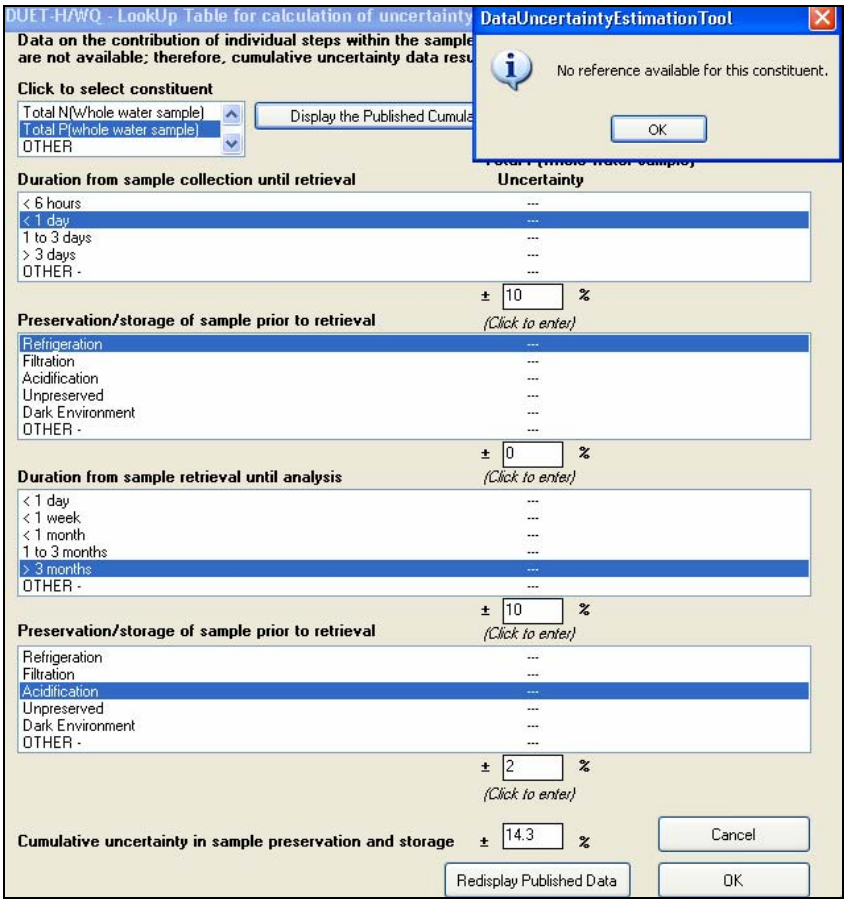


Figure 3. Example DUET-H/WQ Lookup Table for sample preservation/storage.

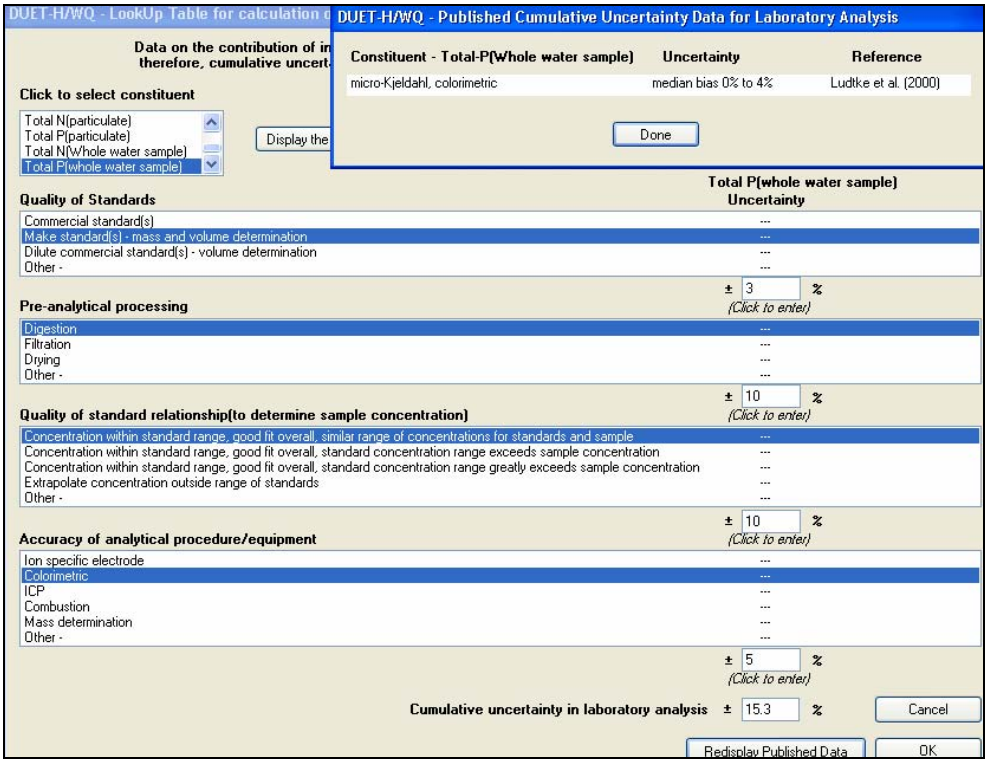


Figure 4. Example DUET-H/WQ Lookup Table for laboratory analysis.

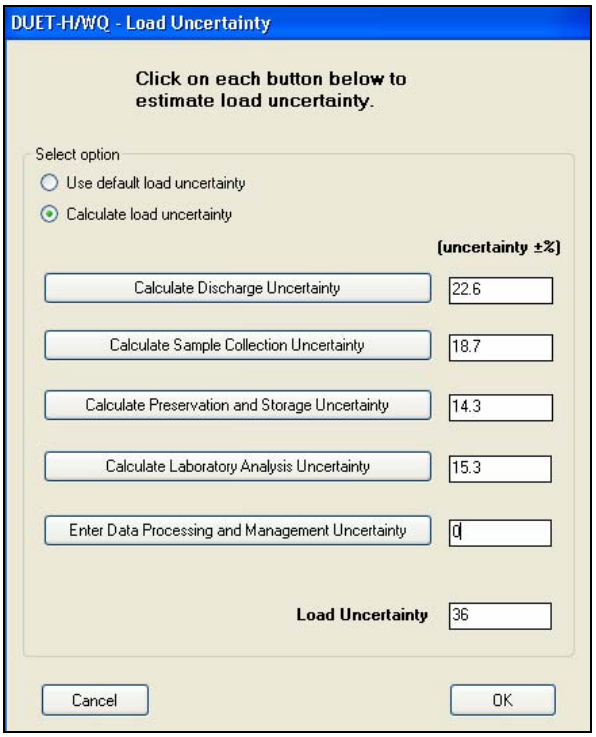


Figure 5. Example DUET-H/WQ summary calculation for load uncertainty.

3. RESULTS AND DISCUSSION

In the five monitoring projects, uncertainties were estimated for measured data from a total of 131 storm events. The uncertainty in discharge measurements for individual storm events ranged from 7-27% with a median of 14% (Fig. 6). For TSS, the load uncertainty ranged from 15-35% with a median of 20%, and the uncertainty in concentrations ranged from 12-26% with a median of 18%. The uncertainty in TSS loads was typically less than other constituents because of limited post-collection transformation, relatively simple analytical procedures, and high concentration values. No equipment malfunction, which would have increased the uncertainty dramatically, occurred related to discharge or TSS data.

It is important to note that the uncertainty associated with measured concentrations is always less than or equal to that of measured loads. This reduction occurs because the discharge measurement procedural category (and it associated steps and uncertainty contribution) is irrelevant in concentration determination. The resulting reduction ranged from 10-37% depending on the relative magnitudes of other sources of uncertainty.

Little difference in uncertainty was evident between dissolved NO_3-N and PO_4-P (Fig. 6). The median uncertainty in NO_3-N and PO_4-P flux measured for individual events were 22-23% for loads and 17-19% for concentrations. Much higher uncertainty (up to 104%) occasionally occurred due to extreme high flows and missing samples. The reduction in uncertainty created by the irrelevance of discharge measurement in NO_3-N and PO_4-P concentrations ranged from 1-35%. The uncertainty in dissolved NO_3-N and PO_4-P loads was typically higher than in TSS loads because of post-collection transformation potential, more complex analytical procedures, and lower concentration values, which counteracted reduced difficulty in sample collection for dissolved constituents.

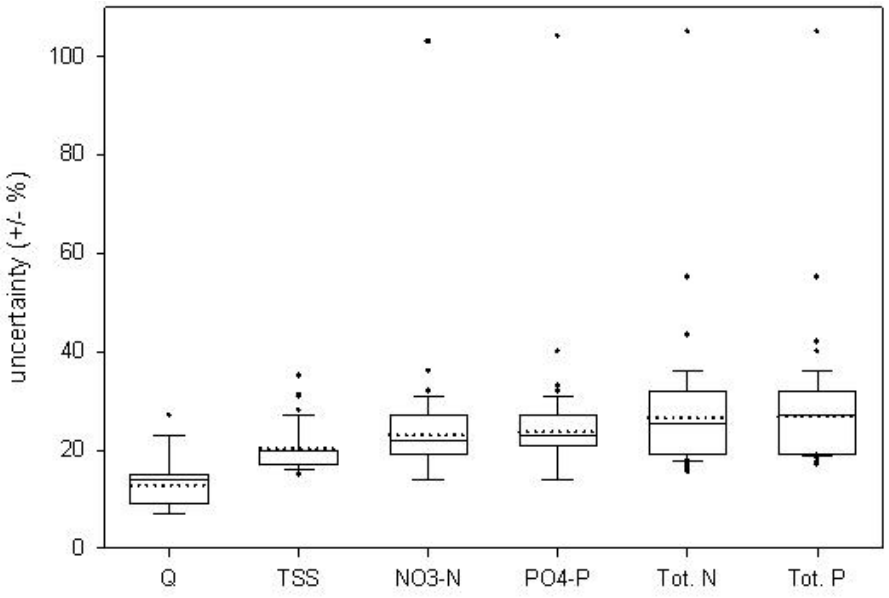


Figure 6. Uncertainty in individual storm discharge and constituent load values.

Similar to dissolved NO_3-N and PO_4-P, little difference in uncertainty occurred between total N and total P (Fig. 6). For individual loads, the uncertainty ranged from 15-105% (median = 25-27%) with increased uncertainties again occurring due to extreme high flows and missing samples. The uncertainties for measured concentrations ranged from 14-104% (median = 23-24%), which represents 0-25% uncertainty reduction compared to load uncertainty. The uncertainty in total N and P loads was typically higher than for TSS loads because of more complex analytical procedures. Total N and P load uncertainty was also higher than for dissolved loads because of increased difficulty in collecting representative particulate samples and additional analytical steps when total N and P were determined by summing dissolved and particulate fractions.

4. INCORPORATION OF UNCERTAINTY INTO MODEL EVALUATION

With uncertainty estimates for model calibration and validation data, this uncertainty can be used to enhance model evaluation. A correction factor was recently developed by Harmel and Smith (2007) to modify the error or deviation term (eq. 2) in several common goodness-of-fit indicators (Nash Sutcliffe coefficient of efficiency, E_{NS}; Index of Agreement, *d*; root mean square error, *RMSE*; mean absolute error, *MAE*). Although this correction factor (eq. 3) better represents model goodness-of-fit in the presence of measurement uncertainty, it does not consider the effect of model (prediction) uncertainty.

$$e_i = O_i - P_i \quad [2]$$

where: e_i = deviation between paired observed (O_i) and predicted (P_i) data.

$$eu2_i = \frac{CF_i}{0.5} \times (O_i - P_i) \quad [3]$$

where: $eu2_i$ = modified deviation considering measurement uncertainty, and CF_i = correction factor based on the probability distribution of each measured value.

In the presence of measurement and model uncertainty, it is more appropriate to evaluate models considering both sources of uncertainty. Thus, correction factors for measurement and prediction uncertainty were developed to enhance goodness-of-fit evaluation by producing realistic estimates of the deviations between measured values and model predictions (eq. 4).

$$eu3_i = \frac{\frac{CF(O_i)}{A} + \frac{CF(P_i)}{B}}{2.0} \times (O_i - P_i) \quad [4]$$

where: $eu3_i$ = modified deviation considering measurement and model uncertainty, $CF(O_i)$ = correction factor based on the probability distribution for each measured value, $CF(P_i)$ = correction factor based on the probability distribution for each predicted value, A = the value of the 1-sided probability of the measurement uncertainty distribution on the side corresponding to the predicted value, and B = the value of the 1-sided probability of the model uncertainty distribution on the side corresponding to the observed value.

These correction factors were developed based on the theory that the deviation between measured and predicted values should be adjusted accordingly to represent their relation to the other's uncertainty distribution (Fig. 7). The degree of overlap between corresponding measurement and predicted probability distribution functions (pdfs) is indicative of the model's predictive ability (Haan et al., 1995). The correction factors are currently being tested for several assumed uncertainty distributions (normal, symmetrical triangular, and uniform) by application to measured and corresponding predicted data sets.

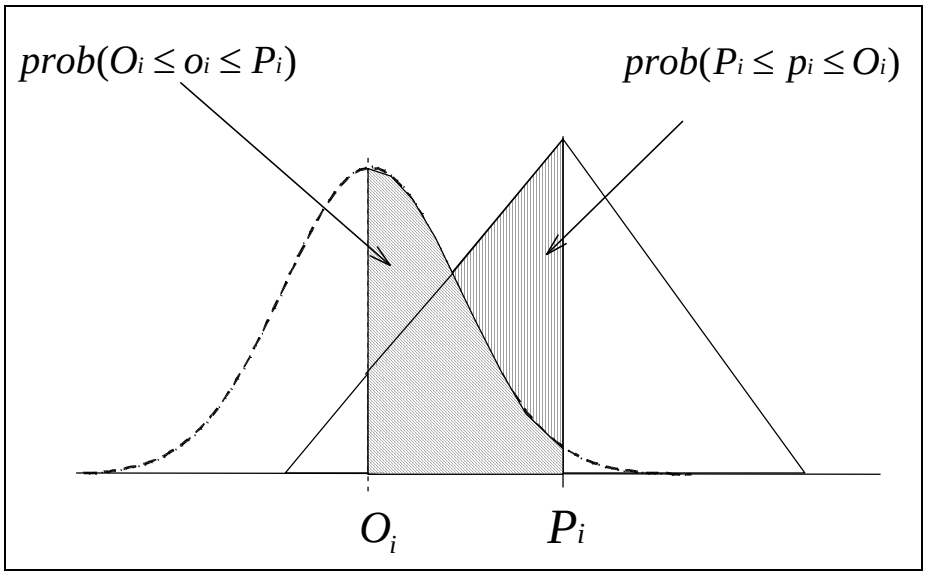


Figure 7. Graphical representation of correction factors that consider both measurement and model uncertainty (for the case $O_i < P_i$).

5. CONCLUSIONS

The multiple benefits of uncertainty estimates corresponding to hydrologic and water quality data (specifically improved monitoring design, enhanced decision-making, and improved model application and understanding) will not be fully realized without a

relatively simple, straight-forward procedure to estimate uncertainty. The Data Uncertainty Estimation Tool for Hydrology and Water Quality (DUET-H/WQ) was developed with these benefits in mind to be a user-friendly tool for data collectors and data users to estimate uncertainty in discharge and water quality data. DUET-H/WQ was designed to estimate the uncertainty contributed by measurement and data processing and management in ***individual*** values. As such, it does not directly account for influences of spatial variability (number and/or locations of sampling points within watersheds), changes in scale, or differences in sampling strategies (such as intensive storm sampling versus monthly grab sampling). DUET-H/WQ can, however, be used with GIS and modelling tools to address such issues.

Because of the economic, societal, and environmental implications of measurement uncertainty, it is important that uncertainty estimation become a routine procedure in hydrologic and water quality data collection and reporting. It is our hope DUET-H/WQ will contribute to this advancement.

The broad applicability of DUET-H/WQ was recently established by its application to data collected in five monitoring projects from a variety of watershed conditions. This application was important because the initial framework development and evaluation estimated the uncertainty only for arbitrary "data quality" scenarios not actual measured data. When applied to individual measured values the estimated uncertainty typically followed a predictable pattern (Q < TSS < dissolved N and P < total N and P).

With estimates of uncertainty for discharge and water quality measurements, which are often used as model calibration and validation data, the effects of measurement uncertainty on model accuracy can now be incorporated into model evaluation. Similarly, a method to incorporate both measurement and model uncertainty into model calibration and validation has been developed and will soon be available once testing is completed.

REFERENCES

Allmendinger, N.E., J.E. Pizzuto, G.E. Moglen, and M. Lewicki. A sediment budget for an urbanizing watershed, 1951-1996, Montgomery County, Maryland, USA. *J. Amer. Water Resources Assoc.* 43(6), 1483-1498, 2007.

Beven, K. On undermining the science? *Hydrol. Process.* 20, 3141-3146, 2006a.

Beven, K. A manifesto for the equifinality thesis. *J. Hydrol.* 320, 18-36, 2006b.

Bonta, J.V., and B. Cleland. Incorporating natural variability, uncertainty, and risk into water quality evaluations using duration curves. *J. Amer. Water Resources Assoc.* 39(6), 1481-1496, 2003.

Brown, J.D., G.B. Heuvelink, and J.C. Refsgaard. An intergrated methodology for recording uncertainties about environmental data. *Water Sci. Tech.* 52(6), 153-160, 2005.

Cooper, R.J. Chemical Denudation in the Proglacial Zone of Finsterwalderbreen, Svalbard. PhD Thesis, University of Bristol, Bristol, UK. 184 pg, 2002.

Collins, G., J.N. Kremer, and I. Valiela. Assessing uncertainty in estimates of nitrogen loading to estuaries for research, planning, and risk assessment. *Environ. Mgmt.* 25(6), 635-645, 2000.

Cuadros-Rodriquez, L., M.E. Hernandez Torres, E. Almansa Lopez, F.J. Egea Gonzalez, F.J. Arrebola Liebanas, and J.L. Martinez Vidal. Assessment of uncertainty in pesticide multiresidue analytical methods: Main sources and estimation. *Anal. Chim. Acta* 454(2), 297-314, 2002.

Gentry, L.E., M.B. David, T.V. Royer, C.A. Mitchell, and K.M. Starks. Phosphorus transport pathways to streams in tile-drained agricultural watersheds. *J. Environ. Qual.* 36, 408-415, 2007.

Haan, C.T., B. Allred, D.E. Storm, G.J. Sabbagh, and S. Prahhu. Statistical procedure for evaluating hydrologic/water quality models. *Trans. ASAE* 38(3), 725-733, 1995.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, and J.G. Arnold. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Trans. ASABE* 49(3), 689-701, 2006a.

Harmel, R.D., K.W. King, B.E. Haggard, D.G. Wren, and J.M. Sheridan. Practical guidance for discharge and water quality data collection on small watersheds. *Trans. ASABE* 49(4), 937-948, 2006b.

Harmel, R.D. and P.K. Smith. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *J. Hydrol.* 337, 326-336, 2007.

Harmel, R.D., D.R. Smith., K.W. King, and R.M. Slade. Estimating the uncertainty in individual and aggregated discharge and water quality measurements. *Water Resources Res.* (In Review), 2008.

Kavetski, D., S.W. Franks, and G. Kuczera. Confronting Input Uncertainty in Environmental Modelling in Calibration of Watershed Models, AGU Water Science and Applications Series, S. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte (editors). Vol. 6, pg. 49-68, 2002.

Keener, V.W., K.T. Ingram, B. Jacobson, and J.W. Jones. Effects of El-Niño/Southern Oscillation on simulated phosphorus loading in South Florida. *Trans. ASABE* 50(6), 2081-2089, 2008.

McCarthy, D.T., A. Deletic, V.G. Mitchell, T.D. Fletcher, and C. Diaper. Uncertainties in stormwater E. coli levels. *Water Res.* 42, 1812-1824, 2008.

Nature. Responding to uncertainty. 437(7055), 1, 2005.

Pappenberger F., and K.J. Beven. Ignorance is bliss: 7 reasons not to use uncertainty analysis. *Water Resources Res.* 42(5), doi:10.1029/2005W05302, 2006.

Reckhow, K.H. Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. *Ecological Modelling* 72, 1-20, 1994.

Reckhow, K.H. On the need for uncertainty assessment in TMDL modeling and implementation. *J. Water Resource Planning Mgmt.* 129(4), 245-246, 2003.

Rode, M. and U. Suhr. Uncertainties in selected river water quality data. *Hydrol. Earth System Sci.* 11, 863-874, 2007.

Sauer, V.B., and R.W. Meyer. Determination of error in individual discharge measurements. Open File Report 92-144. US Geological Survey, Washington, D.C., 1992.

Shirmohammadi, A., I. Chaubey, R.D. Harmel, D.D. Bosch, R. Muñoz-Carpena, C. Dharmasri, A. Sexton, M. Arabi, M.L. Wolfe, J. Frankenberger, C. Graff, and T.M. Sohrabi. Uncertainty in TMDL Models. *Trans. ASABE* 49(4), 1033-1049, 2006.

Silberstein, R.P. Hydrological models are so good, do we still need data? *Environ. Modelling Software* 21, 1340-1352, 2006.

Topping, J. Errors of Observation and Their Treatment. 4th ed. London, U.K.: Chapman and Hall, 1972.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Catchment-Scale Water Quality Modelling and Integration of Collateral Information

Lachlan T.H. Newham[a], John J. Drewry[a,b,c] and Baihua Fu[a]

[a] *The Fenner School of Environment and Society, The Australian National University, Canberra ACT 0200, Australia (lachlan.newham@anu.edu.au)*

[b] *Cooperative Research Centre for Landscape Environments and Mineral Exploration, The Australian National University, Canberra, ACT, Australia.*[c] *Current address: Mackay Whitsunday Natural Resource Management Group*

Abstract: This paper describes progress on techniques in the modelling of sediments and nutrients to inform management effort. It describes a catchment-scale model known as CatchMODS, and the integration of knowledge from a range of collateral sources to inform its development and application. A full description of the model, including its hydrologic, sediment and nutrient modules, is presented. The paper describes how knowledge from field-based studies is incorporated to support the modelling. Knowledge is generated from: (i) water quality monitoring; (ii) sediment tracing of critical source areas and relative subcatchment contributions (using geochemical, and mineralogy techniques); and (iii) field investigation, particularly landscape measurement. This knowledge is used for model conceptualisation, correct parameterisation and performance testing.

Keywords: Water quality modelling; collateral knowledge integration; water quality monitoring; sediment tracing.

1. INTRODUCTION

This paper describes progress on techniques of modelling sediment and nutrients to inform management effort. A catchment-scale modelling framework known as the Catchment-Scale Management of Distributed Sources (CatchMODS) model is at the core of a research program that integrates modelling and field assessment techniques. The latest structure and algorithms of the CatchMODS model are described along with efforts to inform the model development process through integration of collateral knowledge. The structure of the modelling framework is informed by studies of pollutant sources and fate processes. These studies, including monitoring, sediment tracing and field investigation, provide data for correct parameterisation and model performance testing. A discussion of the integration process is made and opportunities to improve the representation of pollutant processes are described.

2. CATCHMENT-SCALE MANAGEMENT OF DISTRIBUTED SOURCES MODEL

The CatchMODS model framework is designed to simulate current conditions and the effects of management activities on the quality of receiving waters at catchment scales (Newham et al. 2004). The framework integrates hydrologic, sediment and nutrient export models with a simple economics submodel, to enable the development and evaluation of the impact of management strategies for reducing nutrient and sediment yields to waterways. The model has been applied in several catchments for various management and research purposes. It was initially developed for application in the Ben Chifley Dam catchment of NSW and has since been applied in several other Australian catchments including the Avon-Richardson in north central Victoria and Cox Creek in the Adelaide Hills, SA.

CatchMODS is based on a series of linked river reaches and associated subcatchment areas. The modelling is lumped at these stream reach and subcatchment units (Newham et al. 2004). The topology of the stream network enables the downstream routing of pollutants with the individual submodels each simulating processes of pollutant attenuation and/or deposition. Reaches and subcatchments are disaggregated using an area threshold to define reaches. The topology of the stream network defines the associated subcatchment areas. The size of a subcatchment in a typical application of CatchMODS averages 30 km^2. Further details on the representation of spatial variability in the model and network topology are provided in Newham et al. (2004). CatchMODS is structured around its hydrologic, economic, sediment and nutrient export submodels which are described in the following sections.

2.1 Hydrology

The IHACRES rainfall-runoff model (Jakeman et al. 1990) is used to estimate both surface and subsurface discharge in CatchMODS. It is applied at a daily timestep with its temperature and rainfall inputs linearly scaled according to subcatchment mean rainfall and mean elevation, respectively. A full description of the IHACRES model and its regionalisation is outside of the scope of this paper but Post and Jakeman (1999) provide an example of the regionalisation of IHACRES.

The modelled daily discharge from the IHACRES model is used to generate several flow-related statistics for each subcatchment. The mean annual flow (Q_{MAF}) for a subcatchment is estimated as sum of the daily flows during the period of record divided by the number of days of record simulated. The bankfull discharge (Q_{BF}) is estimated as the flow corresponding to a specified average recurrence interval. For overbank flows (Q_{OB}), the median flow volume is used for flows greater than Q_{BF}. Baseflow volumes are estimated for each subcatchment using the baseflow filter method of Croke et al. (2001). The local baseflow volume is calculated for each subcatchment i

$$Q_{BF_{local},i} = Q_{BF,i} - \sum_{u} Q_{BF,u}$$

where $Q_{BF,i}$ is the baseflow from subcatchment i ; and $\sum_{u} Q_{BF,u}$ is the sum of the upstream baseflow inputs. A similar approach is used to estimate $Q_{QF_{local},i}$ – the quickflow from subcatchment i. Scenarios of the impact of climate changes are implemented by substituting the climate inputs to IHACRES. Historic or stochastic inputs are used. The effects cascade through the modelling framework via the flow-related statistics.

2.2 Sediment

The sediment submodel of CatchMODS is substantially modified from the SedNet model (Prosser et al. 2001) but retains several of its underlying algorithms. The focus of CatchMODS is on the simulation of the suspended sediment (SS) fraction. This reflects the importance of SS as a source and transport medium for many common stream pollutants e.g. phosphorus. It also enables investigation of contemporaneous SS fluxes and management effects over the more historic perspective of SedNet. SS is defined as the less than 63µm diameter sized particle fraction. Sediment inputs are estimated from hillslope, gully and streambank erosion sources. Road-derived sediment inputs to a subcatchment are soon to be estimated via the WARSEM model (Dubé et al. 2004) with minor modifications made to that model to enable its inclusion in the structure of CatchMODS, see Fu et al. (2007) for an example of the application of the WARSEM model, and Fu et al. (submitted) for a review of road erosion modelling.

2.2.1 Hillslope erosion

Hillslope erosion is estimated via use of the RUSLE (Renard et al. 1997). The application of the RUSLE is lumped at the subcatchment scale to ease scenario construction. The

combined rainfall (R), soil erodibility (K) and length and slope (LS) factors are estimated and multiplied on a cell-by-cell basis. A mean value of the combined factors is calculated for each subcatchment

$$\overline{\lambda_i} = \sum_{k=1}^{n} R_k K_k LS_k \Big/ c$$

where $\overline{\lambda_i}$ is the mean value of the combined factors for subcatchment i; k is the cell; and c is the total number of cells in the subcatchment. The estimated hillslope erosion for each subcatchment (H_i) is the product of $\overline{\lambda_i}$, the subcatchment cover factor ($\overline{C_i}$) and a sediment delivery ratio (SDR), R and the area of subcatchment i (A_i)

$$H_i = \overline{\lambda_i}\,\overline{C_i} \times R \times A_i$$

$\overline{C_i}$ is estimated as

$$\overline{C_i} = \sum_{j} C_j A_j \Big/ A_i$$

where C_j is the crop factor for land use j; and A_j is the area of land use j. A SDR scales the hillslope erosion estimate from the RUSLE to the quantity of sediment that is delivered to the stream network. The SDR takes a value between 0 and 1. In the application of CatchMODS in the Avon-Richardson catchment, a subcatchment-specific sediment delivery ratio is being trialled. Normalised indices of sediment transport based on Prosser and Rustomji (2000), Bhattarai and Dutta (2007) and Warren et al. (2005) are being trialled to scale subcatchment hillslope inputs. Management change scenarios that affect the hillslope erosion estimates are implemented via changes in land use proportions for each subcatchment. Land use changes are simulated via changes of the $\overline{C_i}$ value.

2.2.2 Gully erosion

Gully erosion is assumed to be sourced from erosion of the sidewalls of permanent gullies. The mass of sediment derived from gully erosion is estimated as the product of the volume of gully erosion and the bulk density of the eroded sediment. Gully erosion is broken into multiple categories to reflect differences in sediment production and physical dimensions amongst the gullies of a catchment. The estimated gully erosion for each subcatchment is

$$G_i = \sum_{s} E_{s,i}\, 2 r_s d_s \rho \varpi_g$$

where $E_{s,i}$ is the effective length of gullies of severity class s in subcatchment i; r_s is the annual average lateral erosion rate of gullies of severity class s; d_s is the depth of gullies of severity class s; ρ is the bulk density of eroded sediments and ϖ_g is the proportion of suspended sediment in the eroded gully material. Management change scenarios are implemented via reducing the length of gully erosion for each severity class for subcatchments where gully erosion control works are to be implemented. An effectiveness factor is introduced to reflect the anticipated or measured effectiveness of the control works.

$$E_{s,i} = L_{s,i} - \nu L_{r,i}$$

where $L_{s,i}$ is the length of gullies of severity class s in subcatchment i; ν is the effectiveness factor for the control works; and $L_{r,i}$ is the length of gully control works.

2.2.3 Streambank erosion

Streambank erosion is estimated using the algorithm of Prosser et al. (2001)

$$S_i = aQ_{BF}^b L_{r,i} d_r \rho \varpi_s \left(1 - L_{v,i} / L_{r,i}\right)$$

where a and b are empirical parameters; $L_{r,i}$ is the total stream length in subcatchment i; d_r is the mean streambank height; ϖ_s is the proportion of suspended sediment in the eroded gully material; and $L_{v,i}$ is the non eroding (vegetated) reach length.

2.2.4 Sediment routing and floodplain deposition

The estimation of sediment deposition onto the floodplain uses the algorithm of Prosser et al. (2001). Total sediment contributions to a reach are estimated as

$$X_{in,i} = H_i + G_i + S_i + X_{up}$$

Routing to the downstream subcatchment ($X_{out,i}$) is estimated as

$$X_{out,i} = X_{in,i} - \left(\frac{Q_{OB,i}}{Q_{MAF,i}} X_{in,i} \left(1 - \exp\left(\frac{-vA_f}{Q_{OB,i}} \right) \right) \right)$$

where v is the settling velocity of sediment particles on the floodplain.

2.3 Nutrients

Dissolved and particulate nutrient fractions are simulated separately in CatchMODS. The P and N submodels of CatchMODS are identical in structure. A generation-rate-based or flow-based approach (or a combination of the two) may be used for the simulation of dissolved nutrients. The attenuation of dissolved nutrients though the system is simulated using a simple exponential function.

2.3.1 Particulate nutrient input

Particulate nutrient input to subcatchment i is estimated as

$$P_i = h \sum HS_{i,j} C_j + S_i C_s + G_i C_g$$

where h is a nutrient enrichment factor; and C_j, C_s, C_g are the surface soil, bulk streambank and bulk gully nutrient concentrations, respectively. Samples for the determination of soil nutrient concentration are taken from the top 5cm of the soil profile. The h factor is introduced to represent the effects of the preferential erosion and delivery of finer soil particles that have higher nutrient concentrations on the hillslope. The factor is applied across all subcatchments and land uses in the absence of knowledge of nutrient enrichment for each specific land use and soil combination.

2.3.2 Dissolved nutrient input

Dissolved nutrient input to subcatchment i is estimated as

$$DN_i = (1 - A_{GR})(oQ_{BF_{local},i} + pQ_{QF_{local},i}) + \sum_j A_{GR} G_j$$

where A_{GR} is the area of subcatchment modelled via a generation rates approach, o is the baseflow dissolved nutrient concentration (g/mL); p is the quickflow dissolved nutrient concentration (g/mL); and G_r is the areal generation rate for land use j. The inclusion of a generation rate for specific landuses enables use of the model in areas where nutrient loss per unit area is likely to be much greater than would otherwise be estimated by CatchMODS algorithms e.g. in dairy farming areas. A generation rate for specific land uses also enables use of the model in areas where field-collected water quality monitoring data is unavailable for model parameterisation.

2.3.3 Nutrient routing

Nutrient routing includes an estimate of the attenuation of dissolved nutrients through a reach. Dissolved nutrient inputs to a reach are estimated as

$$DN_{in} = (DN_i + DN_{up})$$

where DN_{up} is the upstream input of nutrient. Routing to the downstream subcatchment is estimated as

$$D_{out} = D_{in} \exp(-zL_i W_i / Q_{MAF,i})$$

where z is a scaling parameter; and W_i is the width of the stream in subcatchment i. The attenuation of dissolved nutrients increases with high nutrient concentrations and for larger stream reaches.

2.4 Point sources

Point source inputs of SS and nutrients are also modelled in CatchMODS but are not described here in any detail.

2.5 Economics

The costs of management change scenarios are also estimated in CatchMODS. Three types of costs are estimated: fixed, ongoing and landuse-related. Fixed costs are those one-off costs which are incurred during the implementation of riparian and gully zone remediation works. Ongoing costs are the maintenance costs required to maintain the effectiveness of riparian and gully zone remediation works for pollutant control. The landuse-related costs represent the change in gross margins associated with the conversion between landuses. In most instances water quality improvements require conversion from higher to lower value land uses.

3. MORUYA AND TUROSS MODEL

The use of the CatchMODS model in the Moruya and Tuross River catchments is presented here as an example of the integration of collateral knowledge in the model development process. The Moruya and Tuross River catchments are located in the Eurobodalla region of the southeast coast of Australia. The catchments are 1420 km^2 and 1820 km^2 in size, respectively. Both catchments are dominated by native eucalypt forests but grazing and agricultural land uses occur predominantly in the lower parts of the catchments. The catchments are typical of many in the coastal zone of south eastern Australia which face pressure for urban development and land use intensification. Both catchments are sources of drinking water for local communities and support a range of agricultural industries. Their estuaries have high conservation and tourism value.

The application of CatchMODS in the Moruya and Tuross River catchments is an element of a broader program which includes in-stream water quality monitoring, sediment tracing and field measurement. The overarching objectives of the program are to assist local policy makers and managers to (i) identify priority sites (subcatchments) for remediation or management change; (ii) identify appropriate remediation measures; and (iii) estimate the cost effectiveness of remediation and/or management changes. Figure 1 shows the elements of the program and their linkages. The following sections briefly describe these elements and how the knowledge generated has informed the model development process.

3.1 Monitoring

An event-based water quality monitoring program was established in the early stages of the research to build understanding of water quality processes and to parameterise and test the model outputs (Figure 1). Three sites are monitored and sediment and nutrient concentration data from several medium sized events have been collected and analysed at two of those sites. Drewry et al. (2008) describes the monitoring program.

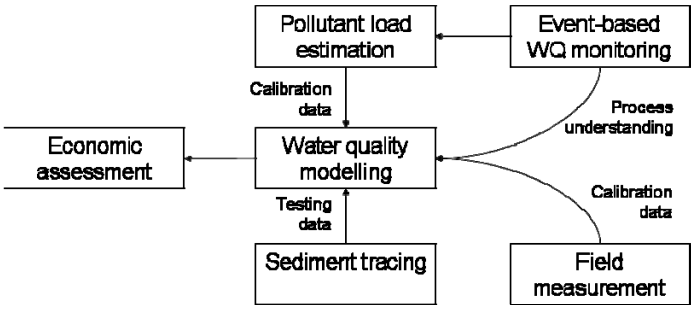


Figure 1. Elements and structure of the research program for the Moruya and Tuross River catchments.

Important process knowledge was generated through analysis of the data collected in the monitoring program. For example, it was found in the Tuross River catchment that during flow events the suspended sediment concentration preceded the flow peak by 9-11 hours indicating that: (i) sediment is supply-limited and that catchment-based sediment control measures are likely to prove effective for water quality improvement; and (ii) sediment sources are likely to be from channel sources or areas close to the monitoring site. The nitrate concentration peak was observed to lag behind the flow peak by several hours and it is likely to be related to groundwater inputs of nitrate. The total nitrogen concentration peak lay between the sediment and nitrate peak indicating a particulate and groundwater sources. Antecedent conditions and storm intensity were both observed to influence pollutant loadings with dry antecedent conditions during drought, and high intensity storms resulting in greater pollutant loadings.

There are several implications of these observations on the structure of the model developed for the Moruya and Tuross River catchments. The first is that effort in estimating channel erosion is warranted as channel sources are likely to comprise a high proportion of the total sediment yield. The dominance of nitrate inputs from groundwater necessitates the representation of hydrologic pathways and this is achieved via the use of the IHACRES model with a relatively simple baseflow filtering approach. The effects of antecedent conditions and storm intensity are not explicitly represented in CatchMODS and this is suggested as an area of further research.

A turbidity meter and data logger installed in the lower Tuross River catchment will aid catchment process understanding, improved load estimation and parameterisation of the sediment submodel. The turbidity data is to be used in conjunction with the water quality monitoring data to determine relationships between turbidity and suspended sediment concentration. Such data will also enable more accurate characterisation of events rather than relying on temporally sparser sampling of pollutant concentration and turbidity.

3.2 Sediment tracing

Sediment tracing research has been completed in the Moruya and Tuross River catchments to identify those subcatchments contributing disproportionally high suspended sediment loadings. The research was based on the geochemical analysis of deposited and in-stream sediment samples at several stream confluences. The objective was to identify the proportional contribution of sediment from selected tributaries and to provide independent data for testing CatchMODS. A description of the method can be found in Fu et al. (2008).

The Donalds and Burra Creek subcatchment were identified as producing disproportionately high sediment loadings to the Moruya River catchment. The Wadbilliga River subcatchment was similarly identified in the Tuross River catchment. Table 1 presents a comparison of the sediment tracing results of Fu et al. (2008) against preliminary results from CatchMODS. With the exception of the Moruya-Deua River – Burra Creek confluence, the sediment tracing results and model estimates are reasonably well matched. The poor comparison at the Moruya-Deua – Burra Creek confluence is thought to be a result of the difficulty in using sediment tracing techniques to compare between confluences of substantially different sizes. Other factors including possibly erroneous

estimates of streambank erosion and the omission of estimates of road erosion require investigation as to their impacts.

Table 1. Comparison of the sediment tracing results of Fu et al. (2008) against preliminary results from CatchMODS.

Confluence name (A-B)	A% by sediment tracing	Mean error (%)	CatchMODS estimate
Burra– Donalds Creeks	30.09	7.36	30
Moruya-Deua River – Burra Creek	45.18	11.33	95
Tuross – Wadbilliga Rivers	54.30	28.37	61
Tuross River – Belimbla Creek	83.95	3.25	70

Further research is required into identifying the relative contributions from the different sediment sources in the catchment. This is to be attempted in a Goodenough Gully subcatchment of the Moruya River catchment where the proportional contribution of road erosion to the total sediment load is to be estimated using geochemical tracing techniques. Opportunities also exist to distinguish between surface and subsurface sediment sources using radionuclide tracing techniques.

3.3 Field measurement and observation

All of the parameters of CatchMODS are theoretically measurable. In practice they are obtained via field measurement within a study catchment or extrapolated from other catchments. In the application of CatchMODS in the Moruya and Tuross River catchments, several catchment-specific data sets were obtained. For example, the location and severity of gullies were field-checked against mapping acquired through air photo interpretation. The accuracy of the mapping was assessed to be of high quality with few errors of omission or misclassification. The dimensions of the gullies (width and depth) were also measured to parameterise the sediment submodel. The gully width divided by the estimated age of the gully provides an upper limit on the lateral erosion rate which is otherwise extrapolated from catchments in the same region.

A range of soil physical and chemical data is required as input to the modelling. This included estimates of sediment bulk densities, suspended sediment proportions and nutrient concentrations for gully and streambank erosion sites. Analyses of soils under different land uses have also been measured. A preliminary description of this work and its incorporation in CatchMODS is provided in Fu et al. (2005). Nutrient concentrations of groundwater were also analysed from accessible bore sites across the catchment.

The data so far collected for the application of CatchMODS in the Moruya and Tuross River catchments is data readily obtainable. Longer term information is required, for example, to improve confidence in measurements of the rates of gully and streambank erosion.

4. SUMMARY

This paper has described progress on the modelling of sediments and nutrients to inform management effort. A full description of the structure of the catchment-scale model known as CatchMODS has been presented. This has included a description of each of its hydrologic, sediment, nutrient and economic submodels. A discussion of the integration of knowledge from a range of collateral sources to inform the development of the model was also made. Useful progress has been made towards improving the model though the integration of this data. Challenges remain to further incorporate collateral knowledge into the model development and application process. Such challenges present need to be actively perused to improve confidence in model outputs.

ACKNOWLEDGEMENTS

This research was undertaken with funding from the Australian Research Council and Eurobodalla Shire Councils as part of project LP0560439. Support from the Australian National University and the Cooperative Research Centre for Landscape Environments and Mineral Exploration were also provided and are gratefully acknowledged.

REFERENCES

Bhattarai, R. and D. Dutta, Estimation of Soil Erosion and Sediment Yield Using GIS at Catchment Scale, *Water Resources Management*, 21, 1635-1647, 2007.

Croke, B.F.W., W.R. Evans, S. Schreider, and C. Buller, Recharge Estimation for Jerrabomberra Creek Catchment, the Australian Capital Territory, In F. Ghassemi, M. McAleer, L. Oxley and M. Scoccimarro (eds), MODSIM 2001 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 555-560, 2001.

Drewry, J.J, L.T.H. Newham and B.F.W. Croke, Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross estuary, Australia, *Journal of Environmental Management*, 87, 2008.

Drewry, J.J, L.T.H. Newham, R.S.G. Green, A.J. Jakeman and B.F.W. Croke, A review of nitrogen and phosphorus export to waterways: context for catchment modelling, Marine and Freshwater Research, 57, 757-774, 2006.

Dubé, K.V., W.F. Megahan and M. McCalmon, Washington Road Surface Erosion Model (manual), State of Washington Department of Natural Resources, 2004.

Fu, B., L.T.H. Newham, and J.B. Field, Influence of particle size on geochemical tracing in Australia, In Sediment Dynamics in Changing Environment, Proceedings of the International Association of Hydrological Sciences, IAHS Publication 325, 2008.

Fu, B., L.T.H. Newham, J.B. Field, Integrating modelling and field-based analysis for improving catchment-scale pollutant management: an application of the CatchMODS model in the Moruya River catchment, In Zerger, A. and Argent, R.M. (eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 2693-2699, 2005.

Fu, B., L.T.H. Newham, J.B. Field and A.J. Jakeman, A Catchment-Scale Model of Road Erosion and Sediment Delivery, In Oxley, L. and Kulasiri, D. (eds) MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 2090-2096, 2007.

Fu, B., L.T.H. Newham, and C.E. Ramos-Scharrón, A review of surface erosion and sediment delivery models for unsealed roads, *Environmental Modelling and Software*, submitted.

Jakeman, A.J., I.G. Littlewood and P.G. Whitehead, Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, *Journal of Hydrology*, 117, pp 275-300, 1990.

Newham, L.T.H., R.A. Letcher, A.J. Jakeman, and T. Kobayashi, A framework for integrated hydrologic, sediment and nutrient export modelling for catchment-scale management, *Environmental Modelling and Software*, 19, 1029-1038, 2004.

Post, D.A. and A.J. Jakeman, Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model, *Ecological Modelling*, 123, 91–104, 1999

Prosser, I.P., P. Rustomji, W.J. Young, C. Moran, and A. Hughes, Constructing River Basin Sediment Budgets for the National Land and Water Resources Audit, CSIRO Land and Water Technical Report, Canberra, 2001.

Prosser, I.P. and P. Rustomji, Sediment transport capacity relations for overland flow, *Progress in Physical Geography*, 24(2), 179-193, 2000.

Renard, K.G., G.R. Foster, G.A. Weesies, D.K. McCool, and D.C. Yoder Predicting soil erosion by water – a guide to conservation planning with the revised universal soil loss equation (RUSLE). Agricultural Handbook No. 703, US Government Printing Office, Washington, D.C. ,1997.

Warren, S.D., H. Mitsova, M.G. Hohmann, S. Landsberger, F.Y. Iskander, T.S. Ruzycki, and G.M. Senseman, Validation of a 3-D enhancement of the Universal Soil Loss Equation for prediction of soil erosion and sediment deposition, *Catena*, 64, 281-296.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

CLUES Spatial DSS: From Farm-Scale Leaching Models to Regional Decision Support

Sandy Elliott[a], Graham McBride[a], Ude Shankar[a], Annette Semandi-Davies[a], John Quinn[a], David Wheeler[b], Liz Wedderburn[b], Bruce Small[b], Allan Hewitt[c], Robert Gibb[c], Roger Parfitt[c] , Brent Clothier[d], Steve Green[d] , Simon Harris[e] and Gerald Rys[f]

[a] *National Institute of Water and Atmospheric Research, Hamilton, New Zealand..*
s.elliott@niwa.co.nz
[b]*AgResearch, Hamilton, New Zealand*
[c]*Landcare Research, Palmerston North, New Zealand*
[d]*HortResearch, Lincoln, Canterbury, New Zealand*
[e]*Harris Consulting, Wellington, New Zealand*
[f]*Ministry of Agriculture and Fisheries, Wellington, New Zealand*

Abstract: Land-use in New Zealand poses risks to water quality in streams and lakes, yet there have previously been few tools used in New Zealand to predict the effects of land-use change at catchment to national scale. The CLUES spatial decision support system has been developed recently to assist with the assessment of land-use change on water quality, farm economics, and employment. The system incorporates a number of existing models from several research providers, ranging from models of leaching at the farm scale to national regression-based models. An early application of the model is the identification of catchments where the receiving surface waterbodies are at risk from nitrate sources. Several extensions of the system are underway to improve the usefulness for local decision-making. These include making the spatial resolution of land-use finer, adding mitigation and land management options, and increasing the number of environmental measures provided by the model, and linking these measures to a values framework. This will improve the utility of the model for integrated catchment management.

Keywords: land use; water quality; decision support

1. INTRODUCTION

Land-use in New Zealand poses risks to water quality in streams and lakes [PCE 2004; MfE 2007a, b]. Although nutrient levels in New Zealand rivers are still low by international standards, there is concern about current and potential eutrophication in lakes, and the impacts of intensification of pastoral agriculture. While field-scale nutrient budgeting models are available, and detailed dynamic models have been applied to specific catchments, there have previously been few planning-level tools available in New Zealand to predict the effects of land-use change at catchment, regional or national scales. The CLUES (Catchment Land Use for Environmental Sustainability) project was initiated in 2003 by the Ministry of Agriculture and Forestry (MAF) and the Ministry for the Environment (MfE). The objective was to tie together a number of models into one GIS platform to facilitate assessments of the progressive effects of land use change on water quality in lakes, rivers, and coastal regions. An early application of CLUES was envisaged in the Sustainable Water Programme of Action, which wanted to identify catchments that are associated with waterbodies at risk of nitrate contamination, at the national level. At the same time as providing a national perspective, it was intended that CLUES would have sufficient spatial resolution so that predictions within particular regions or catchments of interest could be provided. In this sense, the model is dual-purpose. In addition, farm-

economics components have been brought into CLUES to assess the immediate economic implications of land-use change.

The models currently implemented within the Catchment Land Use and Environmental Sustainability framework (CLUES) include predictions of nutrient and microbial flux in streams, and farm employment and economic factors as a function of land-use, climate, and soils. The first version of the system is now complete and has been distributed to national and regional land/water management agencies.

In this paper we describe the CLUES system, discuss ongoing developments, and discuss the applicability of the system to integrated catchment management planning.

2. DESCRIPTION OF THE SYSTEM

CLUES contains databases, a user interface, and graphical display, and models housed within ArcGIS (Figure 1). A national spatial database of all the spatial layers required for the model is provided, although the user may also import land-use layers. A graphical user interface is provided for the model, so that the user may easily identify the study area of interest, set up land-use scenarios interactively, and view results. Users with a knowledge of GIS may also make use of broader ArcGIS functionality to further manipulate the land-use and output displays.

The spatial framework of the model is built around the stream network and associated subcatchments of the REC national stream network. The network has approximately 500,000 stream reaches with subcatchments of 0.5km^2 area on average. The mixture of land-use within a subcatchment is specified, but not the locations of the land-uses within the subcatchment. This level of spatial resolution is suited to studies from the catchment scale to national scale, but not for paddock-scale management.

At present, the system provides predictions of the following measures:

- Total phosphorus (TP), total nitrogen (TN) and *E. coli* loads and yields for each stream reach, and loads generated from each subcatchment. A sediment component is in the process of being added.
- Gross domestic product arising directly from the farming, horticulture, and pine plantation land uses.
- Total employment for farms.
- Cash flow surplus – this is the output from the land use after farm working expenses have been deducted, but before interest, leases, wages of management, and capital expenditure.
- Leaching risk for nitrogen (on a relative scale from very low to very high) for the current land-use, displayed as a grid with 100 m resolution (a departure from the spatial framework described above).

CLUES draws on a number of models from a number of research providers to provide these predictions:

- SPASMO (Soil Plant Atmosphere System Model, HortResearch) [Rosen 2004]. This is a daily model of water and nutrient flux, and is used for prediction of nitrogen leaching from horticultural land-uses. Rather than run the daily SPASMO model within CLUES, a meta-model of mean annual leaching as a function of land-use, region, and rainfall was derived from a number of long-term SPASMO runs. This meta-model is implemented as a look-up table within CLUES.
- OVERSEER Nutrient Budget Model (AgResearch) – is a farm-scale nutrient loss model for various land uses (dairy, sheep/beef lowland, sheep/beef hill country, sheep/beef high country, and deer). It provides annual average estimates of nitrogen losses from these land uses, given information on rainfall, region, soil order, topography, and fertiliser applications. A simplified version of Overseer is

used in CLUES, driven by representative regional inputs (such as herd composition for a particular land-use) where necessary.

- TBL (Triple Bottom Line, Harris Consulting) [Woods et al. 2006] - estimates economic output from different land use types (pasture, horticulture, forestry and cropping), in terms of Cash Farm Surplus (CFS), Total GDP and Total Employment. The calculations are based on the MAF Farm Monitoring models.
- Ensus (Landcare Research) - provides maps of nitrate leaching potential. This is used as an adjunct to interpretation of CLUES results. It is based on studies of nitrogen losses at national and regional scales [Parfitt et al. 2006].
- SPARROW (USGS, NIWA) [Smith et al. 1997, Elliott et al. 2005] – is used for the *E. coli* model, and to assess diffuse sources of nutrients not provided by the OVERSEER or SPASMO models. SPARROW also provides the spatial framework for accumulation and routing of loads, and it includes provisions for loss processes in streams and lakes or reservoirs. The extra source terms and decay parameters are determined by calibration to national and regional water quality data.

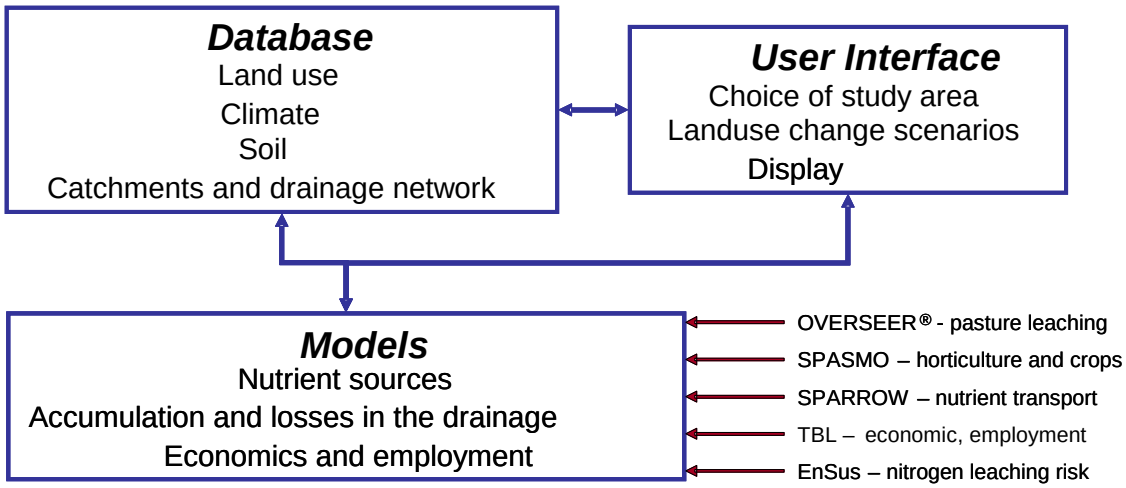


Figure 1. Schematic representation of the CLUES model.

The user interface is written within ESRI ArcGIS, but a custom control panel is used rather than the conventional ArcGIS toolbars, to make the system more appealing to users with little GIS background (Figure 2). The interface manages the reach selection, run control and scenario management, output display, land-use modification, and data or map export. Users familiar with GIS can also use the complete set of ArcGIS features to manipulate the land-use and mapping. The system allows the user to draw polygons of new land-uses, which enables a quick examination of the effect of land-use change. For more advanced uses, land-use can be modified externally and then imported (for example, to specify a fractional conversion of land-use or to make the changes dependent upon other factors such as slope).

Results can be displayed as maps in a variety of forms (Figure 3), or as tabular output and summaries. A tutorial manual is now available from NIWA [Semadeni-Davies et al. 2007]. Values of interest, such as the nutrient load or farm profit can be displayed either as local values associated with a subcatchment or as values accumulated in the upstream catchment.

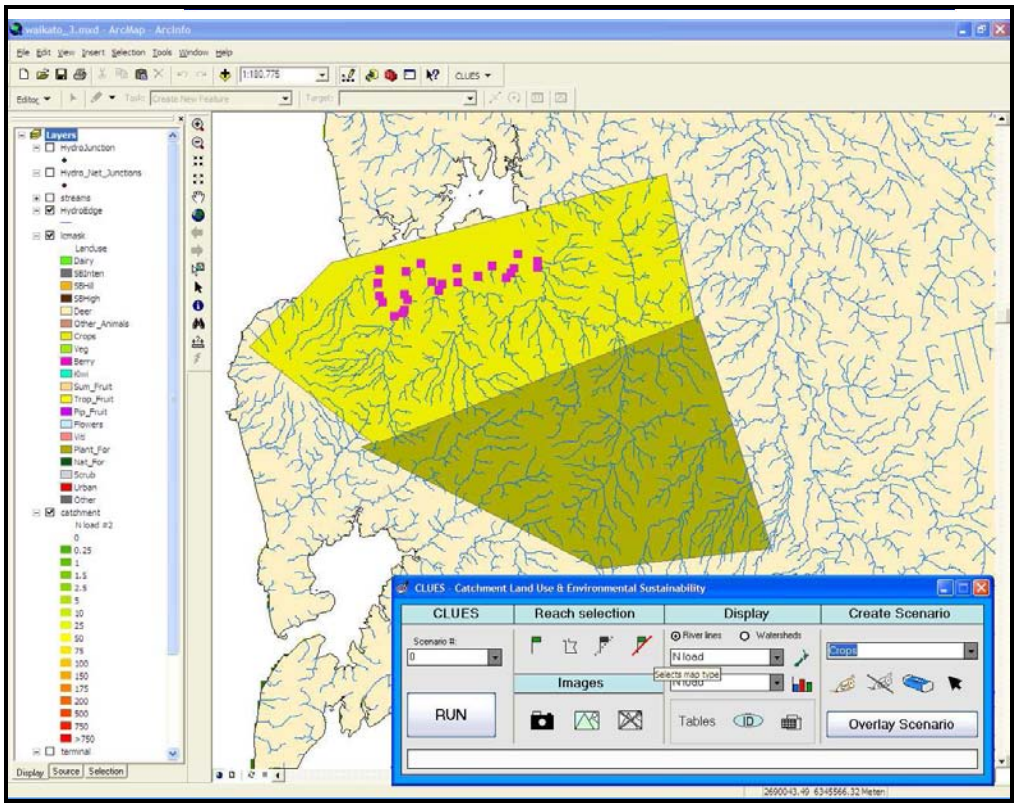


Figure 2. Tool-panel and interactive modification of land-uses

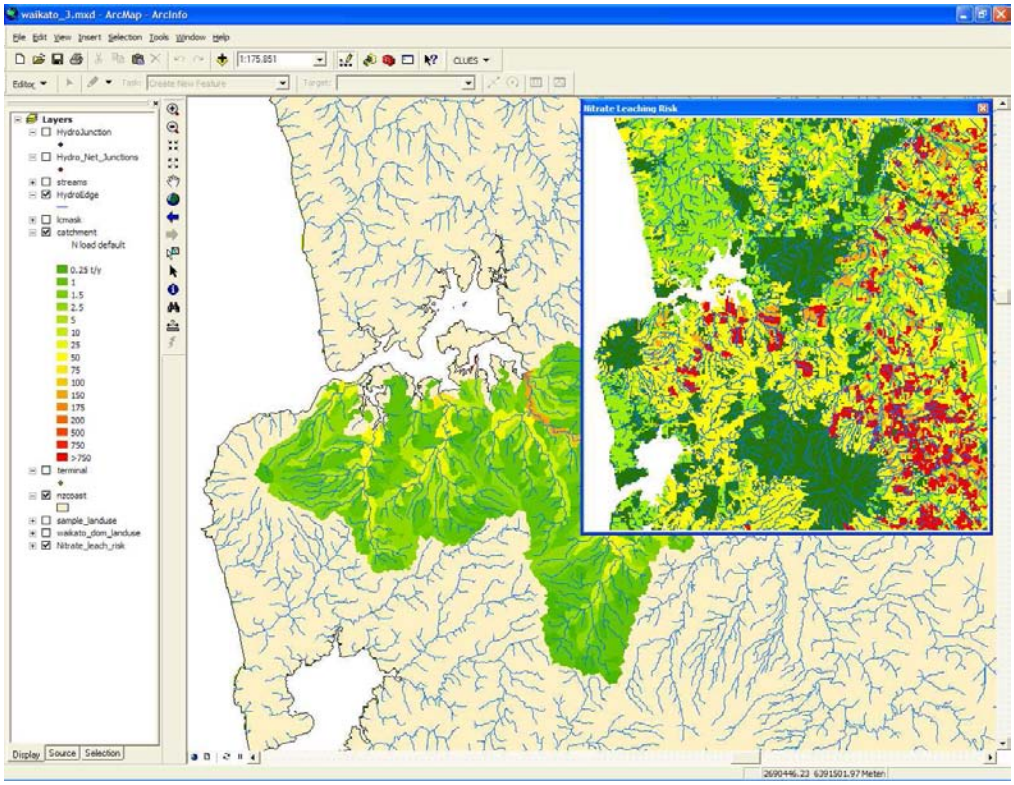


Figure 3. Example map of model predictions. The subcatchments are coloured according to the nutrient load in the associated stream reaches. An accompanying more detailed map of nitrogen leaching risk for the current land-use is also shown.

5. APPLICATIONS AND FUTURE DEVELOPMENTS

The CLUES concept has been embraced by regional and national level agencies, with funding for ongoing development and dissemination. As an example of the support, improved access to national databases on land-use has been provided by MAF. Now that the first version of CLUES is available, various management applications are emerging, and these are driving further development and refinement of the system to make it more suited to these applications.

5.1 Identification of sensitive and at risk catchments

Action 2.2 of the Sustainable Water Programme of Action calls for identification of catchments that are "sensitive and at risk from rural and urban diffuse discharges" [MfE, 2006], and this is a current early application of CLUES. Also, the pastoral agriculture industry are interested in identifying catchments that are sensitive to the effects of dairy land-use, and are co-funding this application of CLUES. The sector seeks to identify locations where dairying is an issue in relation to water quality, so that the scope of the problem can be quantified and mitigation measures targeted and prioritised. This has some parallels to identification of Nitrate Vulnerable Zones in the UK under the EU Nitrates Directive.

At the most basic level, this involves identifying subcatchments with relatively high nutrient yields, and the proportion of the load attributable to dairying – that is, traditional 'hot-spot' identification. This is straightforward to achieve with the current CLUES system. A modification of this approach is to consider the delivered yield or load, that is, the yield or load after delivery to major waterbodies after nutrient attenuation. This 'delivered yield' can be calculated in the CLUES system as the attenuation in streams is modelled, and this will be provided in future versions of CLUES.

An alternative stance on this problem is to identify waterbodies that exceed some threshold concentration. We are using CLUES to determine flow-weighted concentrations in lakes and streams, and comparing this with guidelines for eutrophication or concentration guidelines in streams. By this process, the most-impacted streams and lakes, and their associated catchments, are being identified. Within these catchments, the predominant hot-spots can also be identified.

'Catchment sensitivity' also suggests some consideration of the sensitivity of the receiving environment to increases in nutrient loads. This has several facets. Are the receiving water bodies currently limited by nitrogen or phosphorus? What is the value of the receiving environment, taking into account the use of the water body, uniqueness or representativeness, or contribution to biodiversity? Is the water body already impaired or reaching some threshold in terms of nutrient loading? Work by other agencies to identify waters of national importance (WONI) could feed into this process.

Finally, catchment sensitivity also can be interpreted as the change in state of the receiving water body in response to some anticipated future change in land-use – either an incremental unit change or some future pre-determined land-use scenario. While land-use can certainly be altered in the model, methods to express the results in terms of this forcing-response relationship are not yet developed.

5.2 Modifications for regional applications

The Regional Councils who are interested in application of CLUES have identified that the concentrations of contaminants in streams are in some cases more important than the loads. This is partly due to water quality criteria being expressed in terms of concentrations (which in turn are related to nutrient responses of stream periphyton, bathing suitability criteria for E. coli). This has spurred further research relating loads to concentrations (based on examination of national water quality databases), which will be incorporated into CLUES. Regional Councils are also interested in versions of CLUES with parameters optimised for their particular region; at present, some parameters are available only from a national-level

calibration. Hence, some projects to develop more locally-applicable versions of the model are underway.

5.3 Linking to mitigation and land management

At present, the system does not take mitigation measures into account – it is restricted to consideration of changes in land-use. This limits the ability of the model to identify the benefits of modifications to land management. Hence, mitigation measures are being added to CLUES. For similar reasons, the ability to add more information on management, such as intensification of grazing or modification of fertilise application rates is also being added.

In an effort to link regional-scale planning approaches to farm-scale land management action, we are investigating the means to drive CLUES with the farm or property as a level of spatial unit, hence providing linking between scales. This also provides opportunities to improve the predictions by driving the nutrient modules with more accurate farm-specific data (at present, a range of representative farm types and associated parameters are used, which has been identified as limiting the predictive capabilities for specific small catchment applications). This poses challenges in terms of obtaining and managing the spatial data.

5.4 Linking to values to enhance usefulness for integrated catchment management

At present, CLUES is targeted mostly to assessment of land-use change on water quality, with an additional consideration of economic and employment considerations. This stems from the drivers for the initial development of the system. For use in integrated catchment management, however, it is necessary to consider a wider range of environmental measures.

A systematic assessment of values and suitable environmental metrics has recently been developed by AgResearch. In this approach, sustainability value domains were first established (Table 1) [Small 2008]. Each of the value domains has been broken into a number of ‘value objects’, such as water quality. Indicators and measures for each of these value objects have been identified for international, national, regional, catchment, and farm scales. This has resulted in an extensive list of measures relevant to integrated catchment management. Some of these measures are included in CLUES, while others are would be difficult to include and are likely to remain outside CLUES. This list is being prioritised, with the aim of establishing relations between land-use and values to add to CLUES, or else to relate the current outputs of CLUES to some sort of aggregated values score.

A few water values can be assessed fairly directly from water quality variables predicted by CLUES (e.g., E coli concentrations as indicators of suitability for contact recreation). However, the links between most waterways values (e.g., clear water, native fish biodiversity, safe for swimming, healthy recreational fishery, healthy ecosystem) and levels of N and P predicted by CLUES are typically less straight-forward. This is because the effects of nutrient concentrations or loads on values are conditional upon the local context set by habitat factors, such as presence of riparian shade (high shade prevent nuisance algal growth regardless of nutrient levels), flow regime (highly variable flows can also control plant growth), and lake flushing rates (that influence time for algal bloom development), stratification patterns and history of enrichment. Other biotic factors, such as presence of exotic plant and fish species, also influence how nutrient levels/loads affect water values. This context-dependency of nutrient-water values suggests that the CLUES output will need to feed into other locally tuned models to link directly to common values defined in community consultation processes, for the purpose of integrated management of small catchments. For example, locally-focused holistic models of land use links to waterway values are being developed as Bayesian Network models (initially for Bog Burn catchment in Southland). CLUES output could inform /provide input to such models in various ways.

5.5 Other developments

There are also several future developments which would improve the usefulness of the model for catchment management, including: allowing for temporal progression of land-use rather than a constant future scenario, which introduces possibilities for linking with land-

use evolution models and also prediction of transient responses to land-use change; incorporation of a groundwater component, which is limiting the applicability of the system in some parts of the country; including the effects of land-use change on water resources (such as hydrological statistics); adding a greenhouse gas emissions component (building on the capabilities of OVERSEER); adding uncertainty metrics to the predictions; and providing predictions of biotic impacts in addition to biophysical and economic variables.

Table 1. Sustainability value domains and value objects, from Small [2008].

Sustainability value domains	*Environment*	*Economy*	*Quality of life*	*Culture and identity*	*Participation and equity*
Value objects	Air	Productivity	Safety and security	Identity	Equity
	Land and soil	Prosperity	Health	Culture and heritage	Civic participation
	Water	Employment	Quality of work life		Treaty of Waitangi
	Landscape	Infrastructure	Knowledge and skills		Political/social trust
	Biodiversity	Tourism	Recreation and leisure		Human rights
	Biosecurity		Social connectedness		International treaties
	Kaitiakitanga/ stewardship		Housing		

5.6 Comments on the process of model development

Some of the modifications and improvements discussed above were identified in early round-table meetings of the programme guidance committee, but were not implemented at that time so that efforts could be focused on making the system operational. Other modification needs have become clear following demonstrations of early prototypes. The provision of a second phase of funding, support from regional and national authorities, and a collaborative approach by the researchers has enabled these modifications, which will ultimately make the system more useful for integrated land/water management. There are several elements that have contributed to the success of this project. First, there is a key central-government agency driving the process, with accompanying funding. This not only lends a weighting of government approval to the model, but also encourages the collective engagement of a number of research providers, who might otherwise find it more difficult to collaborate. The involvement of Regional Councils at user-group meetings, as prototype testers, and as funders has supported the programme. Such councils are at a stage where they seem to be receptive to the application of models to catchment planning, whereas this may not have been the case a decade ago. This funding is continuing, which allows the model to continue to be maintained and improved. The CLUES system also serves as a means for various organisations involved in farm-scale modelling to up-scale their models, thus providing a wider range of use of their models. Despite these positive signs, the use of CLUES for actual applications is in its infancy, and the long-term utility and hence viability of the system remains to be confirmed.

6. CONCLUSIONS

The initial national-level impetus to assess the effects of land-use change on water quality has led to the successful development of a spatial decision support system that brings together a range of models with a common national-scale driving datasets and spatial framework. Application of the system to regional and catchment management has highlighted the need for extensions to the system, including refining the spatial resolution, including mitigation measures, and linking to a values framework. Ongoing support of national and regional agencies is crucial in making the system more relevant to integrated decision support.

REFERENCES

Elliott, A.H., Alexander, R.B., Schwarz, G.E., Shankar, Ude, Sukias, J.P.S., and G.B. McBride. Estimation of nutrient sources and transport for New Zealand using the hybrid mechanistic-statistical model SPARROW. *Journal of Hydrology (NZ)* 44(1): 1–27, 2005.

Ministry for the Environment Freshwater for the Future: A supporting document. A technical information paper outlining key outcomes for the sustainable management of New Zealand's freshwater, 2006.

Ministry for the Environment. Environment New Zealand 2007. Ministry for the Environment. Wellington, New Zealand, 2007a.

Ministry for the Environment. Lake water quality in New Zealand Status in 2006 and recent trends 1990–2006. Publication number ME82. Ministry for the Environment. Wellington, New Zealand, 2007b.

Parliamentary Commissioner for the Environment. Growing for good? The sustainability of intensive farming in New Zealand. New Zealand Parliamentary Commissioner for the Environment, Wellington, New Zealand, 2004

Rosen, M.R., Reeves, R.R., Green, S.R., Clothier, B.E., and N. Ironside. Prediction of groundwater nitrate concentration contamination after closure of an unlined sheep feed lot in New Zealand. Vadose Zone Journal 3: 990–1006, 2004.

Small, B. Sustainable development and technology: genetic engineering, social sustainability and empirical ethics. *International Journal of Sustainable Development*. In Press, 2008.

Smith, R.A., Schwarz, G.E., and R.B. Alexander. Regional interpretation of water-quality monitoring data. Water Resources Research 33: 2781–2798, 1997.

Woods, R., Elliott, S., Shankar, U., Bidwell, V., Harris, S., Wheeler, D., Clothier, B., Green, S., Hewitt, A., Gibb, R., and R. Parfitt. The CLUES Project: Predicting the Effects of Land-use on Water Quality – Stage II. Report prepared for the Ministry of Agriculture and Forestry. NIWA Client Report HAM2006-096 NIWA, 2006.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Linking Catchment Environmental Planning to On-Ground Investment: the SCaRPA DSS

N.F. Herron[a], T.W. Barrett[b] and S.M. Cuddy[a,c]

[a] Integrated Catchment Assessment and Management Centre, The Fenner School of Environment and Society, The Australian National University, Canberra ACT 0200, Australia (natasha.herron@anu.edu.au)

[b] NSW Department of Environment and Climate Change, Armidale, Australia (tom.barrett@environment.nsw.gov.au)

[c] CSIRO Land and Water, Black Mountain Laboratories, Clunies-Ross St, Canberra ACT 2600 Australia (susan.cuddy@csiro.au)

Abstract: The Site and Catchment Resource Planning and Assessment (SCaRPA) decision support system (DSS) has been developed for use by Catchment Management Authorities (CMA) in New South Wales, Australia to assist with catchment planning and environmental investment decision-making. The catchment planning module can be used to set targets for environmental outcomes and broad priorities for investment, using multi-criteria assessment and prioritisation tools, scenario building functions and environmental assessment models. Outputs from catchment planning can then be used to set the criteria for environmental incentives funding programs in the site-scale module, and the targeting of investment to key areas. Site-scale models are run to evaluate landholder proposals to undertake environmental works, and the results ranked by benefit-cost ratio to maximise environmental return on investment. When coupled with a well-planned monitoring and evaluation program, the SCaRPA DSS can contribute significantly to an adaptive management framework, in which site-scale investment decisions are informed by a catchment plan, which in turn is informed by progress towards management outcomes and resource condition targets through time via investment in environmental works. This paper describes the SCaRPA DSS and its intended use.

***Keywords**:* catchment planning; environmental investment, integrated assessment; decision support tool (DSS);

1. INTRODUCTION

As population pressures, climate change impacts and competing demands on our natural resources base have increased, there has been a significant shift in the way nations manage their natural resources – i.e. from the relatively unconstrained exploitation of 'unlimited' resources to the establishment of management frameworks, underpinned by guiding principles, for the integrated management and sustainable development of finite resources for future generations. In recognition of the complexity of natural resource management (NRM), due to factors such as institutional arrangements, trans-boundary issues and the inter-connectedness of natural systems, inter- and intra- national initiatives continue to be established around the world to undertake NRM based on principles of ecologically sustainable development (ESD), integrated catchment management (ICM) and adaptive management. In Australia, Commonwealth and State governments have agreed to a national framework for regional delivery of NRM [NNRMTF, 1999] which has seen the establishment of 56 regionally-based catchment management bodies across the country, responsible for the preparation and implementation of catchment management plans and investment strategies that adhere to the aforementioned principles of ESD, ICM and

adaptive management The catchment plans provide the over-arching frameworks for regional investment, and include statements of broad environmental objectives (longer term aspirational goals), accompanied by shorter-term resource condition and management targets, which contribute to realising the longer term objectives of the plans.

To support the implementation of their catchment plans, the catchment management bodies must prepare investment strategies, which provide detail on the activities, timeframes and costs for achieving targets. The effective implementation of a plan, via actual investment in on-ground works, requires the management body to address various issues, including:

- where to invest?
- how to calculate the environmental benefit?
- how much to invest?
- how to engage the community?
- how to ensure an equitable process?
- how to measure the benefit?
- how to link site-based measures back to catchment targets?

Ultimately, each catchment management body aims to achieve its environmental targets in the most efficient way, where efficiency takes into account not just the cost of change, but social considerations, such as the engagement, support and education of the community. Their challenge is to find the right balance between the potentially competing interests of livelihood, cultural values and environmental values.

A decision support system (DSS) can assist in managing the implementation process, and many of the catchment management bodies have developed DSSs to support their investment processes. The result is a plethora of different systems for informing NRM investment, which vary in their sophistication, the issues considered, their data requirements and their prioritization and assessment methods, thus complicating state-wide and nation-wide analysis and reporting. A federally-funded project was commenced in 2005 to develop decision support tools to assist Catchment Management Authorities (CMAs) in New South Wales with their catchment planning and investment processes. The system developed, the Site and Catchment Resource Planning and Assessment (SCaRPA) DSS, links catchment planning and implementation processes into a single GIS-based framework, and includes functionality for scenario modeling, managing data and models, visualizing results, generating reporting and multi-criteria assessments.

This paper looks at how the SCaRPA DSS can assist NSW CMAs to put in place an adaptive management framework which facilitates the cycle from catchment planning to effective implementation and round again.

2. NATURAL RESOURCE MANAGEMENT IN NSW

In 2003, 13 Catchment Management Authorities (CMAs) were established by an Act of the State Parliament to manage natural resources in NSW. The CMAs have been delegated responsibility for investing $436M (AUD) in on-ground works that improve environmental outcomes. To inform the investment process, each CMA has prepared a Catchment Action Plan (CAP), which contains the environmental and community objectives for its region for the life of the plan, including resource condition and management targets for water, land and biodiversity assets, and community outcomes. These CAPs embrace the principles of sustainable development and integrated catchment management, and are subject to periodic review and updating, consistent with an adaptive management framework.

CAPs are accompanied by 3-year investment strategies and annual implementation plans, which outline activities, timeframes and costs for achieving targets at different stages of a CAP's life. To support implementation, most CMAs have developed methods for making investment decisions, which vary in sophistication from paper-based questionnaires to GIS-based assessments to simulation modelling to custom-built decision support systems (DSS). While there is considerable overlap in the issues and methodologies between CMAs, the

lack of a consistent framework for undertaking catchment- and site-scale assessments makes comparisons between regions and statewide reporting significantly more challenging.

3. THE SCaRPA DSS

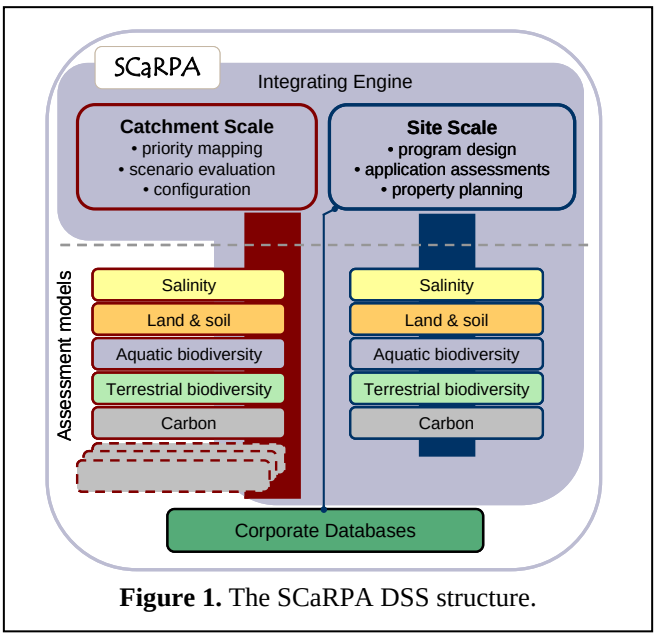


Figure 1. The SCaRPA DSS structure.

The SCaRPA DSS was designed to support current CMA business processes, specifically the process of developing catchment action plans and identifying priority areas for NRM investment, and at the implementation scale, evaluating competing proposals for incentives funding in terms of their cost-benefit ratio. Figure 1 illustrates the broad DSS structure, including the 2 modules – catchment-scale planning and site-scale incentives assessment – and the various environmental impact assessment models housed within each module. The system also includes a configuration tool, which allows each CMA to tailor the system to individual needs, including populating the database with region-specific resource condition and management targets, biodiversity benchmark data and other reference data.

The catchment planning module, SCaRPA-cp, has been designed so that, provided certain protocols are met, a CMA can 'plug in' catchment-scale assessment models of their choice. It contains tools for registering catchment models and data layers, generating priority maps, visualising multiple priority layers, building scenarios and evaluating the environmental impacts of land cover/management changes using the registered catchment models.

The incentives assessment module, SCaRPA-ia, permits CMAs to design funding programs, evaluate the environmental benefit of landholder proposals against criteria specified in a funding program and rank or rate proposals by their cost-benefit ratio, leading to a determination to fund or not fund.

The SCaRPA system supports the full planning process from identifying priority areas for NRM investment and the exploration of scenarios to assist in setting targets, to the evaluation of investment proposals for environmental outcomes via a competitive, credible, transparent assessment process. Links between catchment planning outputs and site-scale assessments are not automated in version 1, but will be a feature of future versions.

Under current deployment arrangements, SCaRPA will be delivered to CMAs as a stand-alone desktop application, with plans to connect it to the corporate network at a later date.

4. THE ROLE OF SCaRPA IN ENVIRONMENTAL INVESTMENT DECISION MAKING

While there is no prescribed process for using SCaRPA, we have assumed that any on-ground implementation program will be informed by some level of catchment planning. Together SCaRPA-cp and SCaRPA-ia assist CMAs to:

1. formulate and refine resource condition targets in terms of evaluation measures provided by the catchment scale tools, and set management targets accordingly;
2. identify spatially explicit priorities for each management target to guide investment at whole of catchment and subcatchment scales;
3. set minimum environmental thresholds for funding and determine the dollar value of environmental services;
4. predict the impacts of alternative planning/investment scenarios on resource condition targets, such as regional biodiversity status and other natural resource targets (e.g. soil, water);
5. respond to the results of monitoring and evaluation (M&E) by updating reference data and models, and reformulating catchment priorities in light of achievements;
6. communicate investment frameworks and assessment criteria to landholders; and
7. follow due process, by providing a transparent, equitable, repeatable, defensible framework for investing public funds.

4.1 Guide Target Setting

Two types of targets are contained within CAPs: *resource condition* targets express the intention to achieve an improvement in the condition of a natural resource by a specified time; and *management* targets set implementation objectives for meeting resource condition targets. Government has given little guidance on how resource condition targets should be set, though it has stated that they should provide a broad indication of catchment health. Management targets are to be more specific and measurable. As a result, resource condition targets tend to be poorly defined, qualitative statements and management targets are more quantitative. Table 1 lists a few examples of catchment and management targets from current CMA catchment plans.

Table 1. Examples of resource condition and associated management targets

Resource condition target	Associated management target/s
By 2016, improve the condition of native terrestrial and aquatic ecosystems	By 2016, 2200 ha of corridor habitat restored to provide connectivity (630 ha by 2009) [NR CMA, 2005]
By 2015 increase by at least 50,000 ha the area of the catchment that is managed to produce a net improvement in soil condition	• By 2015 an additional 15,000 hectares of land will be sustainably managed using industry agreed best management practices in accordance with the Land and Soil Capability system • By 2015 1,500 farmers will have developed property plans and at least 500 of these will have implemented improved farm management measures [BR-G CMA, 2007]

SCaRPA-cp includes priority mapping and scenario evaluation tools, which can be used to set more quantitative resource condition targets. Through an exploration of alternative management scenarios, informed by single and multi-criteria priority layers, the cp module can be used to align resource condition and management targets more closely. With quantitative targets, and a set of models to predict resource condition, progress towards targets from on-ground investment can be evaluated. In the SCaRPA, this necessary involves a comparison of modelled outcomes against a modelled target value, which is useful, although certainly not ideal. Independent monitoring programs are needed that measure the same condition indicators/metrics as those that inform the catchment-scale models within the DSS. Figure 2 illustrates an adaptive management framework in which the SCaRPA is used to assist catchment planning and site-scale investment decision-making steps, leading to the implementation of management actions, and subsequent reporting of progress towards targets via monitoring and evaluation. Ideally, results from the M&E process feed back into the catchment planning and investment decision-making process via iterative updating of model assumptions and reference data in the SCaRPA system.

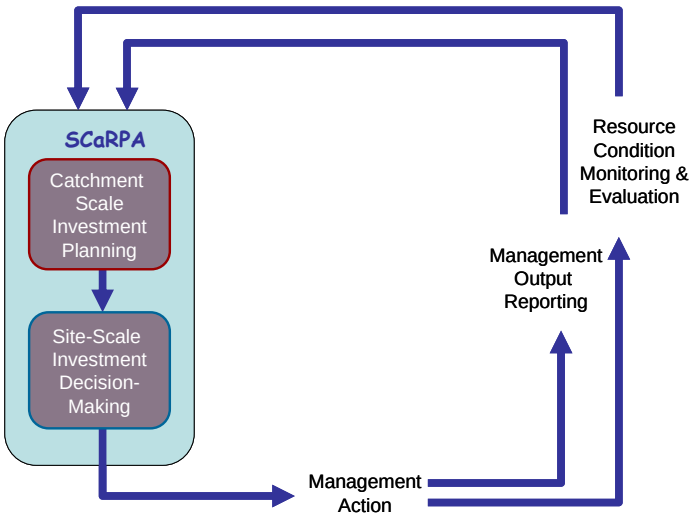


Figure 2. The relationship between the SCaRPA DSS and the cycle of catchment planning, property investment and monitoring-evaluation [adapted from Ferrier, *pers comm.*].

4.2 Guide Investment Location

Decisions about where to invest occur at both catchment and site scales. SCaRPA-cp helps set the broad-scale investment strategy, and is informed by regional, state-wide and/or national data layers, which do not provide sufficient spatial differentiation to pinpoint areas for investment. The site-scale models in SCaRPA-ia are informed by data collected at the sites where management changes are proposed. Thus SCaRPA-ia does not identify the 'best' areas in the CMA management area, but rather the 'best' proposals from a sample population of potential proposals from the CMA management area. Good planning at the catchment-scale can ensure that the sample population of investment proposals is from the 'best' end of the environmental outcomes continuum.

For each management target, spatially explicit priorities indicate where investment in different types of management intervention will maximise resource condition across a range of environmental values. The MCAS-S [BRS, 2007] software, which has been incorporated into SCaRPA-cp, is a spatial multi-criterion assessment tool for combining multiple priority layers into a composite priority surface. With a user-friendly drag-and-drop approach, and the ability to specify different weights for different layers, CMAs can quickly and easily identify areas for investment. If economic data is available spatially, the resultant priority layer can express priority in terms of an environmental cost-benefit ratio.

This information can then be used to establish an investment framework, including directing CMA efforts to engaging landholders in priority areas and/or setting different environmental cost-benefit thresholds for funding in different parts of the CMA area. Where there are competing demands for incentives funding, the priority maps provide a framework for negotiation and communicating the investment strategy. Investment decisions might still be driven by CMA values/policies and socio-economic pressures, but guided by target-driven priorities.

At the site-scale, a decision to invest in on-ground management relies on more detailed site information, which can include an assessment of current condition, local and upstream threats, and details of the current and proposed land uses and management. The site-level models are not automatically linked to priority area information from the catchment planning module, but this information can be entered manually into the data entry window for relevant environmental assessment models or used to frame the investment program within which proposals are evaluated. A funding program design tool is included in the site

module, which CMAs use to specify all the details and assessment criteria pertaining to a new funding program. Priority areas identified through the catchment prioritisation process can be explicitly targeted for investment by calling for applications from those areas.

4.3 Guide Investment Amount

Putting a dollar value on environmental services is a challenging issue, and this project has not attempted to provide a definitive answer to this question. Instead, we have assumed that the value of an environmental service will be determined by the funds available for investment and the market. The funding model options in SCaRPA-ia include tender-based and threshold-based approaches. In the absence of threshold information, a CMA might choose to run a closed tender program and fund the most competitive proposals. The cost-benefit score of the last proposal to be funded can then be used to set a minimum threshold for subsequent funding programs. Over time, as investment information is collected and collated, each CMA will develop a feel for what level of environmental benefit can be achieved per dollar spent. The threshold value for culling proposals from a funding program can be expected to vary over time and with varying assessment criteria – e.g. the threshold value for funding proposals based on providing salinity and biodiversity benefits could differ from that set for proposals that are being assessed for water quality and aquatic habitat benefits, even if the model scores have been standardised to a common range.

If a minimum benefit threshold has been defined, then the threshold-based funding model can be adopted instead of the ranked approach, or the threshold can be used in the ranked approach to cull unacceptable proposals prior to ranking. Figure 3 shows the funding model screen from which the CMA can select their preferred funding model for a program.

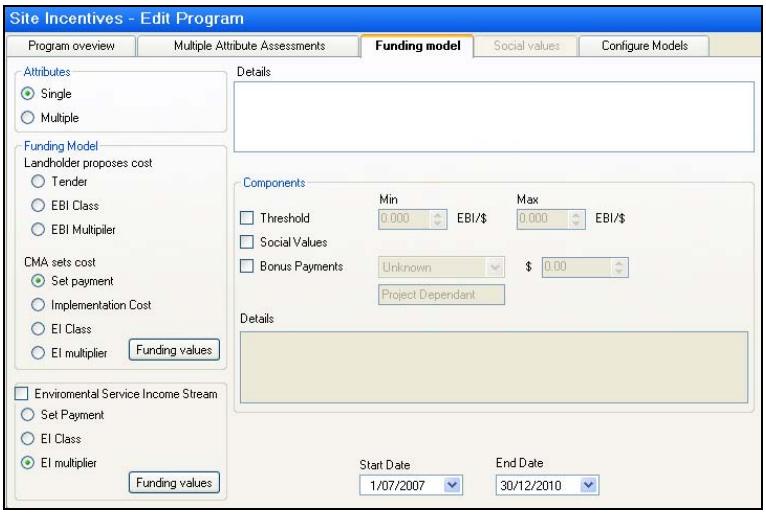


Figure 3. Example of funding models available within SCaRPA-ia

4.4 Determine Environmental Benefit

The environmental outcomes from changing land cover or management can be evaluated at both scales in the SCaRPA, but using different models. At the catchment-scale, models tend to be spatial, utilising existing grid and vector data, to evaluate environmental outcomes. SCaRPA-ia models, which evaluate salinity, terrestrial biodiversity, aquatic biodiversity, land and soil capability and carbon sequestration impacts, are not spatial, although the information captured during site assessments and the model predictions are attributed to spatial units and managed in a GIS. Outputs from these models can be re-expressed as a cost-benefit score, by standardising the environmental benefit score by the proposal cost, and then ranked by cost effectiveness. This ranking process also informs the where to invest question at the scale of implementation.

The metrics/indicators used by the different models, at both scales, do not translate easily into quantitative resource condition targets. However, while resource condition targets are largely qualitative, this is not a big problem. The use of indicators, rather than actual measures of the desired environmental outcome, is commonplace and is a consequence of the many difficulties in quantifying environmental benefits. This paper is not the place to discuss the issue, except to acknowledge that what some of the SCaRPA models predict as the environmental impact might be difficult to express in terms of contributions to resource condition targets. At the site-scale, we have adopted the position that the aim of the incentives assessment module is to assist with ranking competing proposals and that indicators of environmental benefit will suffice for this purpose. To evaluate progress towards targets, the proposals that are implemented can be fed back into the catchment models as a single scenario and evaluated in terms of their catchment-wide environmental impacts.

4.5 Monitoring and Evaluation

The SCaRPA system is not specifically a monitoring and evaluation tool, but progress towards achieving management targets and resource condition targets can be evaluated. The former assessment is relatively easy, since the management targets are expressed in terms of areas or lengths of change (e.g. area re-vegetated; length of stream rehabilitated) and this information is captured spatially via the SCaRPA-ia. The evaluation of progress towards resource conditions targets is more complicated. While the targets continue to be expressed qualitatively, progress towards targets cannot be reported in any meaningful way. Leaving aside the very real issue that any observed improvements might not be caused by the management activity, data collected from an M&E program could only be used to support an assertion that condition had or had not improved. The question of how much improvement or deterioration has occurred can only be determined using models. Thus one role of SCaRPA-cp is, firstly, to set resource condition targets, expressed in terms of the metrics of each of the assessment models and, secondly, to evaluate the environmental impacts of any implemented activities using the SCaRPA-cp models.

As management interventions are implemented, investment priorities will change. By updating the SCaRPA-cp model layers with the management outcomes and resource condition changes identified during monitoring, subsequent rounds of prioritisation and target setting can reflect the changing status through time, consistent with a truly adaptive management framework (Figure 2).

4.6 Engage the Community

The SCaRPA system provides a vehicle for engaging landholders and land managers in property-planning. A CMA officer can sit with a landholder and explore different management scenarios and see how they contribute, or not, to better environmental outcomes. With the aid of air photos or other imagery, maps and site information, a property plan can be developed or an incentives proposal prepared that is consistent with the catchments environmental objectives. The use of a DSS in this way can also contribute to meeting community targets within the CAPs.

4.7 Ensure an Equitable Process

Community engagement will also be facilitated through having a transparent, equitable, repeatable, defensible, scientifically-based framework for managing the assessment process. A CMA must be able to defend their decision to fund or not fund a proposal for incentives dollars, as they are accountable to the Commonwealth and NSW government for the investment of public funds.

SCaRPA-ia includes a program design tool (or template builder), which is where the objectives, details and assessment criteria of a funding program are laid down, prior to

advertising the program and processing applications. The assessment criteria nominated in the program design stage are coded into the system, such that when an application is assessed under that program, an automatically generated set of assessment questions (which might include a selection of social, cultural and risk criteria) and data requirements appear in the application entry windows, and the application is assessed against the rules (e.g. the type of funding model; the minimum benefit threshold, etc.) specified in the program template. If, for example, only salinity and biodiversity models were selected as part of an assessment process, only questions and data inputs needed to inform these assessments would appear in the application pro forma.

If the CMA wants to vary an assessment, we provide a text entry window in the DSS in which the justification for varying the decision from that generated via the assessment process can be recorded.

5. CONCLUDING REMARKS

The integration of catchment planning and on-ground implementation decision support tools into a single framework is a key feature of the SCaRPA DSS design. While version 1 does not fully realise the goal of automated links between priority maps generated in the catchment planning module and the site-scale assessment tools, it goes a considerable way to integrating catchment NRM planning and implementation decisions. Close consultation with the CMA clients has resulted in a customisable framework, which supports their business processes. Used in conjunction with a well planned monitoring and evaluation program, an adaptive management framework for improving natural resource condition can be realised, in which site-scale investment decisions are informed by a catchment plan, which in turn is informed by progress towards management outcomes and resource condition targets through the implementation of on-ground investment in environmental works.

ACKNOWLEDGEMENTS

SCaRPA is a product of the TOOLS2 project, a collaborative R&D project with the NSW Departments of Environment and Climate Change, and Primary Industries, and CSIRO (Land and Water). The authors would like to acknowledge the contributions of their colleagues towards the development of the assessment models and the implementation of the framework. The project operates under an Activity Agreement for use in the devolution of NAP/NHT funds under a Commonwealth and State Bilateral Agreement.

REFERENCES

BR-G CMA. *Border Rivers–Gwydir Catchment Action Plan*, Border Rivers-Gwydir Catchment Management Authority, Inverell, Australia, 2007. Available at http://brg.cma.nsw.gov.au/index.php?page=catchment-action-plan

Bureau of Rural Sciences. *MCAS-S, Multi-Criteria Analysis Shell for Spatial Decision Support*, Commonwealth of Australia, Canberra, Australia, 2007 http://adl.brs.gov.au/mcass/docs/userguide.pdf

Ferrier, S. Conceptual framework for integration of tools across scales - presented at TOOLS2 Workshop, Salamander Bay, NSW 17 Nov, 2005.

NNRMTF (National Natural Resource Management Task Force). *Managing Natural Resources in Rural Australia for a Sustainable Future: A Discussion Paper for Developing a National Policy*. Dept of Agriculture, Fisheries and Forestry, Canberra, 1999.

NRCMA. *Northern Rivers Catchment Action Plan*, Northern Rivers Catchment Management Authority, Grafton, Australia, 2005.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A consistent framework for knowledge integration to support Integrated Catchment Management

A. Holzkämper*, B. Surridge, A. Paetzold, V. Kumar, D. N. Lerner, L. Maltby, J. Wainwright, C. W. Anderson and R. Harris

Catchment Science Centre, Kroto Research Institute, University of Sheffield, Broad Lane, Sheffield S3 7HQ, UK
email: a.holzkamper@sheffield.ac.uk

Abstract: The European Water Framework Directive (WFD) sets out an integrated perspective to water management in river catchments and river basin districts and is a key driver in the movement towards Integrated River Basin Management. Integrated river basin management must deliver objectives related to the WFD in the wider context of various other stakeholder interests, for example related to flooding, water resources, employment and cost. In managing such complex systems, a specific objective can be achieved through different management actions. Likewise, a specific management action can have implications for multiple objectives. Synergies or conflicts between specific objectives and between specific actions are likely to occur, and need careful consideration in order to increase the efficiency of planned management actions. However, such integrated decision making is a very difficult and highly complex task, which cannot easily be accomplished by either single or groups of planners. Integrated modelling tools to facilitate and enhance communication within a group of decision-makers and inform a more objective and evidence-based multi-criteria decision-making process are required. The scope for the development of such an integrated tool is being tested by the Catchment Science Centre (CSC) at The University of Sheffield. The CSC and the Environment Agency are jointly developing a tool termed the Macro-Ecological Model (MEM). The MEM is developed as a consistent framework for the integration of knowledge and information about environmental, social and economic processes and process-interactions that are affected by management actions and have impacts on multiple management objectives. The MEM enables knowledge from various different resources to be integrated, including empirical data, model results and even expert knowledge using a Bayesian Belief Network (BBN) approach. BBNs have the advantage of representing system understanding in an intuitive, graphical format. Furthermore, the approach provides the ability to explicitly account for uncertainties in model predictions. Therefore, the model framework provides a good tool for visualising system understanding and communicating uncertainties. Applied in a participatory process, it can support robust decision making in river basin management. The conceptual model framework is illustrated with examples from the prototyping study. The prototype model captures the process interactions affecting the management objectives "Ecological Status" (composed of both Biological Quality and Physico-chemical Quality) and "Flood Risk". It is planned to be later extended to incorporate further environmental, and also social and economic objectives.

Keywords: Integrated Catchment Management; Integrated Modelling; Decision support; Bayesian Belief Network

1. INTRODUCTION

The actions of many stakeholders influence the processes occurring in a catchment, and as a consequence many of the objectives that are valued within a catchment. For example, flood risk managers, water quality managers and land use planners all have different sets of objectives and they apply different set of measures to achieve these objectives. As the processes in a catchment are connected, it is very likely that interactions between different actions occur. These interactions can either be conflicting or synergetic with respect to the different management objectives. By considering these interactions in integrated management, the opportunity is provided to make management more efficient by focusing on synergetic effects and avoiding conflicting effects. New and innovative solutions can be found to make management more sustainable [Pascual, 2007]. However, integrated management also represents a great challenge. Complex system interactions are involved that are beyond the ability of individual planners and decision makers to grasp. Planners need to work together and objective tools are required to assist planners in making their management decisions [Reichert et al., 2007]. Decision support tools such as integrated models can enhance communication and support decision making by providing insights into the possible impacts of planned management interventions on multiple objectives [Matthies et al., 2007]. Integrated planning takes place at different spatial levels (e.g. national, regional, local). Therefore, integrated modelling tools are required at different spatial scales to support integrated management at these different levels.

This paper presents a conceptual framework, termed the Macro-Ecological Model (MEM), for the integration of knowledge and information about environmental, social and economic processes and process-interactions that are affected by management actions, and that and have impacts on multiple management objectives. The conceptual model framework is illustrated with examples from the prototyping study. The prototype model captures the process interactions affecting the management objectives “Ecological Status” and “Flood Risk”. The model is planned to be later extended to incorporate further environmental, and also social and economic objectives. The full implementation of the MEM would be designed to support the decision-making process in integrated river basin management. The tool is envisaged as a means to enhance communication and system understanding within fora of high-level decision makers in the second and third cycles of the river basin planning under the Water Framework Directive.

2. MODEL CONCEPT

High-level decision making deals with identifying appropriate sets of broad scale management actions whilst considering multiple management objectives in a catchment. The Macro-Ecological Model (MEM) is planned to incorporate a range of different environmental, social and economic management objectives that represent the interests of multiple stakeholders in a catchment. The model will capture the processes and process interactions affecting these different objectives, and it will allow us to make predictions regarding how different management scenarios could affect the status of these objectives in a catchment (Fig. 1).

The processes and process interactions need to be simplified as far as possible to develop a feasible model of the extremely complex catchment system. To accomplish this, the tool must integrate knowledge and information on processes and process interactions from many different disciplines. This knowledge and information will very likely be available in different forms. For example, process-relationships can be derived from empirical data. If no empirical data is available, process relationships could also be derived from existing models or from expert knowledge.

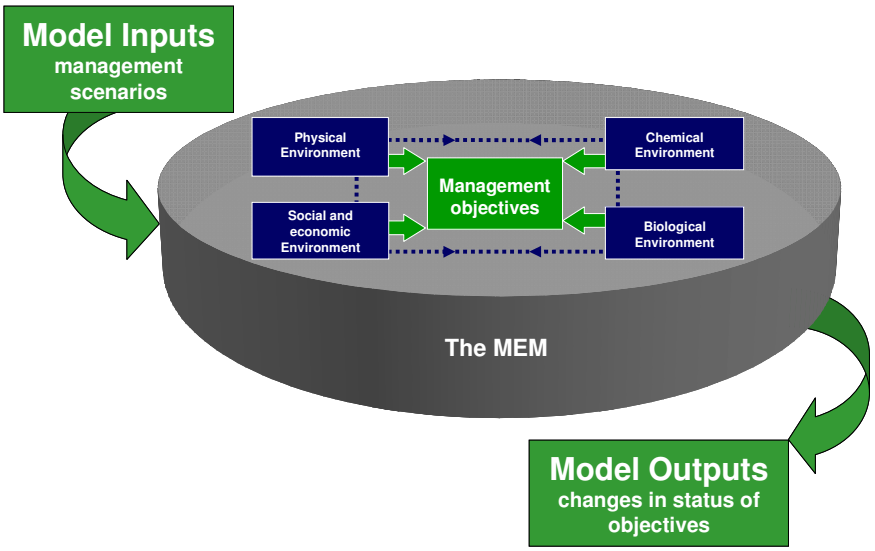


Figure 1. Schematic diagram of the Macro-Ecological Model.

3. MODEL APPROACH

The Macro-Ecological Model is implemented as a Bayesian Belief Network (BBN). A BBN is a graphical cause-effect network, where variables are linked together according to their dependencies [Jensen, 2001]. Associated with each variable is a conditional probability table (CPT), which specifies how this variable is affected by its influencing variables. The CPTs can be derived from data, external model results or expert knowledge, which provides the opportunity to integrate and combine information from different sources in one model. The BBN can be built to any level of detail and thus allows us to simplify complex relationships. Further advantages of the BBN approach are that uncertainties in model predictions can be explicitly considered, and rapid scenario analyses can be performed. The explicit consideration of uncertainties is an important challenge to decision making, particularly in the complex systems involved in integrated river basin management.

4. MODEL CONSTRUCTION

A conceptual framework for the development of the MEM was developed which involves six major steps that are detailed below:

4.1 Identification of index variables

Index variables represent relevant aspects of management objectives within the MEM. The definitions of index variables are ideally based on classification schemes that are currently used within the Environment Agency (EA) in the UK to evaluate the status of management objectives. The index variables must also be a useful basis for decision making. To ensure that these requirements are met, all index variables are defined in consultation with experts from different management functions within the EA during the prototype study. In a full implementation of the MEM the collaborative definition of index variables would be extended to all stakeholders in a catchment. Figure 2 shows the index variables that were identified to represent aspects of the objectives “Ecological status” and “Flood risk” that were chosen to be included in the prototype model to test the development of the MEM.

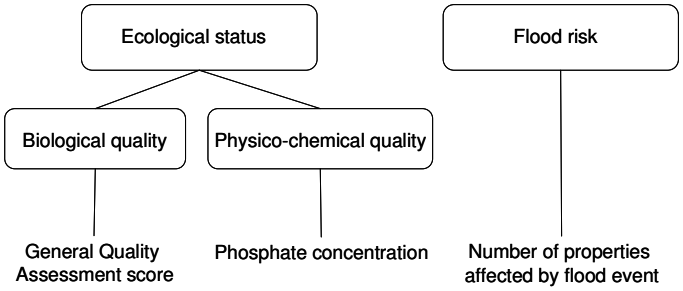


Figure 2. Derivation of index variables to represent relevant aspects of the management objectives "Ecological status" and "Flood risk" in the MEM prototype. The General Quality Assessment score for biology indicates how much the community of macroinvertebrates in the river is affected by organic pollution. The score is currently used in the UK as a basis to the assessment of biological quality as part of ecological status assessments under the WFD.

4.2 Development of conceptual networks

Conceptual cause-effect networks are developed around each identified index variable. Based on literature reviews and in consultation with EA experts as well as experts from the University of Sheffield and other institutions, conceptual cause-effect networks are developed to capture knowledge on processes and process interactions that could have impacts on the index variables. The aim is to reduce the chance that relevant pathways are neglected in the final MEM model. Also, the conceptual models developed can be the basis for a later expansion of the MEM if further data and knowledge become available, or if new management actions need to be integrated. Figure 3 shows the conceptual model that was developed for the index variable "phosphate concentration". The network summarises which processes influence the phosphate concentration in a river and how certain interventions, such as introducing restored wetlands, introducing riparian buffers or applying P-fertilisers in agriculture, could affect the phosphate concentration in the river.



Figure 3. Example of a conceptual cause-effect network linking management variables (green bubbles) to index variable (grey bubble) via a network of intermediate variables (CSO = combined sewer overflow, STW = sewage treatment work).

4.3 Simplification of conceptual networks

To enable the implementation of sub-networks as Bayesian Belief Networks (BBNs), the conceptual models need to be simplified. To do that, cause-effect links and variables with minor relevance are excluded, as well as links and variables that are not influenced directly or indirectly by any of the management actions under consideration. This process of simplification was conducted in collaboration with stakeholders to ensure that they remained confident in the conceptual basis of the MEM. Appropriate management actions to be considered in the MEM prototype are identified in consultation with EA planning experts. Other links and variables that are excluded from the network are those that can not be specified due to insufficient data and knowledge. If such data and knowledge become available at a later stage, the BBN structure could be extended according to the conceptual network (Fig. 3). Current data and knowledge gaps that limit an approach such as the MEM will be documented and reported as a record of the potential barriers to a full implementation of the MEM. Figure 4 shows the simplified network which describes how phosphate concentration is affected by agricultural land management, riparian buffers, phosphate inputs from point sources (i.e. combined sewer overflows and sewage treatment work effluents) and morphological modifications, such as introducing restored wetland, embankments or channel resectioning.

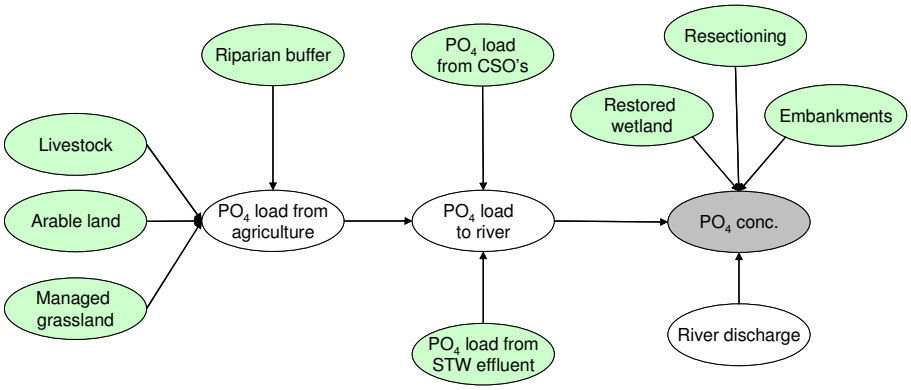


Figure 4. Example of a simplified network suitable for implementation as a BBN (CSO = combined sewer overflow, STW = sewage treatment work).

4.4 Specification of sub-networks

The BBN sub-networks are specified based on information and knowledge from different resources. Where empirical data is available, this will be used to specify network components. If no data, but existing and well-tested models are available, these can be used to generate synthetic datasets, which can then be used to populate parts of the BBN. If only expert knowledge is available to inform network components, CPTs can be elicited from experts to specify the links. If data become available at a later stage, they can be used to update the model and knowledge-based network specification. Experts from various disciplines are consulted to identify data availability, model availability, and finally to gather expert knowledge for the network specification. The network specification can differ for different river types to take into account the fact that many processes are affected by relatively local-scale characteristics (e.g. altitude, soil type, geology). In this way, the definition of a river type would account for significant differences in processes and process relationships due to local-scale characteristics. Suitable river typologies for the MEM are under consideration. For example, the WFD-defined river typology could be included, which is based on the three characteristics “altitude”, “geology” and “catchment size”. This typology would be based on commonly available data and it could also be an intuitive basis for expert knowledge elicitation.

Figure 5 shows the phosphate sub-network, which is planned to be specified based on model simulations of two external models: the PSYCHIC model [Davison et al., 2008], which predicts how land management affects phosphate loads from agriculture and the SIMCAT model [EA, 2006], which predicts how the phosphate concentration in a river is affected by diffuse and point sources of phosphate, and by the river discharge. These two models do not take into account the effects of riparian buffers, restored wetlands, embankments and channel dredging. Therefore, knowledge on the influences of these variables on phosphate loads from agriculture, and on phosphate concentration in the river, respectively, will be elicited from experts. This sub-network has been used to test the approach of using exiting mechanistic models to specify a BBN, and to combine results from these mechanistic models with expert knowledge in a single BBN. Other sub-networks being developed as part of the MEM project are testing different approaches to specification of the BBN, including the combination of data-based and expert-knowledge based networks.

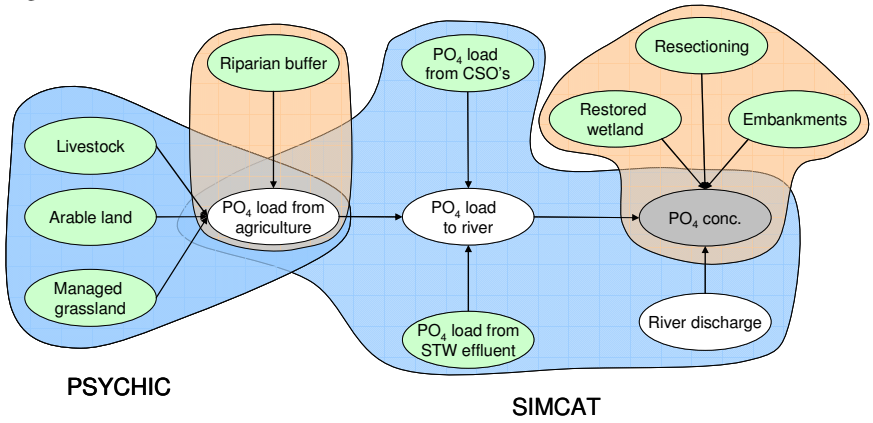


Figure 5. Example of a simplified network to be specified based on external model simulations (blue background) and expert knowledge (red background); (CSO = combined sewer overflow, STW = sewage treatment work).

4.6 Merging of sub-networks

Once the different sub-models for biological quality, phosphate and flood risk are specified and tested, they are merged into a single network. For the prototype model, sub-networks for phosphate, biological quality and flood risk will be linked together. A subset of the combined network is shown in Figure 6. Links between sub-networks are provided via index variables and management variables. For example, the index variable "phosphate concentration" is an input to the biological quality sub-network. Management actions such as introducing embankment, restored wetlands and resectioning have impacts on the biological quality component, the phosphate sub-model and the flood risk sub-model.

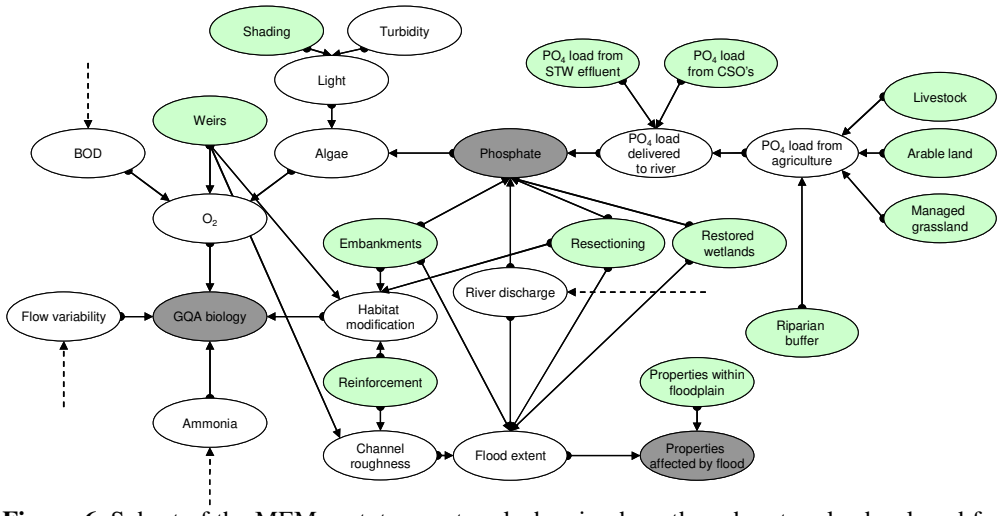


Figure 6. Subset of the MEM prototype network showing how the sub-networks developed for the three index variables are linked together (green bubbles indicate management variables; grey bubbles indicate index variable; dashed arrows indicate connections to further model components; BOD = Biological Oxygen Demand, GQA biology = General Quality Assessment score for biology; CSO = combined sewer overflow; STW = sewage treatment work).

4.6 Model evaluation and updating

The predictions of model components that were specified based on empirical data will be evaluated based on test-datasets. Model parts that were specified based on external model simulations or expert knowledge and for which no data are available will undergo a plausibility analysis that involves an evaluation of model predictions by groups of experts. As the model development is an iterative process, the model specification can be updated as further information and knowledge becomes available. For example, if empirical data become available for a model part that is specified based on expert knowledge, this empirical data can be used to update the conditional probability tables for that model part, which will decrease the prediction uncertainties.

5. MODEL APPLICATION

To predict changes at the catchment scale, the predictions of the river-type specific models have to be aggregated. In a simple approach, each river type-specific model would be applied once and the predictions would be aggregated by assigning weights to the predictions according to the frequencies of the river types in the catchment. In the most realistic case, the model would be applied for every single water body within a catchment. This would allow us to estimate the catchment's response to a highly disaggregated management scenario, because different combinations of management actions could be assumed in each water body. The catchment-scale outputs could then be represented either as averages (e.g. average General Quality Assessment score for biology in a catchment or Basin), or as numbers of exceedences of a certain threshold (e.g. number of water bodies with good biological quality in a catchment or Basin). The model could finally be applied to compare a baseline scenario to different management scenarios for a catchment. In different scenarios the impacts of management actions, such as, introducing wetlands in 20% of the lowland water bodies or reducing livestock numbers by 10%, could be tested and evaluated with respect to the multiple management objectives.

6. CONCLUSIONS

The Macro-Ecological Model project has developed as a consistent framework for knowledge integration, using BBN technology. As such, it is designed to be a tool for high-level decision support in integrated catchment management, which brings together knowledge from different disciplines to support a more holistic evaluation of planning alternatives. The Bayesian Belief Network approach is well suited to integrating knowledge from different resources. It also provides the opportunity to perform rapid scenario analyses, which makes it a very practicable tool to be applied in a planning context. The possibility to take modelling uncertainties explicitly into account enables robust decision support [Schlüter & Rüger, 2007]. Previous research in the area of decision-support systems has pointed out that decision-support tools are only accepted by their potential users if these users are involved in the model development from the beginning [Borowski & Hare, 2007]. Therefore, a close interaction between the model developers and the potential users is promoted in the development of the Macro-Ecological Model. The intuitive model structure of the BBN and the integration of information from trusted sources (e.g. EA data, models and expert knowledge) should support the acceptance of the model amongst its potential users.

ACKNOWLEDGEMENTS

The Macro-Ecological Model project (SC060039) is funded by the Environment Agency and the University of Sheffield. We would like to especially thank all experts from the Environment Agency, the University of Sheffield and other academic institutions who volunteered to support the development of the MEM during various group discussions.

REFERENCES

Borowski, I. and M. Hare, Exploring the gap between water managers and researchers: difficulties of model-based tools to support practical water management. *Water Resource Management 21*: 1049-1074, 2007.

Davison, P., P.J.A. Withers, E.I. Lord, M.J. Betson and J. Strömqvist, PSYCHIC – A process-based model of phosphorus and sediment mobilisation and delivery within agricultural catchments. Part 1: Model description and parameterisation. *Journal of Hydrology 350* (3-4): 290-302, 2008.

Environment Agency, SIMCAT 9.4 – A Guide and Reference for Users. 223 pages, 2006.

Jensen, F.V., Bayesian Networks and Decision Graphs. Springer, New York, 2001.

Matthies, M., C. Giupponi and B. Ostendorf, Environmental decision support systems : Current issues, methods and tools. *Environmental Modelling & Software 22*: 123-127, 2007.

Pascual, P., Avoiding Tragedies of the Intellectual Commons through Integrated Impact Assessments. *Water Resource Management 21*: 2005-2013, 2007.

Reichert, P., M. Borsuk, M. Hostmann, S. Schweizer, C. Spoerri, K. Tocker and B. Truffer, Concepts of decision support for river rehabilitation. *Environmental Modelling & Software 22*: 188-201, 2007.

Schlüter, M. and N. Rüger, Application of a GIS-based simulation tool to illustrate implications of uncertainties for water management in the Amudarya river delta. *Environmental Modelling & Software 22*: 158-166, 2007.

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

MoGIRE: A Model for Integrated Water Management

Arnaud Reynaud [a], **Delphine Leenhardt**[b]

[a]*Toulouse School of Economics, (LERNA-INRA), (areynaud@toulouse.inra.fr)*

[b]*UMR1248 INRA-INSAT AGIR, (Delphine.Burger-Leenhardt@toulouse.inra.fr)*

Abstract:

In a context of water scarcity where competition for water allocation is increasing, modelers are often asked to move from specific water use models (either agricultural, industrial or residential) toward more integrated frameworks including an explicit representation off water users relationships. This is the main objective of MoGIRE ("**Mo**dèle pour la **G**estion **I**ntégrée de la **R**essource en **E**au") which is a model developed by an interdisciplinary team for integrated water management at the river basin level. MoGIRE includes a nodal representation of the water network and allows to optimize water use under several possible scenarios (agronomic, climatic or economic). Agricultural, urban and environmental water uses are represented in MoGIRE using mathematical programming and econometric approaches. The model optimizes at each date the allocation of water across agricultural and urban water demands in order to maximize the social surplus derived from water consumption given the constraints imposed by the water network. An application of the model is proposed for the Neste system located in South-West of France.

Keywords: Integrated water management, optimization-simulation model, agronomic-economic modeling, river basin.

1 INTRODUCTION

Climate change and growing water needs have resulted in many parts of the world in water scarcity problems that must by managed by public authorities. Hence, policy-makers are more and more often asked to define and to implement water allocation rules between competitive users. This requires to develop new tools aiming at designing those rules for various scenarios of context (climatic, agronomic, economic). Models have been developed for each type of water use (see for instance Dalhuisen et al. [2003] for a survey of residential water demand models or Couture and Bontemps [2002] for a model of agricultural water demand). However, very few models link these different uses in the context of a particular area, while such an integrated approach is a relevant stake for designing regional water and land policies.

The lack of such integrated models can be explained by the difficulty of integrating models developed by very different disciplines and by the problem of scale change (collecting data on large area, arbitrate between the computational tractability of models and their level of aggregation). However, modelers are more and more asked to deal with large basin scales while analyzing some policy impacts at very high detailed levels. These contradicting objectives require to develop new modeling tools. The CALVIN economically-driven optimization model developed for managing water in California is a good example of this type of framework, Draper et al. [2003]. Cai et al. [2002] and Cai et al. [2003a] propose another integrated water management model for the Syr Darya Basin (Central Asia). Recent reviews of the literature on integrated water management at the basin level include Cai et al. [2003b], Letcher et al. [2007] or Cai [2008].

We present here an original framework for integrated water management at the river basin scale called MoGIRE ("**Mo**dèle pour la **G**estion **I**ntégrée de la **R**essource en **E**au"). MoGIRE is an integrated framework, currently under development by an interdisciplinary team within the APPEAU project (2007-2010) funded by the French National Research Agency within the "Agriculture and Sustainable Development" program. It is intended to optimize water use at the river basin level and to evaluate scenarios (agronomic, climatic or economic) for a better planning of agricultural and non-agricultural water use. In specific contexts such as in South-West France, water scarcity occurs during the summer when agricultural water needs and urban needs are maximum. Besides, water demand greatly varies within the water basin, with sub regions having competitive uses. Consequently MoGIRE accounts for such intra-annual conflicts and for the spatial heterogeneity of the water basin.

The remaining of the paper is organized as follows. In the next section, we present the integrated river basin model and we describe its main components. In the following section, we provide an empirical application to the Neste system (South-West of France). We conclude by summarizing our findings.

2 ARCHITECTURE OF MOGIRE

MoGIRE may be viewed as a multi-use water allocation model with an explicit representation of the water network at the river basin level.

2.1 Representation of the river basin

A river basin includes the network of the river and its tributaries, the land of the river catchment, and all related land and water users. Because the river basin is large and the users are numerous, we need a simplified representation of this complex system. Following Cai et al. [2003b] or Cai et al. [2003a], it is possible to view the water network as network objects (reservoirs, pumping nodes and confluence nodes) connected by some links (rivers, canals). This nodal network representation is a common framework for considering water allocation in river basins, Letcher et al. [2007]. In this type of representation, water users interact with the stream system in two ways, Letcher et al. [2007]. First, they may affect the generation of runoff and thus the volume of water reaching the stream. Second, they may involve direct extraction or use of water once it has reached the stream. We mainly focus here on this second type of interaction between water uses and the water network.

We used a nodal representation of the river network with a division of the catchment area into regions (see Figure 3 for a presentation of the Neste river basin where the model has been implemented). In rural areas, agriculture is the main water user. This is why we identified the regions using soil, climate and hydrological criteria. The methodology is fully described in Clavel and Leenhardt [2008] and resulted in A regions indexed by $a = 1, \ldots, A$. In these regions, water uses (agricultural, domestic and industrial) must then be modeled.

2.2 Representation of agricultural water demands

The first step of the analysis has been to characterize farmers located within each region $a = 1, \ldots, A$. Some GIS regional data on soil, climate and hydrology have been used. Agricultural regions are then characterized by specific cropping systems, climate conditions, soil characteristics, economic conditions (available agricultural area (in ha), number of farmers, crop prices, etc.). Moreover, each agricultural region is connected to the water network through one or several nodes. Water allocated to any agricultural region allows to produce crops which generates a net profit that will be optimized by the social planner.

The second step of the analysis has been to build for each region an agricultural production model allowing to derive the net profit from water use.

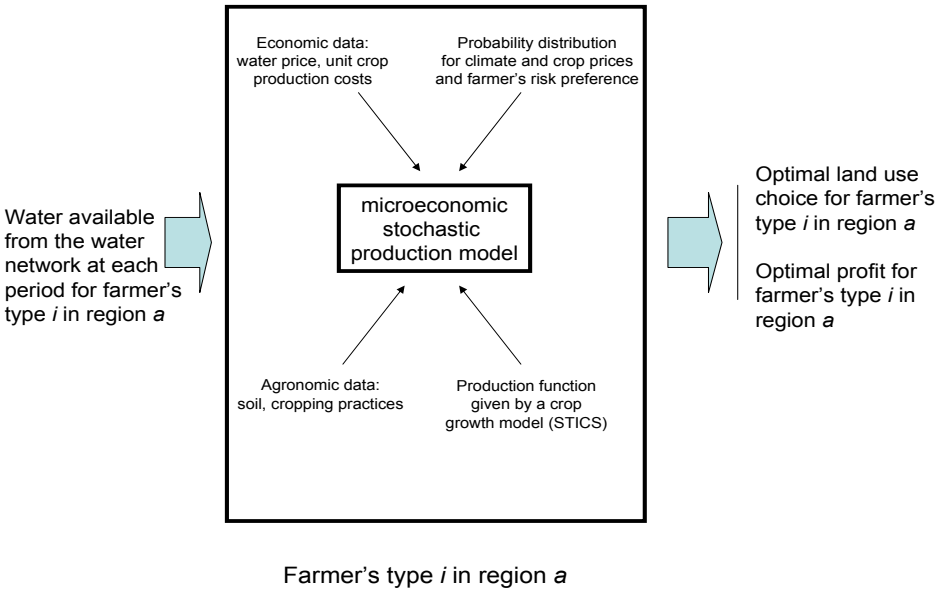


Figure 1: Agricultural production model for farmer's type i in region a

– Since Maton et al. [2005] and Maton et al. [2007] have shown that, at a regional scale, easily accessible farm characteristics could be good indicators of farmers' practices, we first have build, for the whole area, a typology of farmers. Following Clavel and Leenhardt [2008], the typology is based on farm size, cropping systems and irrigation intensity. This typology has allowed us to identify I farmer types we will index by $i = 1, \ldots, I$.

– Second, statistical agricultural databases gave us the number of farmers of each type in each region. The number of farmers of type i in region a is denoted by n_{ai}.

– Third, we have developed an individual production model in order to represent the behavior each farmer type i in each region a, see Figure 1. This individual model incorporates a risk linked to climate uncertainty and impacting land use decisions of farmers. These decisions are : allocating the available land across all possible crops (land use choices), choosing a sowing date for each crop (sowing date choices) and finally allocating for each crop at each date of the growing season the available water (water use choices). We assume that a farmer takes these decisions in order to maximize its expected utility of profit. This individual production model is presented in details in Reynaud [2008]. Notice that the agricultural production functions used in the model are derived from the STICS biophysical model of crop growth, see Brisson et al. [1998]. This crop growth model allows to characterize the relationships between agricultural water use and crop yields.

– Lastly, using the number of farmers n_{ai}, we aggregated the production outcome for all types of farmer within a given region to get regional land uses and regional net profit, see Figure 2.

2.3 Representation of urban water demands

The first step of the analysis has been to characterize urban water consumers (domestic and industrial users) located within each region $a = 1, \ldots, A$. The characterization of urban water consumers is based on socioeconomic data from various sources including population census and water agency files. Those regions are defined, in particular, by their size (number of households) and by the socioeconomic characteristics of households (income, age, housings, etc). All the urban water demands are connected to the water network through one or several nodes.

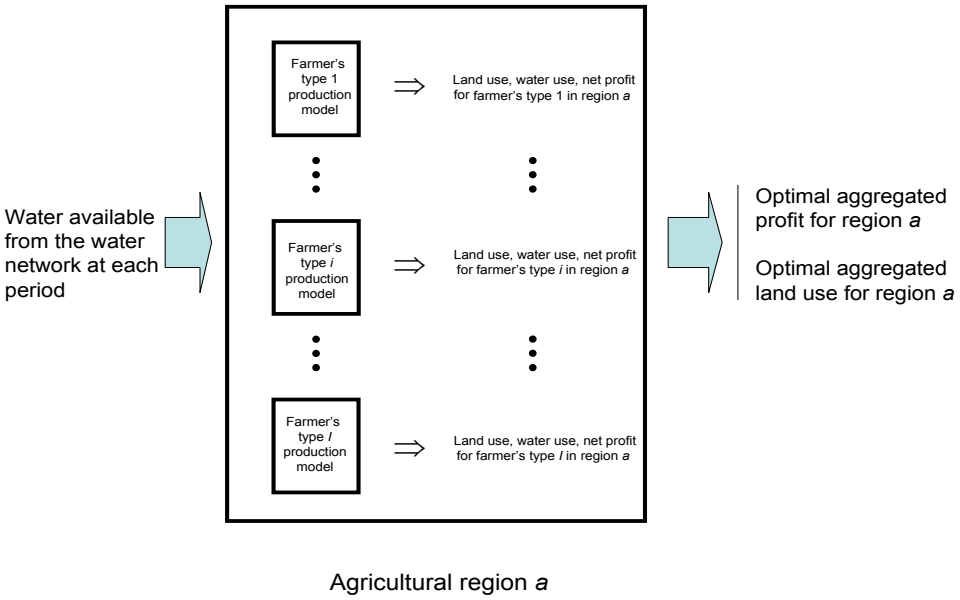


Figure 2: Agricultural production model for region a

Economists have developed frameworks for estimating the value for urban water consumption. Typically, this value corresponds to the marshallian surplus which can be derived from the urban water demand function. Hence, estimating the value for urban water consumption in each region requires first to estimate urban water demand functions, see Garcia and Reynaud [2004]. Dalhuisen et al. [2003] have conducted a meta-analysis on residential water demand based on 51 published articles. They report that 20 articles provide estimates of price and income elasticities based on a logarithmic demand function (either semi or double logarithmic). Following this previous literature (see also Nauges and Thomas [2000]), we assume that the water demand function in a given urban region can be approximated by a log-log form. Denoting by $\hat{Y}_a$ the daily water consumption of the representative residential water user in region a, we have:

$$\ln \hat{Y}_a = C_a + \sum_l \alpha_l \cdot \ln \mathbf{X}_{la} + \beta \cdot \ln p_a \tag{1}$$

where C_a is a constant, $\mathbf{X}_{la}$ represents a vector characterizing the residential water users and determining the water consumption (characteristics of households, income, housings, climate, etc), p_a is the peak period unit water price and β is the price elasticity of the urban water demand which is assumed to be negative. To estimate the parameters of the urban water demand function (Equation 1), we used various sources of data including population census and water agency files.

Having estimated the urban water demand function, we can compute the daily consumer surplus (CS_a) resulting from water consumption in region a. Then, we can compute the variation in urban consumer surplus due to a change in the water price from p to p' in region a. We denote it by $\Delta CS_a(p, p')$. Since the consumer's surplus is, by definition, measured by the area under the Marshallian demand curve, the change in daily consumer surplus following a change in the water price from p to p' is given by (see Strand and Walker [2003]):

$$\Delta CS_a(p, p') = \frac{1}{1+\beta} p' \hat{Y}_a(p') \Big[\Big(\frac{p}{p'} \Big)^{1+\beta_p} - 1 \Big]. \tag{2}$$

2.4 Representation of environmental water demands

Freshwater ecosystems need certain water flow regimes to sustain their animal and plant communities. This environmental water demand, allowing for instance to provide critical habitat for endangered species, increases the competition for water among users especially during the summer season. There is however no consensus on the economic value to be attributed to the environmental water demands. As a result, we have decided to introduce the use of water for environmental purpose as a set of flow constraints that must be satisfied at specific nodes of the water network and at specific periods (typically during the summer). This approach allows us to get the dual variables associated to these flow constraints. In other words, this approach allows us to measure the social cost for all water users from imposing these environmental water flows constraints.

2.5 The river basin water allocation problem

The model, including the representation of the water network, has been coded using the algebraic modeling language GAMS, see Rosenthal [2008]. GAMS is a high-level modeling system for mathematical programming and optimization. MoGIRE then allocates water across uses within a region and across competitive regions. The water is allocated to maximize the agricultural and urban economic value resulting from water use at each time within the year. This pursuit of economic objectives (aggregated social surplus resulting from water consumption) is then limited by water availability, water network characteristics, and minimum flow constraints.

3 Application of MoGIRE to the Neste System (France)

3.1 The Neste system

Within the APPEAU project, MoGIRE is currently applied on the Neste system located in South-West of France, see Figure 3. The Neste system is a system of rivers artificially recharged by upstream reservoirs located in the Pyrenees mountain through a single canal, the Neste canal. The Neste system covers 800,000 ha area and gathers the catchment areas of 14 main rivers. The land is mainly dedicated to agriculture (500,000 ha are cultivated from which 50,000 ha are irrigated).

This system has been chosen for application of MoGIRE first because it is managed by a single operator, the Compagnie d'Aménagement des Coteaux de Gascogne (CACG) and second, since there are significant water scarcity issues in that area that make the development of an integrated water management model relevant. For instance, the relationship between the CACG and farmers is defined by a contract. This contract specifies first, a discharge rate and second, a quota associated to each unit of discharge rate (l/s) subscribed by a farmer. Given the current price for the discharge rate, the demand of discharge rates by farmers exceeds the flow capacity of the system. The CACG must then manage a waiting list for quota allocation.

3.2 The representation of water demands

The regions have been defined as the intersection between CACG management units and Small Agricultural Regions (SAR), Clavel and Leenhardt [2008]. CACG management units are used by the operational water manager to optimize water releases between different rivers. The SARs, defined by the National Institute for Statistics and Economic Studies, are mainly based on a soil and climatic conditions homogeneity criterion. As a result, 67 regions have been identified in the Neste system. Those regions are characterized by specific cropping systems, specific climate and soil characteristics and by their connections to the water network.

601 local communities are located within the Neste system. The local communities have been aggregated in order to represent urban water consumers in each of the 67 region. Data used for estimating the residential water demands come from various sources including population census

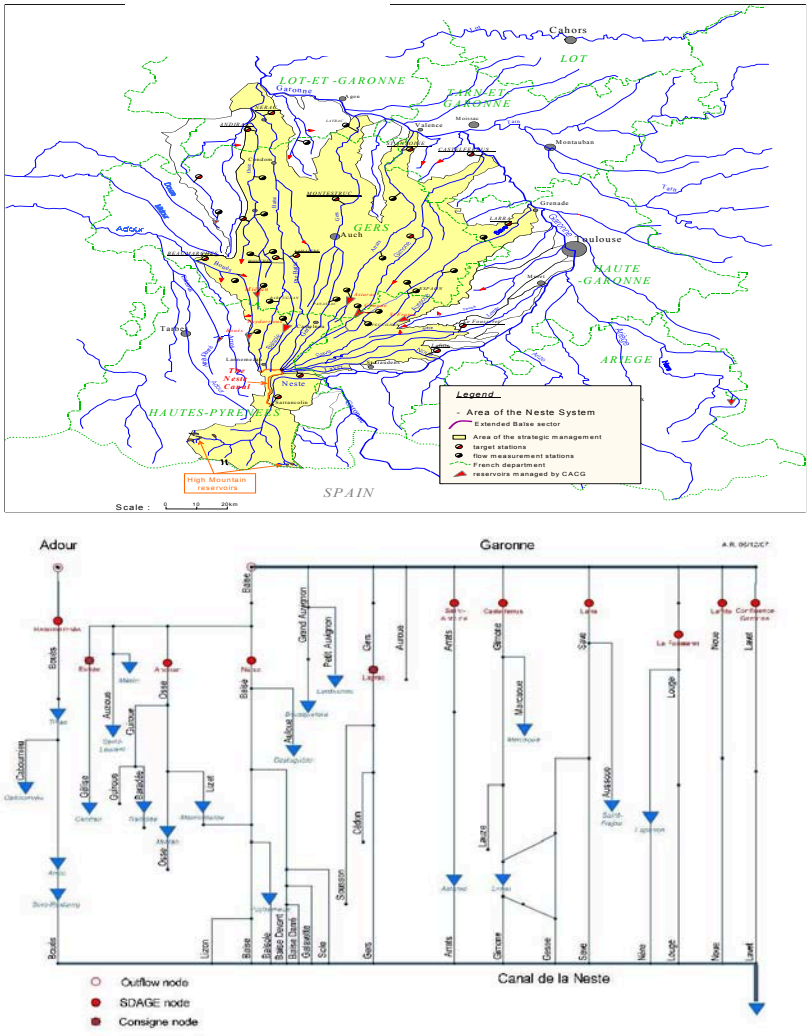


Figure 3: The Neste system (France) and its nodal representation

and water agency files based on local communities declarations. The dataset we have used covers 3 years, 1998, 2001 and 2004.

Concerning the environmental water demands, we have made the distinction between two minimum flow levels: the Crisis Minimum Flow (CMF) and the Target Minimum Flow (TMF). The CMF (in French, débit de crise) defines the minimum flow under which biological life is not possible. When the water flow is lower than CMF, no irrigation is allowed. The TMF (in French, débit objectif d'tiage) defines the water flow level above which biological life is not constrained. When the water flow is greater than TMF, irrigation is not limited.

The model then optimizes at each date the allocation of water across agricultural and urban water demands in order to maximize the social surplus derived from water consumption. MoGIRE also allows to simulate various agronomic, climatic or economic scenarios on the Neste system.

4 Conclusion

We have presented an original framework for integrated water management at the river basin. This model, called MoGIRE, includes a nodal representation of the water network. Agricultural,

urban and environmental water uses have been modeled using mathematical programming and econometric approaches. MoGIRE then optimizes water use at the river basin level and allows to evaluate scenarios (agronomic, climatic or economic) for a better planning of agricultural and non-agricultural water use. The model has been applied to the Neste system in Southwest of France. All the agricultural and the urban water demands have been identified and the whole model, including the nodal representation of the water network, has been coded using the algebraic modeling language GAMS. We are currently analyzing the robustness of the approach through scenario testing. The next step of the analysis will be to characterize the optimal allocation of water across competitive water users at each point of the network and at each date of the year.

Finally, we plan to evaluate several scenarios using MoGIRE (agronomic, climatic or economic). For instance, a relevant economic scenario could be a move toward peak-load pricing. The rational for implementing peak-load water pricing in an integrated river basin framework is that the value of water is likely to vary according the period of the year. In particular, it is likely that the social value is higher during the summer where the high competition across water users might result in scarcity rents. Peak-load pricing could then be used by the water network operator in order to send scarcity signals to water consumers. Agronomic scenarios may include the introduction of new crop varieties more resistant to water stress or the introduction of new cropping practices (e.g. early sowing of early maize varieties Lorgeou et al. [2006], replacing maize by sorghum (Amigues et al. [2006]), etc.). Lastly, since weather conditions are introduced as determinants of agricultural and urban water demand models, some scenarios of climate change may be tested (e.g. SRES scenarios (Terray and Braconnot [2007]). Agronomic scenarios may include the introduction of new crop varieties more resistant to hydric stress or the introduction of new cropping practices. Lastly, since weather conditions are introduced as determinants of agricultural and urban water demand models, some scenarios of climate change may be tested.

ACKNOWLEDGMENTS

This work is a part of the project APPEAU funded by the program ”Agriculture and Développement Durable” from the Agence Nationale pour la Recherche (ANR). The authors gratefully acknowledge support from the ANR grant that has made this work possible.

REFERENCES

Amigues, J.-P., P. Debaeke, B. Itier, G. Lemaire, B. Seguin, F. Tardieu, and A. Thomas. Sécheresse et agriculture. réduire la vulnérabilité de l’agriculture à un risque accru de manque d’eau. expertise scientifique collective. Expertise scientifique collective, Rapport, INRA (France), 380 pages + annexes., 2006.

Brisson, N., B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget, B. Nicoullaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Durr, G. Richard, N. Beaudoin, S. Recous, X. Tayot, D. Plenet, P. Cellier, J.-M. Machet, J. M. Meynarda, and R. Delécolle. Stics: a generic model for the simulation of crops and their water and nitrogen balances. 1. theory and parameterization applied to wheat and corn. *Agronomie*, 18:311–346, 1998.

Cai, X., D. C. McKinney, and L. S. Lasdon. A framework for sustainability analysis in water resources management and application to the Syr Darya Basin. *Water Resources Research*, 38: 21–1, June 2002.

Cai, X., D. McKinney, and L. Lasdon. An integrated hydrologic-agronomic-economic model for river basin management. *Journal of Water Resources Planning and Management*, 129(1):4–17, 2003a.

Cai, X., M. Rosegrant, and C. Ringler. Physical and economic efficiency of water use in the river basin: Implications for efficient water management. *Water Resources Research*, 39(1):1013, 2003b.

Cai, X. Implementation of holistic water resources-economic optimization models for river basin management - reflective experiences. *Environ. Model. Softw.*, 23(1):2–18, 2008.

Clavel, L. and D. Leenhardt. Development of matrices to build scenarios of cropping systems distribution for integrated catchment assessment. a proposition for an irrigated area in south-western france. Mimeo, UMR1248 INRA-INSAT AGIR, 2008.

Couture, S. and C. Bontemps. Irrigation water demand for the decision maker. *Environment and Development Economics*, (7):643–657, 2002.

Dalhuisen, J., R. Florax, H. deGroot, and P. Nijkamp. Price and income elasticities of residential water demand: A meta-analysis. *Land Economics*, 79:292–308, 2003.

Draper, A., M. Jenkins, K. Kirby, J. Lund, , and R. Howitt. Economic-engineering optimization for california water management. *Journal of Water Resources Planning and Management*, 129 (3):155–164, 2003.

Garcia, S. and A. Reynaud. Estimating the benefits of efficient water pricing in france. *Resource and Energy Economics*, 26(1):1–25, 2004.

Letcher, R. A., B. F. W. Croke, and A. J. Jakeman. Integrated assessment modelling for water resource allocation and management: A generalised conceptual framework. *Environ. Model. Softw.*, 22(5):733–742, 2007.

Lorgeou, J., A. Bouthier, J.-P. Renoux, and G. Cloute. Stratégie d'évitement en mas grain dans le centre ouest. adapter le cycle aux contraintes hydriques par la précocité? *Perspectives agricoles*, mars:62–67, 2006.

Maton, L., Leenhardt, and B. J.-E. Geo-referenced indicators of maize sowing and cultivar choice for better water management. *Agronomy for Sustainable Development*, 27(3):377–386, 2007.

Maton, L., D. Leenhardt, M. Goulard, and B. J.-E. Assessing the irrigation strategies over a wide geographical area from structural data about farming systems. *Agricultural Systems*, 86: 293–311, December 2005.

Nauges, C. and A. Thomas. Privately-operated water utilities, municipal price negotiation, and estimation of residential water demand: The case of france. *Land Economics*, 76(1):68–85, 2000.

Reynaud, A. Agricultural land allocation and intra-annual optimization of agricultural water use under climate and price uncertainty. Mimeo, LERNA-INRA, University Toulouse 1., 2008.

Rosenthal, R. E. *GAMS: A user's guide*. GAMS Development Corporation, 2008. Washington DC, USA.

Strand, J. and I. Walker. The value of water connections in central american cities: A revealed preference study. Technical report, University of Oslo, 2003. http://folk.uio.no/jostrand/watervaluepaper.pdf.

Terray, L. and P. Braconnot. Livre blanc escrime. étude des scénarios climatiques réalisés par l'ipsl et météo-france. Miméo IPSL and Météo-France, 70 pages, 2007.

iEMSs 2008:International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Step by step calibration of an Integrated Model System for irrigation management

T. Arnold [a], **T. Berger**[a], **H. Uribe**[b]

[a]*Universität Hohenheim (490d), 70593 Stuttgart, Germany (tarnold@uni-hohenheim.de)*

[b]*INIA Quilamapu, Chillan, Chile*

Abstract To assess how the multiple processes that constitute irrigation land use interact, the basin-level hydrological model WASIM-ETH and a bioeconomic model MP-MAS were embedded into a common framework. Following conceptual integration of theories, we here describe the integration of data and joint calibration and validation of both models. Methodologically, interacting variables are first specified from data, then from model outputs, and then dynamically coupled. Interdisciplinary cross-checks and sensitivity analyses improved calibration. For irrigation management at basin scale, we indicate that physical scarcity of water, which restricts the current land use pattern in drier years, and allocative water scarcity which indicates inefficient water right markets, coexist in one basin, in different irrigation sectors. Also, due to heterogeneity of asset distribution among farmers, effective constraints vary considerably, so that water right consideration are not relevant to many. Thus, policies require in-depth analysis of the target groups at individual level.

Keywords: Catchment scale; Irrigation; Multi-agent System

1 INTRODUCTION

A comprehensive scientific theory of agricultural land use change and resource use (land, water) does not exist. Rather, a multitude of different theories coexist - not only amongst disciplines, but also within disciplines. Phenomena in the real-world are the product of multiple concurrent processes; ecological, soil, meteorological and hydrological conditions constrain the decisions of farmers, who must also deal with market conditions, institutional constraints and incentives. In addition, their own knowledge base also has impact on what is, cumulatively, land use change [Parker et al., 2003]. Multiple processes interact at different scales, each examined by different disciplines and institutions. This produces an abundance of theories and alternative explanations, but data tends to be scarce and disconnected.

The integration of processes into a single framework (and model) poses at least three types of conceptual challenges. The first challenge is the conceptual integration of theories, as a basis for the coupling of models and data, to generate scientifically sound computer code. This challenge includes formal descriptions of exchanged variables, inconsistencies in model assumptions across model components, and sequencing [Argent, 2004]. The second type of challenge, addressed in this paper, is the process of bringing data and integrated models together; the simultaneous calibration and validation of models with one data base, in order to understand interaction. A third challenge is epistemological, questioning under which conditions the resource-intense integration of models can actually generate new knowledge. Beven [2007] relates it to the context of a case study, as well as to the organization of the knowledge-generating research process.

1.1 The project 'Integrating Governance and Modeling'

The integrated model system presented here is being developed within the project *Integrating Governance and Modeling*, a CGIAR *Challenge Program on Water & Food* project. Its objective is to explore policy options that improve the management of water resources at both the local and the regional levels, aided by integrated computer models that resolve the micro-scale. Furthermore, this project analyzes existing governance structures and their ability to distribute water equitably and efficiently (Project website http://www.igm.uni-hohenheim.de/igm).
The case study in Chile assessed challenges in water management through a participatory approach with key stakeholders, and identified policy options to address these challenges, along with policy evaluation criteria. Based on this, model use cases are being developed jointly with farming organizations and with governmental institutions. An interdisciplinary data base was compiled, which combines GIS data, socio-economic data from census and farm surveys, crop production data, plant data, canal data, and registries on land use and water rights. Under a (semi-) predictive modeling paradigm, a model system was built that integrates the basin-scale distributed hydrological model WASIM-ETH [Schulla and Jasper, 2007] and a bio-economic, agent-based model MP-MAS used for agricultural water use analysis [Berger et al., 2007]. To study the relevance of interactions, models can either be used and calibrated separately [Uribe et al., 2008], or dynamically-coupled.
The study area contains the watersheds of the Putagán, Ancoa, Achibueno and Longaví rivers, in the Maule Region of Chile. At a total area of 5300 km^2, this region contains approximately 100.000 hectares of agriculturally used lands, mostly with snow melt from the Andean mountains through a complex distribution system of canals. Water is distributed from river organizations, to smaller user organizations summarized as irrigation sectors, and finally to farmers, according to the water rights system of Chile.

1.2 Integrated modeling: A calibration and validation challenge

The data base of our case study contains maps, constant technical parameters, economic cross-sectional data for two time periods (1996/7 and 2006/7), time series on price- and cost developments, and hydrological and meteorological measurements (time series). Qualitative expert opinions on the transient period exist, but are not always coherent. The analytical challenge is to find a theory-based explanation that is consistent with data for 1996/7 and 2006/7, with transient time series, and which matches expert opinion as well as possible.
Alternative model formulations can lead to the same outcomes ('equifinal models', Beven [2001]), or they can look equifinal because outcome aggregates and coarse empirical data conceal micro-scale logical fallacies. Here, the integration of evidence from empirical data and from theories across disciplines and at multiple scales can greatly reduce uncertainty between alternative explanations, by proving hypothesis as inconsistent or incomplete.
The objective of this publication is to demonstrate the value of an integration framework as methodology, rather than offering policy implications. All data interpretation is still preliminary.

2 METHODS

2.1 The Integrated Modeling System (IMS)

The bio-economic, agent-based model MP-MAS describes and spatially captures land use decisions of farmers [Berger, 2001]. At individual level, a constrained-optimization model mimics production and farm investment decisions, and the resource use involved (land, water, credits, labor). *Inter alia*, modules include land and water markets, soil nutrient dynamics, diffusion of innovation in networks, water sharing within irrigation sectors, and a lumped routing model between sectors. At yearly time steps, MP-MAS recursively updates the asset endowment of farmers,

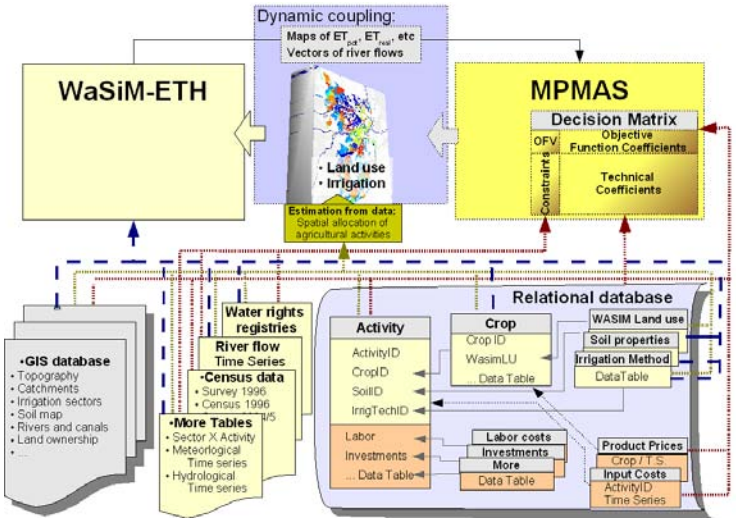


Figure 1: An integrated model system that allows standalone and coupled model runs, with two models WASIM-ETH and MP-MAS, coupling functions, and data

so that aspects of heterogeneity and distribution can be analyzed.
The WASIM-ETH is a process-based, fully-distributed hydrological basin model [Schulla and Jasper, 2007]. For each grid cell, vegetation cover can be parameterized (single or multi-layered). Processes include interception, surface evaporation, infiltration into the top soil layer, surface runoff. The model also includes an unsaturated 1D-vadose zone module, based on Richards' equation, in which plant transpiration is withdrawn while limited by ET_{pot}. Excess water percolates. Groundwater flow can be parameterized, or dynamically modeled with a 3-dimensional advection model. The representation of above-ground vegetation characteristics and the vadose zone make WASIM-ETH a good choice for modeling irrigation. Routing of surface water is based on a sub-basin approach, derived from a topographic analysis. Water channeled between sub-basins is externally parameterized as extractions, inflows or bypasses. The original irrigation module was extended by our working group [Arnold et al., 2008].
Both models MP-MAS and WASIM were integrated conceptually and technically (Fig. 1) and linked to an interdisciplinary data base. In standalone mode, models can be calibrated separately by domain experts. During this calibration phase, inconsistencies within the shared data base were identified and eliminated. The framework also allows for the dynamic coupling of the components, such that agricultural land use is computed from the socio-economic model at a yearly time step and reported to the hydrological model through a translation interface. For the monthly decision on irrigation quantities and its abstraction, the model is run twice for each month: First, WASIM estimates crop water demand and water availability *without* irrigation. Then, the economic model develops a monthly irrigation plan and reports it to WASIM. Agricultural production is continued within MP-MAS and the hydrology in WASIM. After each month, maps (e.g. real evapotranspiration) are reported to MP-MAS. The translation interface consists of a hierarchical sequencer and spatial and temporal rescaling functions, partly drawing on the relational data base. Data is passed between applications using the Typed Data Transfer library. Technical implementation has been finalized and the software code has been technically verified by showing that outputs from a reconstructed standalone version corresponds to coupled results.

A key conceptual integration challenge is the cyclic reuse of irrigation water. Physical-based modeling requires micro-scale data (soil characteristics, methods of irrigation, drainage, scheduling, and the characteristics of the canal system). Hansen et al. [2007] discusses coupled approaches, and concludes that the problem of micro-level heterogeneity for irrigation has not been resolved.

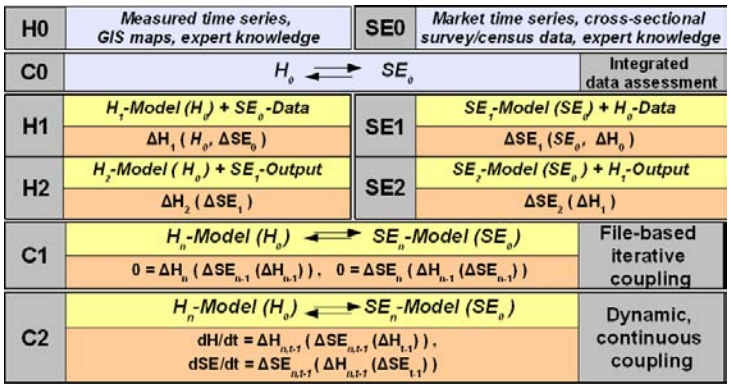


Figure 2: Calibration and validation of the hydrological (H), the socio-economic (SE) and the integrated system (C). Integration levels are data + data (0), model + data (1) and two models (2)

Within our IMS, we apply dual concepts [Arnold et al., 2008]: In WASIM, cyclic reuse within sub-basins is eliminated and *effective* irrigation efficiency η_{eff} is used. MP-MAS uses field-level efficiency η_{field} and estimates the reuse fraction within the EDIC module. This duality facilitates the modeling of water as an economic constraint for production, while maintaining a basin-wide perspective and ensuring consistency at all scales. Return flows from one sub-basin to another can use the natural path, but also artificial canals, so modeling remains problematic.
Within the MP-MAS model, the handling of space was revised, and data handling was improved. The internal time stepping of the code was restructured, which required further modularization.
Within WASIM, the original irrigation module was extended. Potential evapotranspiration can now be computed with a crop-coefficient. Irrigation water can be applied as precipitation, but also directly to the surface. Access water may be directly applied as runoff, or as percolation through parameterization, and scheduled rotational schemes are avoided. Routing and the irrigation restriction was also modified [Arnold et al., 2008].

2.2 Framework for stepwise calibration

Three stages in the calibration and validation of our coupled model, which includes a hydrological component H and a socio-economic component SE, can be distinguished (Fig. 2): (1) The integration of interdisciplinary data and consistency tests on data, (2) the calibration and validation (C&V) of each component H and SE with that data, and (3) the coupling C of calibrated models - first file-based, then dynamic. At each step, inconsistency might require revisiting concepts, further calibration or data improvement.
Both modules H and SE build on data, on expert knowledge, and on 'boundary conditions' from the other domain ('coupled variables'), which are uncertain and mutually dependent.
After data from both domains (H_0 and SE_0) is collected, the first step is the creation of a common information base (C_0). Already, this step requires full conceptual integration. Using C_0, baseline scenarios of coupled variables are then defined, as boundary conditions for standalone calibration. Both domains benefit from the exchange of data and expert knowledge. Within the usual, discipline-specific model analysis of that base line, the sensitivity analysis to the coupled variables gives $\Delta H_1(H_0, \Delta SE_0)$ and $\Delta SE_1(SE_0, \Delta H_0)$. When assessing biophysical-economic interaction, the team can focus analytical and data collection efforts on improving those coupled variables with a strong influence on model outcome indicators.
In the second step (H_2 and SE_2), boundary conditions are created from model results. Logical consistency (correlation structures) are then preserved and finer time resolution makes interpolation unnecessary. Recalibration improves models and model sensitivity $\Delta H_2(H_0, \Delta SE_1)$ and $\Delta SE_2(SE_0, \Delta H_1)$. Procedurally, steps H_2 and SE_2 require revisiting the conceptual integration

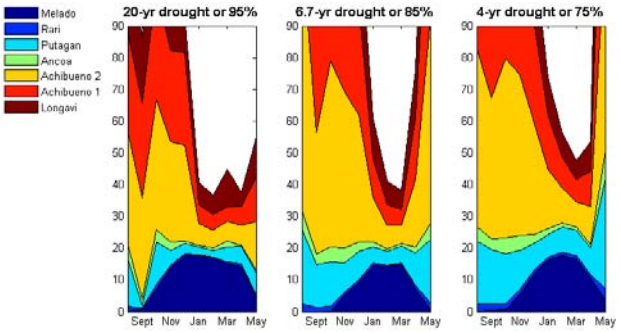


Figure 3: River flows at seven stations in dry years (20-year, 6.7-year, and 4-year drought).

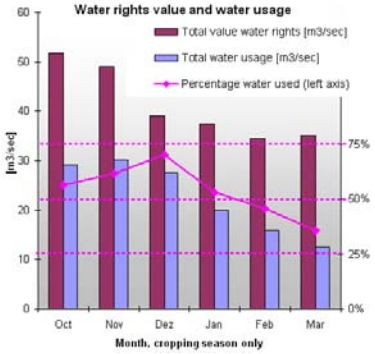


Figure 4: Relative share of water rights used, during dry year.

and data. Technically, data exchange between both models is clarified, data formats are revisited, and data translation functions required for coupling become apparent and can be automatized.
In the third coupling step $C1$, we investigate the first-order impact of systems on each other. File-based coupling of models suffices, which builds on automatized data translation routines. Models are rerun and outputs are iteratively exchanged $SE_n(H_{n-1}), H_n(SE_{n-1})$, until convergence is reached ($H_n - H_{n-1} \approx 0$). Eventually, both models are re-calibrated. It is recommendable to choose the time horizon such that the complexity of interactions remains reasonable. This step already allows for full validation against observed data. As an example, we analyze the impact of a new technology applied by upstream farmers. What is the first-order impact on downstream farmers (first iterative step), and how will the adaptation of downstream farmers further modify the hydrology at convergence?
Finally, dynamic coupling allows for the assessment of higher-order dynamics. Here, the continuous exchange of coupled variables between the two well-calibrated and validated models allows for direct interpretation of model scenarios. Using the above example, after upstream changes of farming practices, what are system dynamics once farmers adapt to hydrological changes?

3 STEP-WISE MODEL CALIBRATION AND VALIDATION

With an integrated and technically verified coupling setup, and finalized data collection and data integration at appropriate scales, the dependencies of coupled variables were analyzed. Using indicative results, the step-wise and integrated calibration and validation procedure is outlined.

3.1 Data analysis for variations in hydrology ΔH_0

The following analysis integrates river flow time series, data on water rights registries, data on agricultural land use by irrigation sector, and technical coefficients on land use activities, as compiled in an interdisciplinary data base (Fig. 1).
As input to SE_1, time series of river flow stations are analyzed for average and dry years (4-, 6.7-, and 20-year frequency, in which water flows exceed 75%, %85 or %95 of the time respectively). Monthly water availability fluctuates strongly over years, especially during the month of January (Fig. 3). During Jan-Feb of "normal" droughts, e.g. during El Nino years, access to the Melado canal, fed from a neighboring catchment, buffers shortages. During extreme droughts, this dependence starts in Nov-Dec.
As an indicator for physical water scarcity per sector, the percentage of available water is calculated by multiplying agricultural land use data [ha/sector] at irrigation sector level with estimations of water requirements for those land uses [$\frac{\text{liter}}{\text{ha}}$]. Results are divided by water availability, calcu-

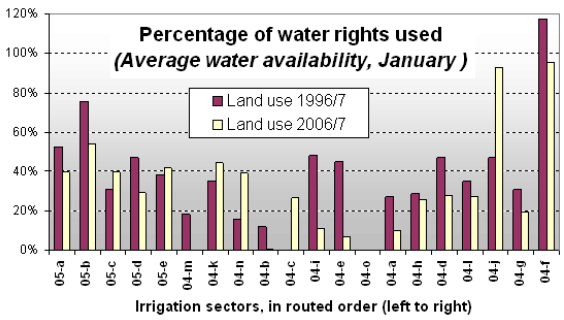


Figure 5: Usage of available water assets from farmers, during average January.

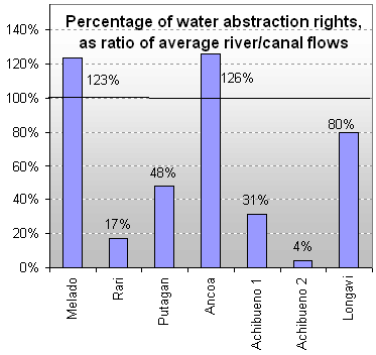


Figure 6: Ratio of river flows to equivalence value of added water rights.

lated with water rights data for each sector [$\frac{\text{units}}{\text{sector}}$] times the value of these rights [$\frac{\text{liter}}{\text{unit}}$]. Results show full usage only in the downstream sector 04-f, which also receives water not used upstream. During average years, all other sectors under-exploit water resources. Possible explanations include over-supply, risk-aversion to droughts, and scarcity due to misallocation of water rights among farmers. The estimated change of water requirements in sectors 04-l and 05-e from 1997 to 2007 reflects the reduction of rice production, in line with falling prices in 2000 (Fig. 5).
For MP-MAS , water rights are expressed as the percentage of river inflows. The irrigation sectors that suffer from the strongest annual fluctuations of inflows are the large rice-producing sector 04-l, which satisfies 20% of January water demand from the Melado canal, sector 04-f ($< 2\%$ Melado water) and the large 04-g ($\approx 45\%$ Melado water), while access to anti-cyclic water supplies (Melado, and to some extend the Longavi River) buffers drought conditions.

3.2 Sensitivity of MP-MAS to variations in hydrology - $\Delta SE_1(\Delta H_0)$

The economic model MP-MAS was used to assess the impact of water variability on individual farm households. Based on empirical data, we modeled farm decisions using Mathematical Programming, and validated results interactively with 10 selected farmers. A statistically consistent population of farmers was then created, and Schilling [2007] estimated shadow prices for water for different years. Shadow prices for water (SPW) range from \$P/m^3 $0.0 - 60.0$ and a maximum of \$P/m^3 120.0. During droughts, the percentage of farmers with low or very low SPW falls from 55% during normal years to 15%. These results clearly highlight that not only water availability, but also access to labor, to capital and markets determine water usage by farmers.

3.3 Data analysis for variations in economic impacts ΔSE_0

To model irrigation at basin scale, while maintaining a field-level perspective, cyclic water reuse needs to be captured properly. It is invisible at macro scales, and data scarcity requires a parametrization approach for return flows. Using the WASIM-ETH specification for 1996/7 land use [Uribe et al., 2008], the net (non-cyclic) abstraction of irrigation water was estimated at $\approx 38.2 \frac{m^3}{sec}$, and a (net) irrigation table of $IRR_{\text{eff}} = 3.76 \frac{\text{mm}}{\text{day,ha}}$.
The routing of irrigation water between sub-basins through canals was initially estimated from water rights. Yet, analysis of land use water requirement suggests that only during the peak month of February is the full routing realized, in other months only shares are utilized (Fig. 4). Further, for January, a comparison of river/canal flows with the value of water rights flows reveals that abstraction rights exceed flow for both the (water-scarce) Ancoa river and the Melado canal (Fig. 6). The additional 23% are explained as return flows, from irrigated fields back into main

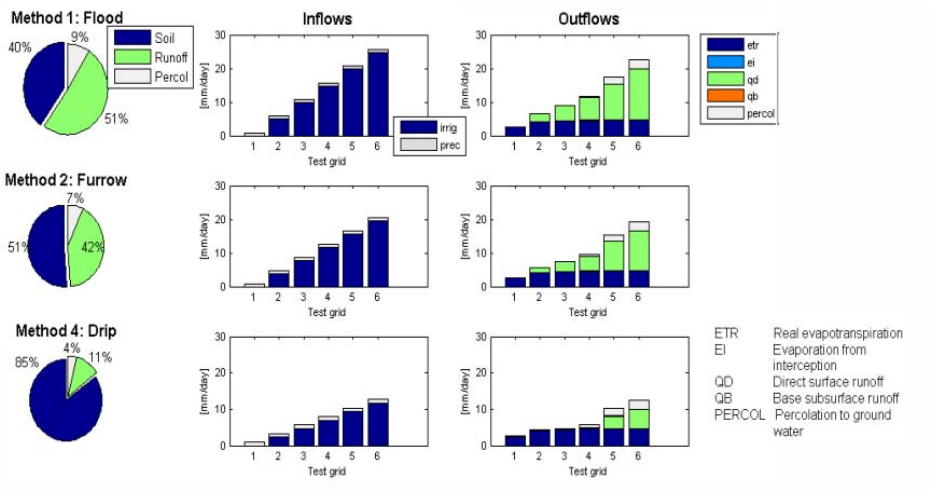


Figure 7: Sensitivity to flow component at grid cell level to irrigation methods, and irrigation water table applied (evaluated at grid cell level).

canals/rivers, which conforms closely to the 22% estimated by MOP [1992].
Using EDIC equations and land use data, field-level data for irrigation methods and their efficiency η [$\frac{\text{mm ETP}}{\text{mm applied}}$] was aggregated to *effective* efficiencies (without direct reuse, but with return flows) to bridge field-scale and basin-scale.

3.4 Sensitivity of WASIM-ETH to variations in irrigation - $\Delta H_1(\Delta SE_0)$

The initial calibration gave a Pearson correlation of $R^2 > 0.9$ for outflows of large sub basins and $R^2 > 0.55$ for those of small sub basins [Uribe et al., 2008], before incorporating the above data. Grid-scale analysis revealed that even with acceptable reproduction of flow station measurements, relevant internal model processes were not properly represented. The heterogeneity of soil water dynamics and data scarcity for ground water dynamics, limits the physical scope of this study (see also [Hansen et al., 2007]). The soil module of WASIM is currently recalibrated at micro-scale, using the extended WASIM-ETH irrigation module with effective irrigation efficiencies, applying water directly to top-soil and using parameterized return flows and percolation. Figure 7 shows model experiments where otherwise homogeneous cells receive different irrigation tables, according to three different methods.

4 DISCUSSION OF PRELIMINARY RESULTS

In accordance with local knowledge, the analysis shows that hydrological constraints are economically relevant only during the summer months (Jan - March). Economic analysis shows that a basin perspective on absolute water scarcity does not suffice to explain why high-value irrigation agriculture (fruits & vegetables) is not expanding further. In average years, the total amount of water utilized is below 65%. Nevertheless, high shadow prices, especially during drought conditions, show that a significant portion of the farm population is constrained by water. The hydrological model requires a consistent representation of surface evaporation, plant water uptake from root zones, interflows and surface runoff at grid scale. Furthermore, the conceptual treatment of return flows, such that irrigation dynamics at the basin scale are correctly represented and properly understand, is also necessary.
Ongoing analysis focuses on consistency checks of data with local experts and on calibration of both the hydrological and the economic model, first at grid (and agent) scale, with grid (and agent) scale parameters, then at meso-scale, with sector-level parameters.

5 CONCLUSION

A calibration and validation framework for integrated, empirical and (semi-)predictive modeling was demonstrated using a case study. The integration of theories and concepts, of models and model source code, and ultimately of data, was shown to be dependent on and mutually beneficial for each other. Reasoning that builds on integrated concepts and data at respective scales reveals inconsistencies in otherwise behavioral model specifications, which are thus inadequate for in-depth or predictive analysis. For irrigation management at the basin scale, we indicate that physical scarcity and allocative scarcity coexist in one basin. In a heterogeneous population of farmers, constraints vary considerably. Therefore, effective policies require precise definition and in-depth analysis of the target groups. As a general note, this paper emphasizes that integrated modeling at the basin scale needs to be seen as a co-evolution of concept development beyond disciplinary boarders and theoretical, computational and data management aspects. This supports the vision of Beven [2007] for model tool boxes to be learning tools, to relate data, to identify knowledge gaps, and to value well-targeted monitoring for new information.

ACKNOWLEDGEMENTS

We thank the CGIAR Challenge Program on Water & Food and Alechile (DAAD/CONICYT) for financial support.

REFERENCES

Argent, R. M. An overview of model integration for environmental applications- components, frameworks and semantics. *Environmental Modelling & Software*, 19(3):219–234, 3 2004.

Arnold, T., H. Uribe, and T. Berger. A distributed hydrological model of irrigation and agricultural land use at the basin scale: Application to the viith region chile, April 2008. Accepted for Conference "21st Century Watershed Technology (ASABE)".

Berger, T. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics*, 25(3):245–260, 2001.

Berger, T., R. Birner, J. Díaz, N. McCarthy, and H. Wittmer. Capturing the complexity of water uses and water users within a multi-agent framework. *Water Resources Management*, 21:, 129–148, 2007.

Beven, K. A manifesto for the equifinality hypothesis. *Journal of Hydrology*, 320:18–36, 2001.

Beven, K. Towards integrated environmental models of everywhere: uncertainty, data and modelling as learning process. *Hydrological & Earth System Sciences*, 11(1):460–467, 2007.

Hansen, J., J. C. Refsgaard, and V. Ernstsen. Problems with heterogeneity in physically based agricultural catchment models. *Journal of Hydrology*, 342(1-2):1–16, 2007.

MOP. Estudio de factibilidad del canal linares. Report, volumen 1 y anexo, Consorcio EDIC-CEDEC for Ministerio de Obras Públicas, Dirrección de Riego, Chile, 1992.

Parker, D., S. Manson, M. Janssen, M. Hoffman, , and P. Deadman. Multi-agents systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93(2):314–337, 2003.

Schilling, C. Interactive modeling of farm-level response to the ancoa dam project, chile. Master thesis, University of Hohenheim, Stuttgart, Germany, 2007.

Schulla, J. and K. Jasper. Model description wasim-eth, new version. Internal report, IAC, ETH Zürich, Switzerland, 2007.

Uribe, H., T. Arnold, J. Arumí, T. Berger, and D. Rivera. Aplicacion del modelo wasim-eth para el estudio de la hidrologia en un sistema de riego, 2008. submitted.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Development of an integrated framework for assessing the impact of urban planning on water quality

G. Fu[a], S.-T. Khu[a], F. di Pierro[a], A. Doglioni[b] and F. Primativo[b]

[a] *Centre for Water Systems, School of Engineering, Computing and Mathematics, University of Exeter, North Park Road, Harrison Building, Exeter EX4 4QF, UK (g.fu/s.t.khu/f.di-pierro@exeter.ac.uk)*
[b] *Technical University of Bari, via del Turismo no. 8, Paolo VI, 74100, Taranto, Italy (A.doglioni@poliba.it; francescaprimativo@tele2.it)*

Abstract: Recently, environmental concerns have gained increasing attention in the urban planning process. This paper aims to address the effects of urban planning on the surrounding water quality by developing an integrated modelling framework. This framework provides a conceptual and quantitative tool to quantify the impacts of urbanisation patterns on water systems, including urban drainage, wastewater treatment plant, and receiving rivers. The development of the framework (known as SSDIM), including individual components and their interactions, is described in this paper, and its functionality is demonstrated by a semi-hypothetical case study. The results show the effects of different planning strategies on receiving water quality, and also show that the framework is a flexible and useful tool for town planners and water specialists, to explore sustainable paths in the urban planning process.

Keywords: Impact assessment; Integrated modelling; Sewer system; Urban planning; Water quality

1. INTRODUCTION

Urban development has grown significantly both in the developing and developed countries in recent years in response to demographic changes and economic growth. For example, the UK government [2007] has recently pledged to build 3 million new homes by 2020, many of which will be in the already densely populated south-east. The development of housing and its associated infrastructure is one of the main driving factors for changes in land cover to urban land uses, which have long been recognised as a principal contributing factor to many environmental problems, including water pollution, air pollution, aquatic environmental deterioration, flooding and many others. Particularly, development of a new housing area can severely affect water flow and water quality of urban water systems in several aspects [Butler and Davies, 2004], including the increased pollutant loadings from intensified human activities and the runoff generated from impervious areas during rain events. These will impose increasing pressure on existing wastewater infrastructure, and increase the likelihood and risk of combined sewer overflows (CSO) in a combined sewer system.

These environmental impacts from urban land use changes have gained increasing attention in the urban planning process, along with many other concerns in the regional, political, social-economic and cultural aspects. Urban planning is a complex decision making process, and its overall effects are not always obvious, and might only be revealed with the aid of specific simulation models. Integrated assessment and modelling is a promising tool for providing good guidance in such complex realms [Jakeman and Letcher, 2003]. This paper describes the development of an integrated framework that provides a conceptual and quantitative tool to investigate qualitatively and, ultimately, quantify the short- and long-

term impacts of urban planning on ambient water quality. The functionality of this framework is demonstrated by a semi-hypothetical case study, in which the impacts of three planning scenarios are analysed and compared in terms of the river dissolved oxygen (DO), biochemical oxygen demand (BOD), ammonium concentrations and river flow.

2. DEVELOPMENT OF THE FRAMEWORK

The framework acts as an interactive platform between the policy level and the natural environment, herein an urban water system. It provides a simulation and decision support tool to assess both short- and long- term potential impacts of urban planning on the surrounding water systems, to suggest a broad class of preliminary options to scope, and to identify the best options given physical, economic and social management objectives.

The framework consists of four key subsystems, including land use, urban drainage/sewer, wastewater treatment plant, and river systems. It was implemented in the SIMULINK® environment, which allows a system dynamics modelling of the different processes in various components and the interactions between them. Figure 1 shows the information-flow diagram of the framework, and each of the components is described below.

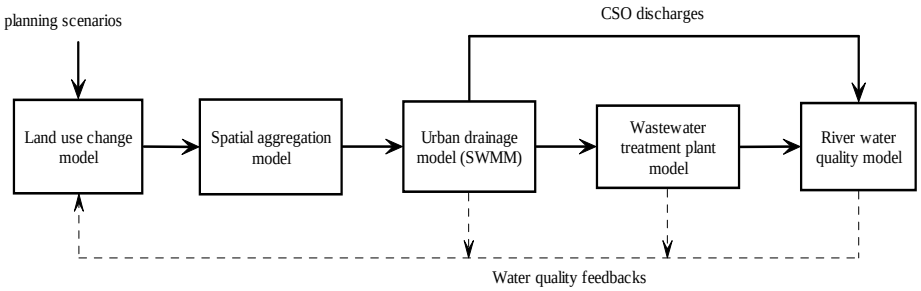


Figure 1. The information flow diagram of the integrated framework (SSDIM).

2.1 Land use change model

The land use change model represents the land planning process and is used to simulate the urban growth over a given planning horizon. This model is a cellular automata-based model, in which the development state of each cell is determined in terms of its own state and the states of its nearby neighbours at the previous time step, according to some transition rule. A few types of transition rules have been implemented, and new rules can be easily integrated to allow for the analysis of the impact of different expansion patterns on the water quality.

Table 1 shows the parameters used in the land use model. Driven by population growth, the model distributes the population spatially over a predefined development area according to the transition rules chosen and provides the land use type, number of houses, and number of people for each residential land use cell. The cell-based information is transformed and linked to the node-based sewer network in the spatial aggregation model according to the geographic information of the catchment and the topology of the newly designed/expanded network. The new network branch was defined a priori in this paper, and an automatic procedure for expanding automatically the network in the implementing process will be investigated at a later stage. For the sake of simplicity, a uniform square grid was used to represent land parcels.

Table 1. The land use model parameters.

Parameters	Description
Number of grids	The total number of grids for the new housing area
Grid resolution	The size of the grids used
The initial land use type	The land use types of each cell at the initial time
The existing houses	The number of houses of each cell at the initial time
Land use class	The land classes of each cell specifying the priority for development
The maximum housing density	The maximum number of houses per ha specified for each cell
The development pattern	The transition rules specified for the new development
Housing parameters	The house type, occupancy, and landtaken

2.2 Urban drainage and river models

These two components are used to simulate the various water quantity and quality processes in the river and sewer system, which expands to the new urbanised areas for wastewater discharges. The EPA Storm Water Management Model (SWMM) [Huber and Dickinson, 1998] was chosen because of its robustness and popularity.

Complex biochemical processes can occur while pollutants are transported in the sewer networks, which have a significant effect on the urban water system. A water quality model [Lijklema et al., 1996] was chosen here to model the river system, which mainly focuses on the oxygen balance in the river. The organic matter was modelled in two fractions, i.e., slowly and readily biodegradable. This has an advantage in modelling the effect of CSO discharges and treatment plant effluents that might contain very different organic matter in terms of biodegradability [Schütze, 1998]. This model also considered DO and ammonium.

2.3 Wastewater treatment plant component

This component is used to simulate the biological processes in the municipal activated sludge wastewater treatment plant, and is based on the Benchmark Simulation Model No. 1 [Copp, 2002]. This benchmark model represents a denitrification plant, consisting of a five-compartment reactor with an anoxic zone and a secondary settler. A basic control strategy is used to control the dissolved oxygen and nitrate levels. The Activated Sludge Model No. 1 (ASM1) is used to simulate the biological processes in the reactors [Henze et. al. 1986] and the secondary settler is modelled as a non-reactive 10-layer process using a double exponential settling velocity model developed by Takács et al. [1991]. As shown in Figure 2, there are two internal recycles: Nitrate internal recycle from the 5th tank to the first tank and return sludge recycle from the underflow of the secondary settler to the first tank. In addition, waste sludge is pumped away continuously from the settler.

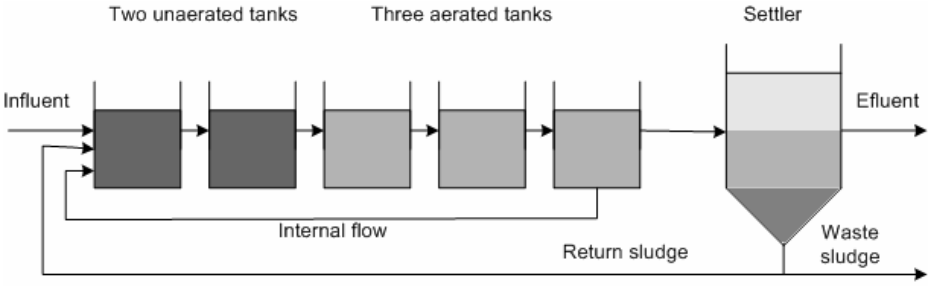


Figure 2. Schematic representation of the benchmark plant.

2.4 Model interactions

In this framework, the components of the framework were run at different time steps to allow for an accurate enough description of the processes involved. In general, the time step of land use change model was set to one year; the urban drainage and treatment plant models are simulated at a much smaller step (for example, 5 or 30 mins). The underlying assumption for the different time steps is that the urban wastewater system can be simulated for a rain events based period, within each annual time step for land use change model. This enables the effective representation of system dynamics arising from rain events. One of the important characteristics in a combined sewer system is the CSO discharges, which is generated when the inflow exceeds the maximum treatment plant influent. These discharges are directly discharged into receiving rivers, and could have a significant impact on river water quality.

System interactions also exist between the pollutants considered in the sewer system, treatment plant and river, and these are usually modelled using different pollutant variables. In the urban drainage system, the pollutants considered include suspended solids, volatile

suspended solids, chemical oxygen demand (COD), soluble COD, ammonium and nitrate: this enables simulating the impact of land use changes on the runoff water quality. In the ASM1 model, there are 8 processes incorporating 13 different components, which are classified into soluble components and particulate components. In the river, the pollutants considered are readily and slowly biodegradable BOD, ammonium and DO. A factor-based conversion method is used here to convert different pollutants [Schütze, 1998].

Outcomes from urban drainage, wastewater treatment plant and river system will be used to provide urban planners with feedbacks with the view of identifying alternative and more sustainable development patterns. These feedbacks may include the occurrence of sewer flooding, treatment plant effluent compliance, and river water quality exceeding a predetermined threshold.

3. SET UP OF A CASE STUDY

The functionality of this framework is demonstrated by a semi-hypothetical case study, consisting of a real combined sewer system in UK, a hypothetical treatment plant and river system. The network has 346 manholes, 7 Weirs and 2 outfalls. The treatment plant influent is controlled by a weir, and the flow exceeding the maximum influent is directly discharged into the receiving river. The hypothetical treatment plant was adjusted according to the capacity of the sewer system. The river considered is about 40 *km* in length, and it was equally divided into 40 reaches for simulation. A constant upstream inflow with constant pollutant concentrations was assumed for this river in order to clearly illustrate the impact of a rainfall event with a 10-year return period. The CSO overflows are discharged into the river at Reach 2 and treatment plant effluent at Reach 8.

An area of 100 *ha* was identified for possible future residential development, and a grid cell size of 100 *m* × 100 *m* was used in the land use model for population distribution. Three planning **scenarios** assessed in this paper were defined below:
Scenario I (normal increase): the population increase by 10,000 in 20 years, and the maximum housing density is set to 40 households per *ha*.
Scenario II (low density increase): the population increase by 20,000 in 20 years, and the maximum housing density is set to 40 households per *ha*.
Scenario III (high density increase): the population increase by 20,000 in 20 years, and the maximum housing density is set to 60 households per *ha*.

In the three scenarios, the population increase was assumed to be equally distributed over each year, for instance, in the base case scenario, 10,000/20 = 500 persons are distributed to the new development each year. These urban planning scenarios were evaluated in terms of a set of water quality indicators including the DO, BOD_5 and ammonium concentrations at Reach 11, i.e., 3 *km* downstream of the treatment plant effluent and 9 *km* downstream of the CSO discharges in the receiving river. Reach 11 was chosen because it is the most critical location in terms of the chosen water quality indicators.

4. RESULTS AND DISCUSSION

4.1 The impact of Scenario I

Figure 3 shows the effects of the base case scenario on river water quality for three selected years, i.e., Years 1, 10 and 20. The river water quality in terms of DO, BOD_5, and ammonium has deteriorated continuously with the population increase in the new development, while the river water flow increases steadily over time. The increased river flow results from the new development, in which more runoff is generated from the rain event because of the increase in impervious area and more domestic wastewater is discharged because of the population increase. Most of the increased volume of water is discharged into the river through CSO overflows because the maximum treatment plant influent has been reached for the simulated rain event, even for the existing catchments, and this results in the deterioration in the receiving water quality.

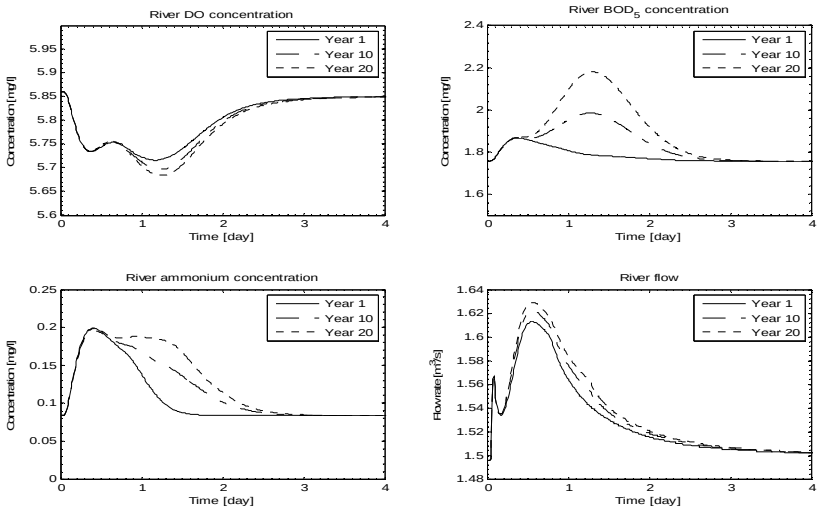


Figure 3. The impacts on river water quality for years 1, 10 and 20.

4.2 Comparison of the scenarios

Figure 4 shows the comparative results of the three scenarios at the end of the planning period, i.e., Year 20. Compared with the base case scenario, it is clearly shown that the river water quality of the low density scenario deteriorates when the population size is doubled and housing density is kept unchanged. However, for all the water quality indicators, the impact of the high density scenario is roughly the same as the base case scenario.

Both the population size and housing density for the new development have a clear impact on the receiving water quality. However, increasing the housing density has a positive impact on river water quality. This is probably due to the fact that less land is transformed into urban land uses at a higher housing density so that less runoff are generated and discharged into the urban wastewater system. Thus, the impacts of population increase can be reduced to some extent as illustrated by the high density scenario. This may be suggested as an effective mitigation option for land use planners to reduce the impact of new developments and thus achieve a better water environment.

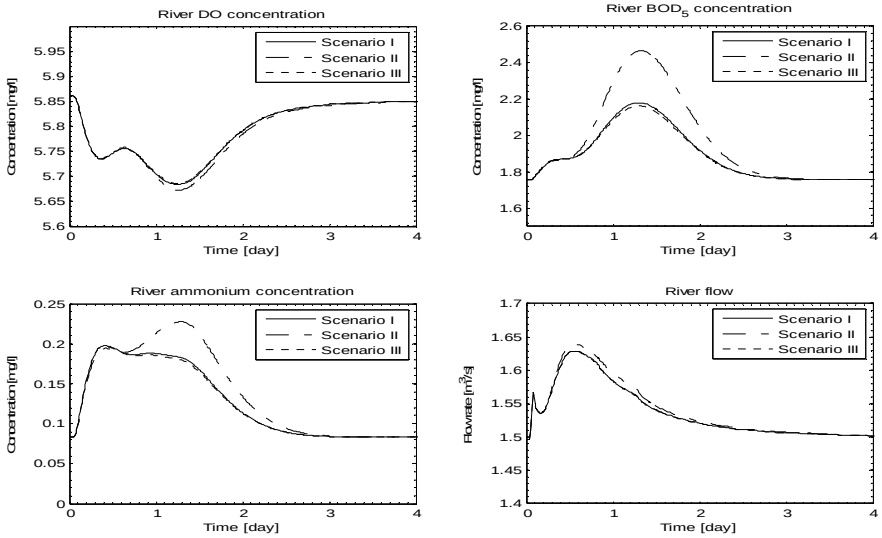


Figure 4. The impacts of different urban planning scenarios on river water quality.

5. CONCLUSIONS

Urban land use planning is a crucial task that has large impact on our lives. It drives economic development, affects social structures, shapes the urban landscape and modifies the environment. The impact of urban planning on the surrounding environment is not always obvious and difficult to assess, particularly for the long term impact, and it usually needs specific computer simulation models to support in the process.

In this paper, an integrated modelling framework was developed to assess the effects of alternative urbanisation patterns on the surrounding water quality. This framework provides a conceptual and quantitative tool to quantify the impacts of urban land use changes on urban water systems, including urban drainage, wastewater treatment plant, and receiving rivers. To demonstrate the functionality of this framework, a semi-hypothetical case study was set up and three planning scenarios were evaluated for a new housing area. The comparative results show that the population growth rate has obvious effects on receiving water quality. However, these effects can be reduced to some extent when the housing density is raised to a higher level. This may be suggested as an effective mitigation option for land use planners to reduce the impact of new developments and thus achieve a better water environment.

ACKNOWLEDGEMENTS

The study is partly funded by UK EPSRC grant no GR/S86846/01 and is also part of the Integrative Systems and the Boundary Problem (ISBP) project (www.tigress.ac/isbp), supported by the European Union's Framework 6 Programme New and Emerging Science and Technology Pathfinder initiative.

REFERENCES

Butler, D., and Davies, J.W., Urban Drainage, 2nd edition, Spon Press, London, 2004.

Copp, J.B. (ed.), The COST simulation benchmark- description and simulator manual. ISBN 92-894-1658-0, office for official publications of the European communities, Luxumbourg, 2002.

Henze, M. Grady, C.P.L., Gujer, W. Marais, G.v.R., Matsuo, T., Activated sludge model No.1. LAWQ Scientific and Technical report 1, IAWQ , London, UK, 1986.

Huber, W.C., Dickinson, R.E., Storm Water Management Model. User's Manual Version IV, US Environmental Protection Agency, 1988.

Jakeman, A.J., Letcher, R.A., Integrated assessment and modelling: features, principles and examples for catchment management. *Environmental Modelling & Software*, 18, 491-501, 2003.

Lijklema, L., AAlderink, R.H., de Ruiter, H., Procesbeschrijvingen DUFLOW zuurstofhuishouding in stromende en stagnante watersystemen. Landbouwuniversiteit Wageningen, 1996

Schütze, M., Integrated Simulation and Optimum Control of the Urban Wastewater System. PhD thesis. Department of Civil Engineering; Imperial College of Science, Technology and Medicine, University of London, London, 1998.

Takács, I., Patry, G.G., Nolasco, D., A dynamic model of the clarification thickening process. *Water Research*, 25(10), 1263-1271, 1991.

The UK government, The Government's plans for reform, 2007. Http://www.number-10.gov.uk/output/Page12422.asp.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Representing river system behaviour in UML to enhance development of DSS

J. A. E. B. Janssen[a] **and Ioannis M. Dokas**[b]

[a] *Water Engineering and Management Group, University of Twente Enschede, The Netherlands, j.a.e.b.janssen@utwente.nl*

[b] *Cork Constraint Computation Centre, Department of Computer Science, University College Cork, Ireland, i.dokas@4c.ucc.ie, jdokas@yahoo.gr*

Abstract: Water management decisions often have far reaching consequences for the every day life of the inhabitants of a river basin. To take these consequences into account during decision making processes, we aim to develop a software tool that can indicate the expected effects of different management strategies on a number of river functions. This requires knowledge domain representation that fits the perception of stakeholders, and at the same time provides a starting point for modelling. We found that UML provides a language that on the one hand facilitates the communication and is intuitively understood, and on the other hand provides a high level of consistency as well as decomposition into elements that can easily be modelled. Moreover, UML provides the space required to deal with both quantitative and qualitative information, which is desirable to incorporate knowledge of different scientific fields / disciplines.

Keywords: River basin management, Maas river, UML, DSS.

1. INTRODUCTION

The development of software models for environmental management is to a large extent an interdisciplinary activity. During the development of the model, many persons with different backgrounds are involved. These include for instance scientific experts in different fields, policy makers at different administrative levels, inhabitants, software developers, consultants, and so on. Therefore the successful development of these models depends to a large extent on communication [De Kok et al., 2006]. The general approach is the use of some kind of common modelling language, such as a verbal or graphical description of model components. This language is then used to construct a model of reality, which is understood or not by the receiver, depending on whether he or she is familiar with the language, and depending on the clarity and specificity of the syntax (rules, 'grammar') and semantics (meanings) used in the language. Different languages may be required during the different stages in the design process of tools for environmental management.
Here we focus on the development of a conceptual model for the exploration of strategic river management in the Dutch Maas River. The research is in a phase where the relevant actors are known, the indicators have been selected, and management measures have been proposed. Now the conceptual model ought to provide a framework for the software implementation, i.e. the translation to programming language. In this stage we focus on the communication between the software designer and domain experts. While having in mind that the systems need to be implemented in a programming language, it is desirable to

provide an early link with the world of software developers and programmers. This means that a number of technical and social issues need to be addressed, among which adoption of standard specification, notation and documentation on models [Argent, 2001]. For this specification, one can distinguish between a) informal, b) semi-formal or c) formal specification. These specifications vary in the level of precision in notations and semantics. In general, the more formal the method is, the more difficult it will be to apply it. In this paper we propose to use the Unified Modelling Language (UML). UML is a widely used standard language for the specification, visualization, construction, and documentation of the components of software systems. We aim to explore the potential use of UML for the development of a DSS for river management, involving particular issues such as the involvement of different decision-criteria and different stakeholders.
We hypothesize that UML can help overcoming a number of problems that are generally encountered during the domain analysis, like issues in communication and ad hoc decisions during the design. The problems have been formulated as requirements, which the UML needs to satisfy to be of added value. We tested the extent to which UML satisfies the requirements by an empirical investigation concerning the Dutch Maas river.

2. STATE OF THE ART

The development of software tools to support river management begins with analysis of the system at hand. Often, graphical representations are used, such as system diagrams [Matthies et al., 2007]. The most common approach is to strongly link the graphical user interface to the code representing the environmental model [Papajorgji et al., 2004b]. Recently several model developers have started using UML. Barthel et al. (2005), for instance, have successfully used model conceptualizations in UML for the development of DSS. UML has also been applied to water to soil water-balances and irrigation- scheduling modeling [Papajorgji, 2004a]. The object-oriented approach is also used to enable the assembly of simulation models from previously and independently developed component models [Hillyer et al., 2003]. Pahl-Wostl et al. (2008) applied UML in environmental modeling and found the following advantages:

- The enforcement to be precise and concrete forces domain experts to clarify their own concepts and models and to make choices in otherwise fuzzy approaches.
- The incorporation of concepts as part of and interactive implementation process requires mutual explanations of concepts as part of a new project communication.
- Complex system structures can be communicated unambiguously.
- It can facilitate the comparison of competing concepts.

The most undisputable argument for using UML however is the lack of an alternative method that can cover such a broad and easy to understand variety of views [Pahl-Wostl et al., 2008]. In addition, other languages, like the Modular Modelling Language (MML) [Maxwell et al., 1997], which have been used in developing DSS do not have the level of international standardization support such as UML. The use of UML for the development of DSS seems promising. However, a structured approach of the evaluation of domain-specific requirements has not yet taken place. This work aims to provide such an evaluation, and add an assessment of the suitability of UML for this type of domain to the existing knowledge.

3. MATERIALS AND METHODS

In this work we have used knowledge from the water management domain in order to formulate the requirements of a DSS. UML was used as modelling method and it was evaluated empirically in a case study.

3.1 UML

On the surface, UML is a visual language for capturing software designs and patterns. It can be applied to a large number of different areas and can capture and communicate everything from company organization to business processes to distributed enterprise software. It is intended to be a common way of capturing and expressing relationships, behaviours, and high-level ideas in a notation that's easy to learn and efficient to write. UML is visual; just about everything in it has a graphical representation [Pilone & Pitman, 2005].
First and foremost, it is important to understand that UML is a language. This means it has both syntax and semantics [Pilone & Pitman, 2005]. UML provides a way to capture and discuss requirements at the requirements level, sometimes a novel concept for developers. There are diagrams to capture what parts of the software realize certain requirements. Also, there are diagrams to capture exactly how those parts of the system realize their requirements. Finally there are diagrams to show how everything fits together and executes [Miles et al., 2006].

Obviously the focus of UML is modelling. It is a means to capture ideas, relationships, decisions, and requirements in a well-defined notation that can be applied to many different domains. Modelling not only means different things to different people, but also it can use different pieces of UML depending on what you are trying to convey. The use case view captures the functionality required by the end users. The concept of end users is deliberately broad in the use case view and includes the primary stakeholders, the system administrator, the testers, and potentially the developers themselves. The use case view is often broken down into collaborations that link a use case with one or more of the four basic views. The use case view includes use case diagrams and typically uses several interaction diagrams to show use case details.

3. 2 Description of the reference system

Various management measures (river works) can be applied to optimize the functions of rivers. In the Netherlands, of these functions, the conveyance of water while maintaining a safe catchment is considered the most important one. The safety is expressed by an exceedance probability, indicating the expected occurrence interval of extreme discharges. In general for the Netherlands' rivers the interval is 1250 years. The water levels corresponding to peak discharges with a frequency of 1:1250 years are not supposed to exceed an indicated maximum level corresponding to the dike height minus a safety board. In order to maintain these water levels, even when the extreme discharges increase due to climate change, river engineering measures are taken. These engineering measures however, may have side effects on other functions during normal discharges. It is desirable that the negative side-effects on other functions, such as nature, agriculture, shipping and habitation, remain limited. For simplification purposes, we only include three user functions here; landscape impact, agriculture suitability and safety. Also, only three engineering measures are taken into account (Figure 1).
The impact on safety can be assessed by calculating the probability that levees are overtopped. This only occurs during extremely high discharges for which a maximum probability is legislatively determined. Landscape impact is relevant to the inhabitants of the area. It is determined by a combination of measure and area features, such as the changes in width, location in the river

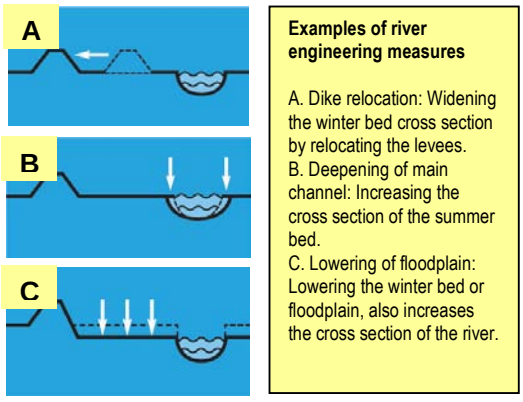


Figure 1. River engineering measures [www.ruimtevoorderivier.nl, 25-04-07]

bed, and the original river cross-section dimensions. The impact on agriculture is determined by calculating the annual effects of drought and effects of floodplain inundation (which is where the agriculture is assumed). Table 1 shows an overview of the relevant indicators in this system.

Table 1. Overview of indicators

	Time scale	**Spatial scale**	**Nature**
Safety	Probability [-]	River [~km's]	Crisp
Landscape impact	Impact [-]	Measure [~m's]	Fuzzy inputs and relations
Agriculture	Process[yr]	Location [km n^{o}.]	Fuzzy inputs, crisp relations

3.3 Evaluation process

The domain of this research is strategic river management. This refers to the early phases of river system planning, in which a trade-off has to be made between different objectives of the river planning process. From this, a number of requirements to the conceptual model can be derived. In fact the evaluation process has started with the identification and selection of evaluation criteria. The aim was to evaluate UML based to these criteria. The second step of the evaluation process was to design the river model. In this step the Maas river was used as case study. The third step was to discuss and evaluate the use of UML with domain and software experts to evaluate UML based on the criteria defined in the first step.

3.3.1 Evaluation Criteria

The identified evaluation criteria are listed below.

- *Different temporal and spatial scales.* A variety of functions is taken into account, and depending on their accompanying evaluation criteria the system description contains many different temporal and spatial scales.
- Because in our case, stakeholders or end-users determine what needs to be included the descriptions of the relations vary from *qualitative to quantitative.*
- *Different actors require different types of information.*
- Implementation of the design in a software tool requires a *consistent* design.
- *Flexibility with respect to model extension.* The stakeholders' interests represented in the model may change over time. We want to be able to anticipate such changes.
- For iterative development of the model, *the communication with people from different backgrounds needs to be supported.*

3.3.2 River Model Development

The model was developed based on a readily existing system diagram. The problem of this diagram however was, that it is not sufficiently specific to provide a sound basis for programming. The system was redesigned by the authors, from a river engineering and software engineering background, respectively. Part of the design is based on literature research, and on study of the case documentation. To further specify the system, additional knowledge was collected during meetings with river experts, both with a practical and research background. By bringing together the knowledge, a number of diagrams applicable to the river domain were developed.

3.3.3 Expert Evaluation

After designing the model in UML, the eventual product and the process underlying it were again discussed. During a workshop with model designers in environmental modelling a number of pros and cons of using UML came to light. The actual designs were again discussed during face to face meetings with people with a background in river research, both from their practical experience as well as from research experience. The focus of these discussions was both on the use of the diagrams during the design of the system, as well as on the later 'usability' of the diagrams. The expert opinions were categorised corresponding to the requirements that were outlined earlier.

4. EMPERICAL STUDY RESULTS

Our research aims to develop a conceptual model that outlines how a software tool can support various information needs in the trade-off between different river management measures. A number of UML diagrams will be used. Some indicative examples are shown below.

4.1 Use cases

In this example, there are three main groups of actors who will use the information provided by the tool (Figure 2). In the first place, there is a river manager. He has a set of measures at his disposal. He may want to implement different measures at different locations, and with varying parameters. His initial interest in proposing measures, in this case, is an improvement of the flood safety. He wants to know how safety changes, at different locations along the river, as a result of the measures. Second, inhabitants are interested to know what the impact of the river works (measures) on the landscape will be. They prefer measures that fit in the landscape to measures that severely disrupt it. Third, farmers want to know how the measures affect the ground water levels in the area, and the frequency with which floodplains are inundated. Also the duration and depth of this inundation, as well as the season during which it takes place, is important in the evaluation of preferences.
This diagram clearly illustrates which information is required by which actor, and it shows at a very abstract level how the information coheres with the river system. Consistency is approached from two angles here; first, all the relevant actors need to get a place in the diagram, and second all information needs require an actor perceiving it. In the initial stage of the DSS design this is a very helpful diagram to draw, because it shows which information needs be derived from the eventual model. By keeping the diagram very abstract in this stage, one prevents looking into the relations too much. Too much detail in the relations may shift the focus of the design to the availability of knowledge, models and data, something not yet important in this stage.

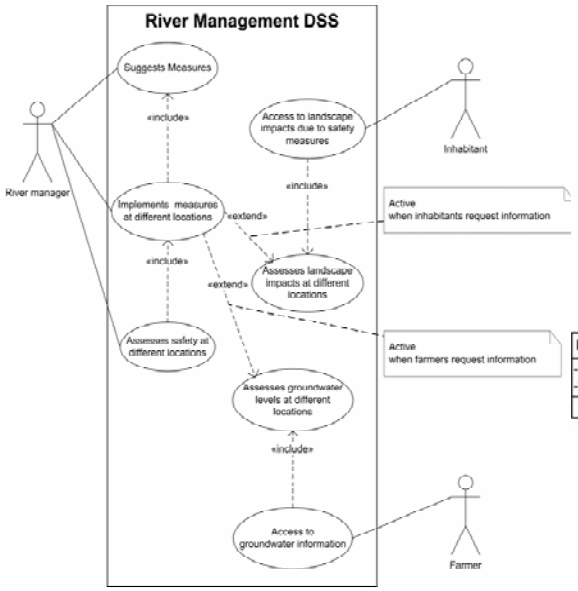


Figure 2. Use case diagram

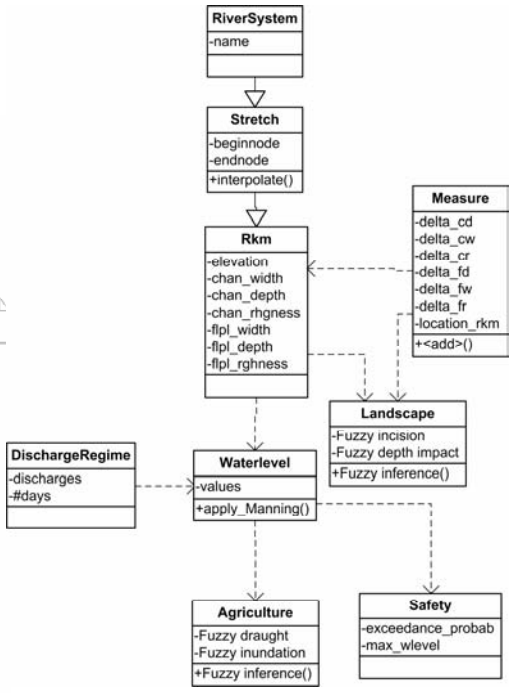


Figure 3: Static structure of the river system

4. 2 Static structure

The static structure, or class diagram, depicts the time-independent relations between the objects in the model. Classes are collections of objects. In our case, the model applies to a certain river system. Although it is common to omit attributes and relations in a conceptual model, we have added some of them here to clarify the model structure (Figure 3; previous page). Including attributes allows to more easily check the consistency of the conceptual model; if attributes are not included it is easier to define overlapping or inaccurately defined classes.
The river system consists of different, geologically distinct, stretches. Stretches have a begin- and end-node. Within each stretch a number of river kilometres (rkm) is located. For each rkm a cross-section is described along with roughness data. The measure class contains the different measures. Attributes hereof are the changes (delta) in cross section (channel depth and width, floodplain depth and width) and roughness data (channel roughness, floodplain roughness). The measures perform operations on the cross sections, for the rkm's at which the measure is located. Measure and location dimensions are the input for a fuzzy inference [Zadeh, 1973] on landscape impact. The new cross section, combined with the discharge regime of choice, determine the new water level. The water level, next, is the input for safety and agriculture indicators, one of which is evaluated using probability theory, and the other using another fuzzy inference.
The static structure diagram shows how information in the system at hand is related. It clarifies which properties are inherited, and it indicates how the different objects relate to each other. Thinking about accurate classes and subclasses provides a better insight into the system and provides a sound basis for model implementation. Moreover, it leads to a modular structure, which makes it easier to attach new functions or indicators.

5. SYSTEM LEVEL MODELING

An illustrative example showing how the UML models were transformed into code (i.e. a system level representation), using Matlab, is shown below.

```
(1)   river = load ('stretches.txt');
      measure = xlsread ('measure.xls');
      Q = load('QBorgharen93.txt')';
(2)   ini_prof = initiate (river); % interpolate profiles between nodes
(3)   wlevels1 = manning(Q,ini_prof); % Q in rows, nodes in cols
```

In step (1), the data-files area read. The stretches.txt file contains a number of river stretches nodes, defined between the "beginnode" and the "endnode", each characterized by the attributes summarized in the UML static structure. Example of the data set of the text file is shown in Table 2. Other data-sets (not shown in the example code above due to space limitations) are the "measures" and "discharges" which are required for the calculations, as it is depicted by the static UML model. A series of average discharges per day is given, in this case for the year 1993. In step (2) the initial profiles per stretch node (shown in the model with the class "Rkm"; river kilometer) are interpolated from the stretches, by the subroutine 'initiate'. Step (3) then applies the subroutine 'manning' to calculate "Waterlevels" based on the given discharges and on the profiles per "Rkm". In the following steps, not depicted here, the measures are applied to their respective locations (Rkm's).

Table 2: Example of dataset 'stretch'

Elevation [m+NAP]	Channel width [m]	Channel depth [m]	Channel roughness [-]	Floodplain width [m]	Floodplain depth [m]	Floodplain roughghness [-]
45.62	130	7.03	0.03	10	14.38	0.029
18.22	90	10.04	0.03	2590	2.78	0.026
…	…	…	…	…	…	…

Figure 4 shows an example plot produced by the DSS of the original river stretch, and the river stretch after measure implementation. On the latter the 'manning' subroutine is applied again to calculate changes in water levels. The water levels, and old and new river descriptions, next provide input to the fuzzy agriculture and landscape modules.

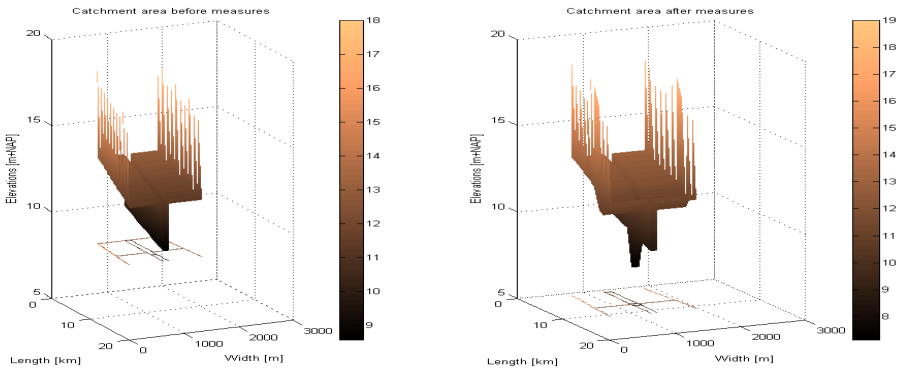


Figure 4: 3D plots of a river stretch before (left) and after (right) measure implementation

6. CONCLUSIONS

To conclude this paper, we present the results of our empirical evaluation that describes the "performance" of UML with respect to the criteria defined in section 3.3.1.

Ability to handle different spatial and temporal scales.
The static structure shows the coherence between objects on different spatial scales, and the temporal scales can be addressed in attributes of the objects (which for that purpose need be included in the conceptual model). Inclusion of differently scaled variables potentially increases the solution space and supports the comprehensiveness of the conceptual model.

Ability to deal with quantitative and qualitative relations.
UML semantics provides sufficient accuracy to come to a unified understanding of the meaning of the relations in the system without comprising the need to be overly formal. A disadvantage is the fact that in the natural system most relations are of a causal nature, requiring the incorporation of diagrams representing dynamics to clarify the way the system works.

Ability to deal with different actors and their respective information requirements.
The application of the use case diagram is the initial step that helps clarifying which actors need information from the model, and how the different kinds of information cohere with the natural system. It requires an accurate specification of the model users and possible other actors involved.

Design consistency.
Unlike many other graphical representation methods, the UML syntax and semantics strongly force one to think about the consistency of the design. Inconsistencies lead to overlaps, incompleteness or in the worst case to a failure to include all components in the same diagram set. The occurrence of either of the three therefore forces one to rethink the design, and in this manner avoid inconsistency issues during further implementation.

Flexibility with respect to model extension.
UML provides a sound basis for software documentation. This makes it easier to extend the model, than in comparable cases in which the documentation is less specific. Using the object-oriented approach allows treating objects as 'libraries', which can easily be extended.

Once matching w.r.t. attributes and operations, new classes, or for instance 'user functions' can easily be added. A disadvantage is that changes in relations between objects or users still require extensive re-engineering.

Support the communication with people from different backgrounds.
A disadvantage of using UML is that although the diagrams are relatively easy to intuitively understand, building them is not so straightforward. Particularly when more different diagram types are involved, quite some study may be necessary before the benefits of UML can be gained to the fullest.

ACKNOWLEDGEMENTS

We would like to thank Dr. Jean-Luc de Kok of University of Twente for his comments on the manuscript of this paper. We would also like to thank Georg Holtz and Dr. Marcela Brugnach of Osnabrück University for organizing a discussion session on UML, and we are grateful to all participants for sharing their experiences and commenting on our work. During the last phases of this work, Dr. Dokas, was supported by the research project SCEWA (Grand No 2007-DRP-2-S5), funded by the Irish Environmental Protection Agency under the DERP grand scheme. Miss Janssen was supported by the Cornelis Lely Stichting in the framework of her PhD research.

REFERENCES

Argent, R. ad B. Houghton. Land and water resources model integration: software engineering and beyond, *Advances in environmental research*, 5, 351-359, 2001

Barthel, R., D. Nickel, A. Meleg, A. Trifkovic, and J. Braun. Linking the physical and the socio-economic compartments of an integrated water and land use management model on a river basin scale using an object-oriented water supply model, *Physics and Chemistry of the Earth*, 30, 389-397, 2001

De Kok, J.L and S. Kofalk. Towards a user-oriented design of a DSS for Integrated River-Basin Management: the Elbe Prototype DSS. In: A. Voinov, A. Jakeman, and A. Rizolli (Eds.), Proc. of the IEMSS Third Biennial Meeting: "Summit on Environmental Modelling and Software", July 9-12, 2006, Burlington VT, USA, 2006

Hillyer, C., J. Bolte, F. van Evert and A. Lamaker. The ModCom modular simulation system, *European Journal of Agronomy*, 18, 333-343, 2003

Maxwell, T. and R. Costanza. A language for modular spatio-temporal simulation, in *Ecological modelling*, 103, 105–114, 1997

Matthies, M., J. Berlekamp, S. Lautenbach, N. Graf and S. Reimer. Systems analysis of water quality management fort he Elbe river basin, *Environmental modeling and software*, 21, 1309-1318, 2006

Miles, R. and K. Hamilton '*Learning UML 2.0*', O'Reilly, 2006

Pahl-Wostl, C., J. Möltgen, E. Ebenhoeh, and G. Holtz. The NeWater management and transition framework – state and development process, in: C. Pahl-Wostl, P. Kabat and J. Möltgen (eds.) 'Adaptive and integrated water management – coping with complexity and uncertainty', CAIWA-conference, Nov. 12-15 2007, Basel, Switzerland, 2008

Papajorgji, P. and T. M. Shatar. Using the Unified Modeling Language to develop soil water-balance and irrigation-scheduling models, *Environmental modeling and software*, 19, 451-459, 2004a

Papajorgji, P., H. W. Beck and J. L. Braga. An architecture for developing service-oriented and component-based environmental models, *Ecological modeling*, 179, 61-76, 2004b

Pilone, D. and N. Pitman. UML 2.0 in a nutshell, 234pp., O'Reilly , 2005

Zadeh, L. A. Outline of a new approach to the analysis of complex systems and decision making, *IEEE Transactions on systems, man and cybernetics*, SMC-3, 28-44, 1973

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Model development for integrated water management in the influence area of phosphate mines

Kathy Bru[a], Anne-Gwenaëlle Guézennec[a], Sandra Lanini[b], Nina Graveline[b], Konstantinos Soulis[c], Fatma Karaouli[d] and Anis Guesmi[d]

[a]*BRGM Environmental & Process Division, Orléans, France (k.bru@brgm.fr, a.guezennec@brgm.fr)*

[b]*BRGM Water Division, Montpellier, France (s.lanini@brgm.fr, n.graveline@brgm.fr)*

[c]*Agricultural University of Athens, Department of Natural Resources Management and Agricultural Engineering, Athens, Greece, (soco@aua.gr)*

[d]*National Agronomic Institute of Tunisia (INAT), Rural Engineering Department, Tunis, Tunisia, (elmaainat@yahoo.fr)*

Abstract: The phosphate industry is a major contributor to the economy of some Mediterranean countries such as Morocco, Tunisia and Jordan. Large volumes of water are required by the mining industry in areas where water resources are scarce or limited. The increasing water demand, due to many factors such as industrial, agricultural and socio-economic development, combined with a pronounced climatic variability threatens groundwater resources in terms of quantity and of quality. This water scarcity is alarming as it represents a limiting factor to the sustainable development of these regions and it may result in conflicts between the water users. This paper presents the methodology currently applied in the European research project Elmaa to develop a Decision Support System (DSS) in two case studies, one in Tunisia and one in Morocco. The main objective of these DSS is to help water managers in improving water resources management at a regional scale. To achieve this objective, an integrated approach is implemented. This approach includes a realistic description of the socio-hydrosystem of the studied area, definition of scenarios for water resources and water demands evolution for the next 25 years, results of hydrologic and hydrogeologic modelling and results of socio-economic studies. By comparing water scenarios, it allows determining ways to optimise water consumption and to improve water resources conservation.

Keywords: Integrated water management; DSS development; Phosphate mines; Mediterranean countries.

1. INTRODUCTION

The phosphate industry is a major economic sector for some Mediterranean countries (i.e. Morocco, Tunisia, Jordan), with a contribution between 3 and 5% of their Gross Domestic Product. This industry demands large volumes of water, its total fresh water consumption being evaluated in 2001 to 57, 20 and 15 million m^3 in Morocco, Tunisia and Jordan respectively. An increase of fresh water used by the phosphate industry is expected for the next years due to the fast-growing phosphate production. Combined with the population growth, the drinking water access improvement and the extension of irrigated areas for agriculture, this leads to a continuous increase of water demand in countries which have to face water scarcity due to their arid or semi-arid climate. The pressure on water is so serious that it results in competition between drinking water supply and economical activities such as phosphate industry, agriculture and tourism. This brings the need for an integrated water

resources management including a more efficient use of the existing water resources, a control of water demands, and the search of new water resources.
The European project Elmaa aims to reduce tensions on water resources around phosphate mines at regional scale through the development of a Decision Support System (DSS). A DSS is an interactive computer-based system intended to help decision-makers in using data, documents, knowledge and/or models to identify and solve problems [Power, 2004]. Regarding water management, it generally involves at least two building blocks among [Denzer, 2005]:

- Integrated models able to simulate the behaviour of socio-hydrosystems, coming from different fields of science such as water sciences (hydrology, hydrogeology), agronomy (crop needs, irrigation), hydraulic (drinking water supply network), socio-economy (human activities), etc;
- Geographical Information System (GIS);
- Database management system (data, information, results);
- Scenario evaluation module (multi-criteria analysis, economical analysis (cost-effectiveness and cost-benefit analysis), etc);
- Convivial Graphical User Interface (GUI).

Some examples of DSS are 'WaterWare' [Jamieson and Fedra, 1996], the 'MULINO' DSS [Mysiak et al., 2005], and more recently 'WaterStrategyMan' DSS [Maia and Schumann, 2007].

After a description of the methodology adopted in the Elmaa project to develop a DSS, this paper briefly presents the work performed to implement it for two case studies namely the Gafsa mines area in Tunisia and the Khouribga mines area in Morocco.

2. METHODOLOGY FOR INTEGRATED WATER MANAGEMENT

2.1 Overview

Integrated water management involves evaluating the evolution of the water resources and the satisfaction of the water demands on the studied area for the next years. For this purpose, an integrated model has been proposed in the Elmaa project, which consists in:

- a conceptual model,
- scenarios of water resource evolution and water demand evolution,
- a calculator which performs water balance and gives useful information for water management.

This model constitutes the heart of the final Decision Support System (DSS) which aims at defining and ranking the most pertinent actions regarding water management. In the Elmaa project, DSS building is carried out with the support of stakeholders who are representative of the major water users in the studied zones, namely drinking water and sanitation, agriculture and industry (here mainly mining).

2.2 The conceptual model

A conceptual model is a simplified but realistic representation of the socio-hydrosystem of the studied area. It identifies and describes water resources called "Hydrological units" (climate, groundwater, surface water), water demands called "Socio-economical units" (agriculture, drinking water, mines, other activities) and the water flow transfers between the different units. Wastewater treatment plants are considered as water resources which belong to socio-economic units. The principle of the conceptual model is presented in Figure 1.

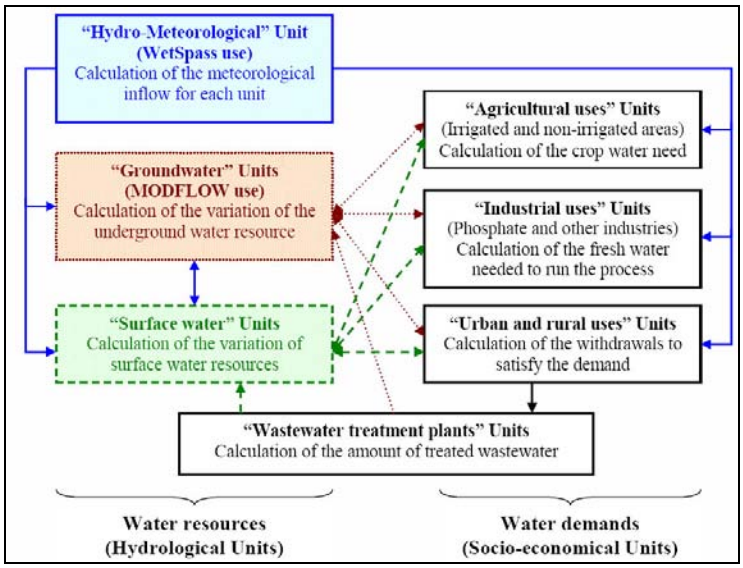


Figure 1. Scheme of the conceptual model

2.3 The scenarios

Scenarios of evolution of water resource and water demand, called here "user scenarios", were defined in the Elmaa project for the next 25 years. They are composed of one climatic scenario, one scenario for water demand evolution and one water management option.

A climatic scenario consists in 25 years-long sequences of daily precipitation and evapotranspiration which show the same statistical properties (auto-correlogramm, return period of rainy events, etc.) as historical series. Three climatic sequences were created (average, dry and wet). Each of them was simulated with "Wetspass", a physically based water balance model for the estimation of long-term average spatial patterns of groundwater recharge, surface runoff and evapotranspiration [Batelaan and De Smedt, 2001].

For the *scenario of water demand evolution*, several social, economical, industrial and agricultural parameters which could impact the water demand were identified (e.g. population growth rate, drinking water consumption, irrigated area and phosphate production…). For each of them, a limited set of potential values is listed according to literature review, prospective exercise and expert opinion. Crossing all the possibilities leads to a set of potential water demand values for each use for the next 25 years.

Lastly, *water management options* were defined by the stakeholders of the project to face the problem of water scarcity. A wide range of options was proposed such as modification of the water allocation scheme, use of non-conventional resources (treated wastewater for example) or technological improvements in the mine process. An economical study has been conducted to analyse these options and to evaluate their global cost, their efficiency (in term of saved m^3 of fresh water) and their potential environmental and sanitary impact.

The choice of one scenario of water demand evolution and of one water management option determines the total withdrawals needed to satisfy each use and their links to a given resource (according to the supply scheme which can be defined by the management option). Using these data and the Wetspass results as recharge input, the evolution of groundwater levels and base flow to the rivers are simulated for each user scenario with a groundwater flow model developed with MODFLOW [Merritt and Konikow, 2000]. An extension of the MODFLOW model allowing combination with the WetSpass model is used here [Batelaan et al., 2003]. For a given user scenario, each "Groundwater" and "Surface water" unit is then associated to a dataset describing its inflows and outflows whereas each "Socio-economical" unit is associated to a dataset describing its water demand and its water supply.

2.4 The calculator

By integrating all the information described above, the calculator gives an overview of the evolution of water resources and the satisfaction of water demands on the studied area for the next 25 years. For this purpose, it performs water balances at a monthly time step for each unit. Since the calculator is statically linked to the unit models of the conceptual

model, calculator running on a user scenario involves simulations with the WetSpass and Modflow models are completed for this scenario. For resource units, this water balance aims to estimate the available water (storage or river flowrate) by difference between inputs and outputs and for demand unit it aims to compare water demand and water supply. The calculator computes four variables to describe the results of these water balances:

- the water volume available in each groundwater V:

$$V = \varepsilon \cdot S \cdot H \quad \text{with} \quad \begin{array}{l} S\text{: groundwater area (m}^2\text{)} \\ \varepsilon\text{: storage coefficient of the groundwater} \\ H\text{: piezometric level (m)} \end{array} \tag{1}$$

- the flow of the main downstream river $Q_{dowstream}$ (if the case study has a perennial river):

$$Q_{downstream}(m^3) = Q_{upstream}(m^3) - \sum surface_withdrawals(m^3) \tag{4}$$

- the satisfaction of the total needs (monthly and annually) S_T:

$$S_T = \frac{\sum withdrawals(m^3)}{\sum needs(m^3)} \tag{2}$$

- the satisfaction of the needs per use (monthly and annually) S^i:

$$S^i = \frac{withdrawals(m^3)}{needs(m^3)} \tag{3}$$

The calculator also gives data and information concerning economical, environmental and sanitary aspect relating to the chosen water management option.

2.5 The Decision Support System (DSS)

The DSS is the final tool which will be used by the actors of water management to explore the impact of various water management strategies combined to climatic scenarios by comparing them against a baseline scenario. The baseline scenario is defined as the current state of the water system under the assumptions that in the coming years water availability and water demand will follow the currently observed and forecasted trends. The comparison between simulated and baseline scenario is made through indicators which are calculated from data given by the integrated model. These indicators are related to water resources characteristics, effectiveness in water demand coverage, water consumption evolution, environmental impacts and costs. Examples of indicators are the following: amount of withdrawals and available water resources, piezometric level of each aquifer, satisfaction of the total water demand, satisfaction of the water demand per use, evolution of the total water consumption with respect to the consumption in 2005, costs, etc.

A Graphical User Interface (GUI) will be created to facilitate the interactions of the DSS user with the integrated model, namely the selection and setup of the simulated scenario, and to provide him with different forms of results presentation, including tables, graphs, maps, etc.

By evaluating different ways to optimise the water consumption and its re-use and by measuring the potential impacts of changes in the mining and agricultural practices related to technological innovations or evolutions in the institutional or regulatory framework, this DSS will help water managers to adopt plans to improve water availability, its conservation and its sharing between the different users in the influence area of phosphate mines.

2.6 Stakeholders involvement

Stakeholders were involved from the early stage of the project within the ‘stakeholders committee’ in order to create and to keep a permanent link between research consortium and local water management actors. While the ‘resource’ part of the system is knowledge and project research based, the ‘user system’ is defined with the support of a stakeholder committee. Three types of consultations were organized. The first type consisted in their participation to meetings of the consortium for them to express their opinions and validate the work of the working groups (all case studies together). The second type, which was individual consultation, aimed at collecting their point of view and above all their indicators

for a good water management, these indicators being a major component of the DSS. The third type of consultation is to gather them in workshops (stakeholders of the same case study together) in order to induce discussions for a common validation of management options. The objective was not to select management options accepted unilaterally but to sweep across a majority of possible management options to evaluate their impact and then to find a consensus among them. This appears to be a real distinct task for which the results of the modelling platform should be useful (i.e. the assessment of the impact of each management option should help for decision making). The different forms of consultation experimented are complementary in the sense that they bring all different information, thus they should not be neglected.

3. CASE STUDIES

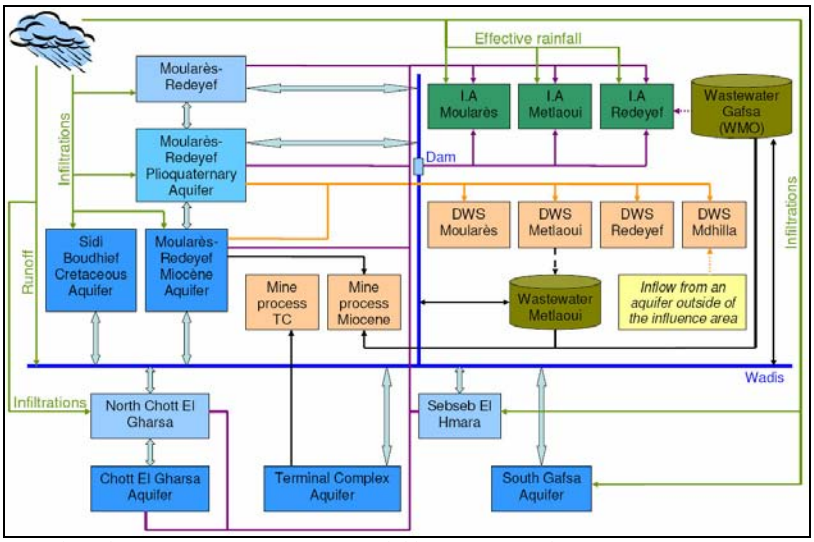
The following sections present the implementation of this methodology to develop a DSS for a Tunisian and a Moroccan case study, namely the Gafsa mines area and the Kouribga mines area.

3.1 Gafsa mines area, Tunisia

The Gafsa mines influence area is situated 350 km south of Tunis near the north-eastern part of the desert. This region is known for its industrial activities related to phosphate (extraction, production of phosphoric acid and of fertilizer) and farming which both compete for water resources. The influence area has a population of 323,709 inhabitants, including 73% living in urban zones and 27% in rural zones. It has a complex hydrographic network mainly made of small non-perennial rivers. It contains 9 aquifers including 3 water tables (Moularès-Redeyef, North Chott El Gharsa and Sebseb El Hmara) and 6 deep aquifers (Moularès-Redeyef Plioquaternary and Miocène aquifers, South Gafsa aquifer, Sidi Boudhief Cretaceous aquifer, Terminal Complex aquifer and Chott El Gharsa aquifer).
Gafsa region, which is classified in the bioclimatic arid level, has a dry subtropical climate characterized by very cold winter, dry hot summers and low and irregular precipitation. The continuous increase in water demand combined with these unfavourable climatic factors result in depletion and degradation of the water resources. Indeed, water demand is increasing due to the population growth, the extension of irrigated areas and the fast-growing mining exploitation. Currently, the phreatic aquifers, which are entirely allocated to agriculture, are entirely overexploited with an exploitation rate of 113%, while the exploitation rate of the deeper aquifers reaches 80%. The withdrawals in deep aquifers are mainly used by industry (76%), the rest being divided between agriculture (13%) and drinking water (11%) [DGRE, 2001a; DGRE, 2001b; DGRE, 2005a; DGRE, 2005b].
The objectives of the DSS developed for the Gafsa mines influence area are to allow a better allocation of the water resources between all the water users and to improve water management, e.g. by proposing technological innovations for water savings. The DSS is still under construction but several steps have already been performed: building of the conceptual model, definition of the parameters for the water demand scenarios and description of the water management options.

The conceptual model derived from the analysis of the Tunisian studied area reveals the complexity of the hydrology of the region and of the links between water demand units and water resource units. As it can be seen in Figure 2, the socio-hydrosystem is made of one unit of surface water, nine units of groundwater resource interacting one another and several units related to water drinking supply, wastewater treatment plants, industrial activities (phosphate mining processes) and irrigated areas for agriculture.

Parameters influencing the water demand in Gafsa area were defined during workshops with the stakeholders (Table 1). For each parameter, two or three values were fixed: a “low value” corresponding to a decrease in water consumption, a tendency value (following the Business As Usual) and a “high value” corresponding to an increase in water consumption. Combination of all parameters fixed at the tendency value defines the baseline scenario. The set of scenarios of water demand evolution which will be included in the final DSS will be constructed by combining all the possibilities of these parameters.



Legend: I.A = Irrigated Area DWS = Drinking Water Supply
TC = Terminal Complex WMO = Water Management Option

Figure 2. Conceptual model of the Gafsa mines influence area (Tunisia)

Table 1. Parameters table

Parameters	Low value for the water consumption	Tendency value (Business As Usual)	High value for the water consumption
Population growth rate	- 0.3% per year	0.51% per year	1.21% per year
Drinking water consumption	BAU decrease of 2% i.e. 4.1 Millions m^3/year	4.2 Millions m^3/year	BAU increase of 2% i.e. 4.3 Millions m^3/year
Efficiency of the waterworks system (Efficiency * wear rate)	Improved efficiency: 90 %	Current efficiency: 80%	
Connection rate with the wastewater collection system		82%	100%
Irrigated area (ha)	Current value: 3371 ha	Projected value: 6027 ha	
Cropping pattern		Current cropping pattern: Perennial crops: 55% Cereals: 3% Fallow: 18% Vegetables: 22%	Other cropping pattern with a higher water consumption
Irrigation efficiency	Improved irrigation system: Improved furrow irrigation: 70% Sprinkle irrigation: 80% Drip irrigation: 90%	Current irrigation system: Furrow irrigation: 50%	
Phosphate production		8 Millions tons/year	BAU increase of 30% i.e. 10.4 Millions tons/year
Volume of fresh water used to produce 1 ton of phosphate		2.25 m^3/ton	
Water recycling rate in the washing phosphate plant	60%	40%	
Water consumption by the other industries		3 Millions m^3/year	BAU increase of 2% i.e. 3.1 Millions m^3/year

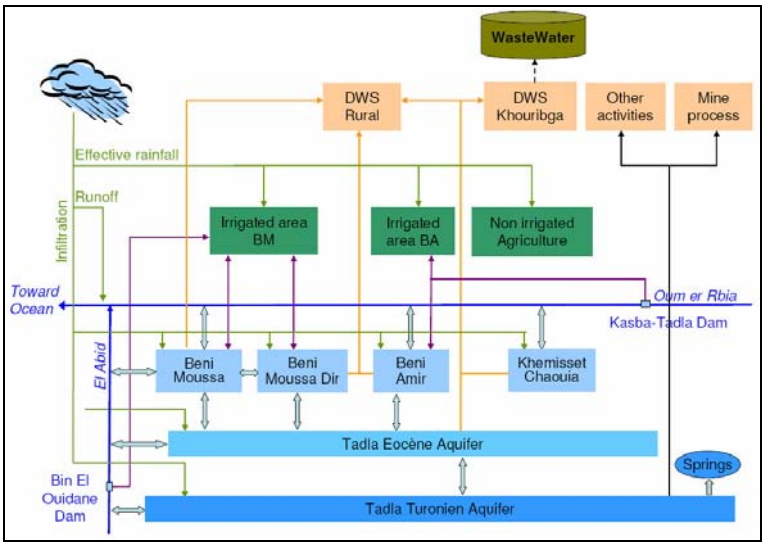
Definition of water management options, carried out with the stakeholders, includes an exhaustive description of each option (technical aspects, implementation, constraints and incentives) and an evaluation of its potential impacts for the environment, the economy and its acceptability. For the Gafsa mines influence area, 9 water management options were defined:

- 2 management options regarding the hydrology:
  - construction of groundwater recharge systems,
  - generalization of water spreading systems.
- 3 management options regarding the agricultural uses:
  - use of water saving systems in the agricultural areas,
  - use of treated wastewater for irrigation,
  - use of phosphate schlamms for agriculture.
- 3 management options regarding the industrial uses:
  - use of treated wastewater for phosphate washing,
  - improvement of the water recycling in the mine process,
  - increase of the charge on the water used in the mine process.
- 1 management option regarding the drinking water supply
  - use of systems to decrease household water consumption.

3.2 Khouribga mines area, Morocco

The influence area of the mining site is located 140 km southeast of Casablanca and covers three provinces: Settat, Khouribga (region of Chaouia Ourdigha) and Beni Mellal (region of Tadla Azilal). The zone extends over 21,075 km^2 which represents about 3% of the national surface area. It is included in the Oum Er Rbia large river basin. Even if this area is characterized by an arid climate with extreme variability of precipitations, the Oum Er Rbia is a very regular stream, its lowest water flow being about 40m^3/s. Two main aquifer systems can be distinguished on the influence area (see figure 3):

- Tadla multilayer system which comprises 3 aquifers partly confined. Only two of them are exploited, due to the poor capacitive properties of the second one.
- Plioquaternary phreactic aquifers (Khemisset Chaouia, Beni Moussa, Beni Moussa Dir, Beni Amir).



Legend: DWS = Drinking Water Supply BM = Beni Moussa BA = Beni Amir

Figure 3. Conceptual model of the Khouribga mines influence area (Morocco)

The Khouribga mines influence area comprises 111 towns including 101 rural and 10 urban towns. Its density is 80 inh./km^2 for a population of 1,300,000 inhabitants. Drinking water is distributed in cities whereas the rural districts are not yet all equipped. The drinking water access rate is about 60% in Khouribga and Settat provinces and 90% in the Beni Mellal province. The studied region is characterized by diversified activities: agriculture (65% of the area) including rainfed areas (69%) and irrigated areas (31%), drinking water supply (urban and rural), mining industry, agro-food industries (sugar industries) and tourism. All these activities are dependant on water resources. Behind the evident problem of water quantity, the water resources quality is also a major issue since agricultural intensive practices (use of fertilizers and pesticides) and lack of treatment for most of the domestic and industrial wastewater threatens the water resources of this area [M'Hamdi et al., 2007].

Similar to the Tunisian case, the DSS regarding the Khouribga mines influence area aims at improving water availability, its quality and its sharing. The development of this DSS is still in progress but the conceptual model (Figure 3) and the description of the water management options have already been carried out. The defined water management options are the followings:

- 4 management options regarding the industrial uses:
  - improvement of the water recycling in the washing plant of Daoui,
  - use of water from dams for phosphate plant,
  - use of nitrated water for phosphate washing plant,
  - use of urban wastewater for phosphate washing plant.
- 1 management option regarding the drinking water supply
  - use of water from dams for drinking water supply.
- 1 management options regarding the agricultural uses:
  - use of water saving systems for irrigation.

4. CONCLUSION

The methodology implemented to develop a Decision Support System (DSS) is presented in this paper. This DSS aims at improving water management on the influence area of two phosphate mining sites considered in the Elmaa project, namely the Gafsa mines area in Tunisia and the Kouribga mines area in Morocco. These DSS building, developed with an integrated approach and in collaboration with stakeholders, has involved performing a realistic description of the socio-hydrosystem of the studied area, defining scenarios for water resource and water demand evolution for the next 25 years, performing hydrologic and hydrogeologic modelling and conducting socio-economic studies. Several water management scenarios will be simulated with these DSS and will be compared through the indicators given by each simulation. Even if these DSS are still in progress, it can already be concluded that they will constitute a powerful tool for helping water decision makers in improving water availability, its quality and its sharing between the different users.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to the European Commission for funding the Elmaa project (Integrated water management of Mediterranean phosphate mining and local agricultural systems) (INCO-CT-2005-015410) and all the participants for their help and commitment.

REFERENCES

Batelaan, O. and F. De Smedt, WetSpass: a flexible, GIS based, distributed recharge methodology for regional groundwater modelling, paper presented at the Sixth International Association of Hydrological Sciences (IAHS) Scientific Assembly, IAHS, Maastricht, The Netherlands, 2001.

Batelaan, O., F. De Smedt and L. Triest, Regional groundwater discharge: phreatophyte mapping, groundwater modelling and impact analysis of land-use change, *Journal of Hydrology*, 275, 86-108, 2003.

M'Hamdi, R., A. Chik, A. El Kanit and J. Briuni, D1-1 Annex 1: Moroccan contribution to the D1-1 (Synthesis Report on the Moroccan, Tunisian and Jordan case study), written by CERPHOS, Elmaa project, 65 p., 2007.

Denzer, R., Generic integration of environmental decision support systems - state of the art, *Environmental Modelling & Software*, 20, 1217-1223, 2005.

DGRE : Direction Générale des Ressources en Eau, Situation de l'exploitation des nappes profondes de Tunisie, *Publication DGRE,* 2001a.

DGRE : Direction Générale des Ressources en Eau, Situation de l'exploitation des nappes phréatiques profondes de Tunisie, *Publication DGRE,* 2001b.

DGRE : Direction Générale des Ressources en Eau, Situation de l'exploitation des nappes profondes de Tunisie, *Publication DGRE,* 2005a.

DGRE : Direction Générale des Ressources en Eau, Situation de l'exploitation des nappes phréatiques profondes de Tunisie, *Publication DGRE,* 2005b.

Karaouli, F., Etude sur modèle hydrogéologique du système aquifère de Moulares-Redeyef, *Master's thesis*, Institut National Agronomique de Tunisie, 2005.

Jamieson, D.G. and K. Fedra, The "WaterWare" decision-support system for river-basin planning. 1. Conceptual design, *Journal of Hydrology*, 177(3-4), 163-175, 1996.

Maia R. and A.H. Schumann, DSS Application to the Development of Water Management Strategies in Ribeiras do Algarve River Basin, *Water Resources Management*, 21, 897-907, 2007.

Merritt, M.L. and L.F. Konikow, Documentation of a computer program to simulate lake-aquifer interaction using the MODFLOW ground-water flow model and the MOC3D solute-transport model, Water-Resources Investigations Report 00-4167, 146 p., 2000.

Mysiak, J., C. Giupponi and P. Rosato, Towards the development of a decision support system for water resource management, *Environmental Modelling & Software*, 20(2), 203-214, 2005.

Power, D.J., Decision Support Systems Web Tour, *http://dssresources.com*, version 4.3, 2004.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Aspects of the application of decision and information support tools in an integrated water resources management practice

R. Hranova

Civil Engineering Department, University of Botswana, Gaborone, Botswana
(hranova@mopipi.ub.bw)

Abstract: The paper presents some problems with the use of modelling, decision, and information support tools (MDISTs), for the implementation of integrated approaches to water resources management. The problems arise because in many cases current monitoring practices do not provide sufficient amount of reliable data for model calibration and verification. The paper emphasises the need for a system analysis approach at all levels of magnitude in a water system - considering all elements, subsystems and their interactions – and the integration of the MDISTs with the corresponding monitoring practices, regulatory instruments and management activities in a closed-loop cycle. Finally, we discuss how MDISTs could be used in the particular case of developing countries.

Keywords: water systems; monitoring; regulation;

1. INTRODUCTION

The integrated approach to water resources management (WRM) practices has been accepted as a leading concept in the field. This approach requires a catchment orientated analysis and the consideration of factors and interactions in the following directions: 1) All natural aspects of the system: surface and ground water, the physical behavior of water, environmental water requirements, impacts and interactions with other environmental media, such as soil and air; 2) Simultaneous consideration of both quantitative and qualitative aspects of water resources; 3) All sectors of the national economy that depend on water, with corresponding engineering structures; 4) The relevant national and local objectives and constraints: legal, institutional, financial, and environmental; 5) The institutional hierarchy and arrangements at international, national, provincial, and local levels, and their corresponding interactions; 6) The spatial variation: upstream and downstream interaction, basin-wide analysis, inter-basin transfer.

It is clear that the implementation of an integrated water resources management (IWRM) practice is a multi-objective task that requires the consideration of numerous factors, and conditions, expertise from different fields of specialization, as well as a large amount of information that needs to be collected, organized, analysed and presented in clear and understandable form. This task cannot be achieved in a sustainable and objective way without the use of MDISTs. The aim of this paper is to emphasize the need for practical application of MDISTs and to discuss specific shortcoming (gaps) in the process.

2. THE SYSTEM APPROACH TO IWRM

The system analysis approach requires the development of the system architecture which identifies the major elements, subsystems and their interrelationship. Often, the formulation of the system is highly subjective, and according to Lendaris [1986], "the system is in the eye of the beholder". Still, the formulation of the conceptual framework can be defined in terms of individual tasks or problems. From the perspectives of IWRM, the basic system under consideration is the catchment basin of given water body [WFD, 2007], including all physical elements: natural water bodies, and man-made structures that influence the water resources status. This paper concentrates on the physical aspects of the catchment elements, as a basis for the system framework formulation, but also emphasizes interconnections among system elements and some "soft" issues, such as institutional and legal arrangements, public involvement and socio-economic development.

An example of a system framework at basin level is presented in Figure 1. The elements of the system include surface and ground water natural resources, as well as, man-made water subsystems that use and influence natural water bodies. A sustainable approach to IWRM requires assessment and simultaneous consideration of both quantitative and qualitative parameters which characterize the current status, trends and variations due to human interaction with enough accuracy. The major quantitative parameters under control are flow rates, rainfall volumes or intensities, water volumes per given period of time and surface/ground water levels. The major water quality parameters under control - such as BOD/COD, solids, specific ions, nutrients, toxic substances, microbiological characteristics - indicate the environmental health of the system. Qualitative assessments of the system behaviour are often estimated based on pollution transport principles and are represented as pollution loads (the product of flow rates and pollutant constituents' concentrations). Mass-balance principles are applied to assess the status of the system and alterations due to human interference.

Under an IWRM approach, all factors shown in Figure 1 should be considered during the decision-making process. Each one of the natural water bodies (rivers, lakes/dams, and ground water aquifers) are usually analysed and modelled at basin or inter-basin levels. One of the most widely applied tools for such analysis is BASINS, developed by USEPA, which provides a framework for integrated basin analysis, based on national water quality, soil and climate data, GIS and different environmental modelling and assessment tools.

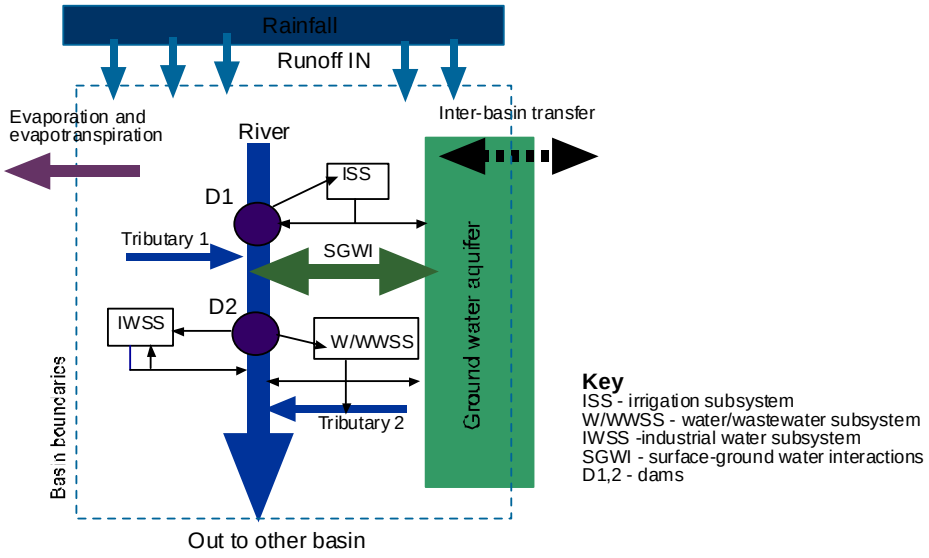


Figure 1 Conceptual framework at basin level

It incorporates the QUAL 2, the HSPS and alternatively the SWAT models. QUAL 2 simulates contaminants transport and transformations in rivers and streams, while the HSPS and SWAT models simulate and allow the evaluation of the impact of different management scenarios on the basin hydrology, water quantity and quality status (point and diffuse pollution), including surface water, sediment transport, ground water and their interactions. HSPF is a lumped parameter model (FORTRAN), while SWAT is based on a command language and includes the subdivision of the basin on hundreds to thousands sub-basins, thus allowing great spatial detail. Comparison of the performance of both models applied to a rural area with intensive agricultural activities [Saleh and Du, 2004], and to a predominantly forest area with urban development [Im et.al, 2003], shows that HSPF requires a bulk amount of data and a more complex calibration process. Regarding their predictive accuracy, both models show more or less similar characteristics, except that HSPF gives better prediction of the variation of daily flows and sediments, while SWAT estimates more accurately nutrient loadings, except for phosphorous.

The implementation of the IWRM approach in the European Union (EU), under the Water Framework Directive [WFD, 2000], also promotes the need for a systematic analysis. It sets a staged approach to an IWRM implementation, with a goal to achieve a good status of the EU's water resources by the year 2015. The DHI group offers an integrated tool to support this goal by a set of software products such as MIKE BASIN (water allocation), MIKE 11 (rivers and streams), MIKE SHE (surface/ground water interactions) and MIKE LOAD (a pollution load estimator). In addition, it provides software to incorporate the analysis of engineering subsystems in urban and rural areas.

The engineering subsystems are entities on their own, involving complex configuration and processes involved. Therefore, a different level of analysis (in terms of spatial magnitude) is necessary in order to the define inputs and outputs affecting the basin system. Figure 2 shows an example of a water engineering subsystem related to the use of water in population centres with a combined sewerage. It comprises of elements and third level subsystems, such as water/wastewater treatment plants and distribution or wastewater/storm water collection networks. Under a separate wastewater system, storm water conveyance and flood prevention structures would be represented as an independent subsystem. It should be noted, that the subsystem in Figure 2 reflects centralized solutions of both water and wastewater conveyance and treatment. Decentralized solutions regarding water supply or wastewater structures would lead to a different system framework and to different levels of magnitude of space – suburb, neighbourhood, individual building.

One important aspect during the engineering subsystems' analysis is the consideration of existing and planned land use patterns in the catchment and the specific urban development plans. The link between land use planning and water resources management is often overlooked. New development projects, especially in urban areas, relay on classic supply-orientated water management solutions. They do not explore and apply thorough analysis and innovative concepts that consider improved water demand and storm water management practices, and reduction of pollution loads to the environment.

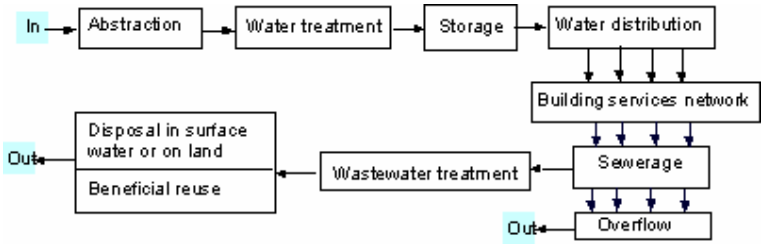


Figure 2 Conceptual framework at population centre level

The implementation of demand orientated approaches, decentralized wastewater systems, rainwater harvesting, and water reuse alternatives could reduce significantly future water demand estimates, and would lead to more sustainable water use scenarios, but only if anticipated during the planning stage of urban development plans and after thorough evaluation of all technical, economic and social factors [Hranova, 2006a]. In this respect, Hurley et. al., [2007] describe an urban development projects' assessment tool (Flexiframe), which integrates information across disciplines, stakeholder groups and professions, and also considers economic, environmental, social and technical criteria. Another example of the application of integrated approaches to urban planning and urban water systems management is the development of sustainability assessment methods and tools [Carden et.al., 2007]. Spears [2006] underlines some drawbacks in the policy formulation and the decision-making process regarding a sustainable urban water development, and points out that the application of new information technology tools could speed up the process and lead to a "leap frog" change towards sustainable solutions.

In general, the systems analysis approach to IWRM requires the definition of all systems elements and sub-systems at different levels: starting from the smallest systems in terms of spatial magnitude and following up the different levels up to basin level of major rivers (continental scale). Such an analysis is often narrowed or not implemented, because in many cases local and national administrative boundaries do not coincide with river basin boundaries, and also, in many cases of major rivers, the basin spatial magnitude involves a number of countries. Lack of coordination and optimal use of available information is aggravated by the fact that systems at the basin level are managed by national authorities, while engineering subsystems are managed by local authorities, and the corresponding information is collected, used and controlled by different institutions. In addition, water resources managers are usually focusing only on quantitative aspects only and their major concern is to provide the estimated demand for water. On the other hand, qualitative aspects are the domain of environmentalists and their major objective is to protect the status of the water resources. Both aspects, if not integrated under the system approach, often lead to conflicting management decisions. The implementation of the system approach avoids this by allowing for an integrated analysis of all aspects of the system at its different levels, and contributes considerably to the formulation of sound and sustainable objectives of water resources management programs.

3. INFORMATION NEEDS AND MONITORING PRACTICES

Water management institutions at different levels perform regular monitoring programs regarding quantitative and qualitative parameters in order to assess the status of water resources and the impacts of man-made engineering structures. The data collected is the basis for calibration and validation of any modelling or decision-support tool, and in addition, serves as a benchmark or as an environmental indicator. Therefore, the successful application of such tools depends heavily on the amount and reliability of such data.

3.1 Gaps in the monitoring practices

The structure of any monitoring program depends on the formulated objectives, but the major characteristics are: monitoring network (sampling locations), types of monitored parameters (with specified testing methods and accuracy levels), and frequency of measurements. Under an integrated and systematic approach to WRM, monitoring programs should supply reliable, relevant and accurate information to support the system analysis process and the application of MDISTs. In practice, several problems related to monitoring can be mentioned.

Monitoring networks should be designed to provide information regarding both quantitative and qualitative parameters, at specific points of interest, as inlets and outlets to systems and subsystems, as well as, data about background (natural) pollution. Unfortunately, existing monitoring networks often do not conform to these requirements

and it would be advisable to reassess the existing networks and to modify them in order to fit the system analysis process.

Testing methods and level of accuracy of measured parameters should be uniform throughout the system analysis process, in terms of space and time, so that it allows for comparison and estimation over the whole system. However, it is common for different laboratories to apply different testing procedures for the same parameter, even in the same population centre, thus leading to ambiguity of the data obtained. Standardization of the testing procedures throughout the system and the creation of common databases for system analysis would contribute considerably towards reducing the errors and uncertainties related to MDISTs applications and, in addition, would contribute to the optimal use of data obtained by different institutions. Also, it would eliminate ambiguity due to the differing magnitude levels of accuracy and ranges of measurements regarding monitored parameters. This applies specifically for water quality parameters: suspended solids, nutrients and BOD/COD are measured in mg/l; toxic metals are measured in µg/l, while some emerging constituents such as pharmaceuticals or endocrine disruptors are measured in nanograms/l. For example, a MDIST tool that supports the water quality management practice of Lake Biva [Ichiki et.al., 2005] provides information regarding the spatial variation of internal and external COD concentrations, with values varying between 0.13 and 4.4 mg/l; however, the American Standards [Standard methods, 1992], recommend a testing procedure with an acceptable level of accuracy between 5 and 50mg/l, and more accurate testing methods for concentrations above 50 mg/l. This shows that the use of such a data set to validate the MDIST would be unreliable and, as an alternative, the TOC (Total Organic Carbon) or other parameter could be used to investigate the organic matter variation in the lake, which is much more sensitive regarding low concentrations, with corresponding change in the monitoring practice.

The frequency of sampling incorporated in the monitoring program is another important point to be considered during the development, validation and application of MDISTs. In principal, the frequency of sampling should reflect with enough accuracy the status of the water body during the period when sampling is not done. For example, if measurements are taken once per month (monthly frequency), it should be anticipated that during this period the actual value of the measured parameter will not change significantly. Unfortunately, in surface water, both quantitative and qualitative parameters vary more often, especially during rainfall events. Increasing the sample frequency will make the monitoring program more expensive but the information provided will be more accurate. Ideally, automated monitoring stations would be the best alternative. However, they are more expensive and also, a number of important parameters can not be measured automatically. Therefore, the choice of monitoring frequency is a trade-off between the level of reliability of the data obtained and the cost of the information obtained.

3.2 Field data and model uncertainty

Model uncertainties are related to: input data and data used for calibration/validation, structure of the model, transformation functions and numerical operations. This section discusses uncertainties related to the use of data. In general, MDISTs are universal tools, based on scientifically proven assumptions, methods and computational techniques. However, their application in specific conditions requires locally obtained data that reflects with enough accuracy the operations and processes involved. Therefore, a reliable adaptation of given MDIST to the specific site conditions in the area would require a thorough validation process, based on locally obtained data from well established monitoring programs. If such data is missing, a data base from other locality with similar conditions might help, but such an approach should be used with caution. Thus, the adequacy of the available local data will determine at large extent the viability of the application of a MDIST.

Hydrological models estimate runoff, based on hydrological and climatic data and land use patterns, and logically, consider uncertain events (rainfall) as a random variable. However, MDISTs applied in water quality often don’t consider the randomness in water

quality data (as a consequence of rainfall), but work with fixed values, usually averaged annual parameters [Hranova, 2006b]. This approach might be acceptable for ground water quality modelling, where, due to the slow pollution transport process, pollutant concentrations do not show high variability. In contrast, surface water, storm water discharges and some point source discharges show much higher variability, and this needs to be considered during the monitoring process and during the models' development and application. Bertrand-Krajewski [2007] reports the result of an analysis of the influence of monitoring frequencies of a given data set on the calibration and verification of a simple model for evaluation of TSS Event Mean Concentrations (EMCs). It was observed that a significant difference in the frequency of measurements does not influence significantly the results of the calibration procedure. However, a significant impact was observed regarding the model verification procedures, such that the model outputs vary significantly, based on increased frequency, and furthermore, it indicates to a possible change in the model structure, which could not be observed, based on a data set with a smaller number of measurements.

4. REGULATORY APPROACHES AND THE APPLICATION OF MDISTs

The regulatory and legislative basis (legal and administrative documents, criteria and guidelines) is a key factor in IWRM. In respect to environmental and public health protection, the philosophy or the basic approach applied to the development and implementation of regulatory documents is the "Water Quality Objective (WQO)" approach, where the main goal is to maintain the status of natural water bodies at a required and specified level. This approach allocates different permissible pollution loads from man-made subsystems (considering both point and diffuse sources) at different locations, based on the current status of the water quality and its assimilative capacity. In the USA, a more detailed version is applied – the "Waste Assimilative Capacity Concept (WACC)", which requires the estimation of two parameters: the total maximum daily load (TMDL) from man made and natural sources (including background pollution), and the loading (assimilative) capacity of the water body under consideration. The comparison of the specific values of these parameters allows the determination of an admissible load (if any) to the surface water body, and based on this estimation, management decisions and plans may be developed. It is clear that both approaches require a systematic analysis at basin level, the collection of massive data sets, and the application of reliable tools for data analysis and processing. It should be emphasised, that the determination of the assimilative capacity of surface and ground water natural bodies could not be determined without the use of models describing with enough accuracy the pollutants transport and transformation processes. The proper development and application of such tools requires reliable and adequate monitoring data to reflect the specific conditions. However, many existing regulatory documents fail to provide enough specific information and detailed requirements, linking the management objectives to a well-specified monitoring practice (including networks, parameters with standardized testing methods and corresponding frequency of measurements) and a to recommendable MDIST, thus leaving gaps in the process of application of such an integrated and systematic approach.

The above-mentioned regulatory approaches allow for the estimation of impacts as a result not only of discharges from point or diffuse sources, but also, the impacts of significant volumes of water withdrawals from natural water resources. Another important aspect in this regard is the consideration of the environment as a legitimate user of water, thus providing the basis for a sustainable use of water resources and their preservation for future generations. In this regard, Breckenridge [2007] points out that current regulatory documents are focused predominantly on management actions in order to prevent pollution, but do not give enough emphasis on activities related to water bodies' rehabilitation, and suggests institutional changes, where local authorities or agencies need to be custodians of both the man-made subsystems and the natural systems under their jurisdiction. However, this raises the need of institutional arrangements, which should be structured in such a way that their authority could reflect the system as a whole.

5. MANAGEMENT PRACTICES WITH EMPHASIS ON DEVELOPING COUNTRIES

In the light of IWRM, the planning process of any activities related to natural or man-made water structures should be based on the outputs of the system analysis. It should incorporate "hard" (engineering structures) and "soft" activities, such as institutional arrangements, economic/financial analysis and public involvement, in conjunction. During this process, the application of MDISTs cannot be over-emphasised. They should be applied at all levels of magnitude in the system. In this regard, the development and analysis of several viable scenarios for a potential solution should be made with corresponding cost estimations and the choice of optimal solution, considering multiple factors, and often, multiple objectives. Thus, the selection process would be based on measurable and concrete criteria. For this purpose, linear, non-linear and evolutionary programming and optimization techniques, as well as, MDISTs based on them, are essential. Such approach to the evaluation of different alternatives is relatively simple to apply, and should be adopted not only by managing authorities but on a broader basis, by consultants working in the field. Furthermore, requirements for the development of multiple alternatives and a scientifically based choice of optimum alternative should be included during the tendering process of large water projects.

The considerations discussed in this paper are valid for all countries; however, developing countries usually face considerable challenges during the implementation process, due to lack of economic, technical and human resources. It is a common practice that management decisions are taken based on very limited local data, thus leading to subjective estimates and costly solutions. In addition, development goals often contradict environmental considerations and an acceptable trade-off should be made. Despite of these limitations, MDISTs should be implemented in the WRM practice because of the following considerations: 1) it would be an incentive for the implementation of proper monitoring and other data collection practice; 2) it would provide reliable means for record keeping and data accumulation; 3) it would provide for a sustainable and cost effective development; 4) it would help in an optimal use of available resources, which in the vast majority of the cases are limited; 5) it would provide for shearing of knowledge and general scientific information and methodologies among developing and developed countries.

5. CONCLUSIONS

The implementation of the IWRM practice requires an extensive, costly and technically intensive background, and many countries do not have the capacity to implement it on a broad basis. However, even if it would be implemented on a limited basis, in terms of spatial application and/or specified objectives, it requires a closed-loop cycle: system analysis at all levels → objectives→ conceptualization → regulatory basis→ MDISTs → monitoring programs → MDISTs → management practice→ system analysis. Such a cycle would contribute towards avoiding gaps in the implementation, helps to reassess the strategy periodically, and thus allows for a phased approach in the implementation process and for a sustainable use of the available resources.

REFERENCES

Bertrand-Krajewski, J.L., Field data requirements for monitoring and modelling of urban drainage systems, in Novotny, V. and Brown, P. (ed.) Cities of the Future: towards integrated sustainable water and landscape management, IWA Publishing, 2007.

Breckenridge, L., Ecosystem resilience and institutional change, in Novotny, V. and Brown, P. (ed.) Cities of the Future: towards integrated sustainable water and landscape management, IWA Publishing, 2007.

Carden, K., N. Armitage, S. de Carvalho, T. Stoeckigt, and M. Snoek, Development of a "sustainability index" for integrated urban water management in South Africa, in Ulanicki B., Vairavamoorthy, K., Butler, D., Bounds, P.L.M., Memon, F.A., (ed.) Water Management Challenges in Global Change, Taylor and Francis, pp 569-575, 2007

Hranova, R., Centralized municipal wastewater reuse projects in arid areas with emphasis on the planning process, *Proceedings of the IASTED International Conference - Environmentally sound technology in water resources management (ESTW) – 11-13 October 2006, Gaborone, Botswana.* (Proceedings in CD format), 2006a

Hranova, R., Specific aspects of diffuse pollution monitoring and implications on diffuse pollution management programs, *Proceedings of the 10th International Specialized Conference on Diffuse pollution and sustainable basin management", 18-22 September 2006, Istanbul, Turkey*. (Proceedings in CD format), 2006b

Hurley, L., S. R. Mounce, R. M. Ashley, and C. K. Makropoulus, Support for more sustainable decision making in urban water management, in Ulanicki B., Vairavamoorthy, K., Butler, D., Bounds, P.L.M., Memon, F.A., (ed.) Water Management Challenges in Global Change, Taylor and Francis, pp 577-584, 2007

Ichiki, A., A. Sakata, K. Sasaki, N. Nakakura and Y. Tai, Water quality estimation in consideration of pollutant runoff and internal production in Lake Biwa, Japan, *Proceedings of IWA-WISA Specialized conference on Diffuse pollution, 9-12 August 2005, Johannesburg, RSA* (Proceedings in CD format), 2005

Im, S., Brannan, K., Mostaghimi, S., and Cho, J., A comparison of SWAT and HSPF models for simulating hydrologic and water quality responses from an urbanizing watershed, ASAE meeting, paper No 032175, St Joseph, Michigan: ASAE, 2003

Lendaris, G.G., On systems and the problem solver: tutorial comments *Systems, man and cybernetics*, Vol SMC-16, 4, 604-610, 1986

Saleh, A., and B. Du, Evaluation of SWAT and HSPF within BASINS program for the Upper North Bosque River watershed in Central Texas, *Transactions of the ASAE*, 47 (4), 1039-1049, 2004

Speers, A., Water and Cities – overcoming inertia and achieving a sustainable future, in Novotny, V. and Brown, P. (ed.) Cities of the Future: towards integrated sustainable water and landscape management, IWA Publishing, 2007.

Standard Methods for the Examination of Water and Wastewater, 18th edn, American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC, USA, 1992

WFD – Directive 2000/60/EC of the European Parliament of the council establishing a framework for the community action in the field of water Policy, 2000, accessed on 11/02/2007 at
http://ec.europa.eu/environment/water/water-framework/info/decision_en.htm

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Integrated modelling of water and nitrate fluxes in large catchment areas – assessing the needs and efficiencies of agro-environmental measures for selected groundwater bodies in Germany.

R. Kunkel[a], M. Eisele[b], B. Tetzlaff[a] and F. Wendland[a]

[a] *Research Centre Jülich, Institute of Chemistry and Dynamics of the Geosphere, Agrosphere Institute (ICG-4) D-52425 Jülich, Germany, r.kunkel@fz-juelich.de*

[b] *State Authority for Mining, Energy and Geology, Lower Saxony, Germany, Stilleweg 2, D-30655 Hannover, Germany, michael.eisele@lbeg.niedersachsen.de*

Abstract: Within the EU-Life project WAgriCo (Water Resources Management in Cooperation with Agriculture) nitrogen management options adapted to hydrological and agro-economic site properties are developed and implemented for three pilot areas in the Federal State of Lower Saxony using new participation approaches and technologies suitable for programmes of measures to reduce diffuse pollution from agriculture. As a target value for water protection measures a nitrate concentration in percolation water of 50 mg/l as an average for groundwater bodies and their hydrogeological subdivisions has been defined. An integrative emission model is used to simulate the interactions between agricultural practice, nitrogen surpluses and the nitrogen flow through the soil and aquifer to the outflow into surface waters. The actual nitrate concentrations in percolation water are calculated for the entire Federal State of Lower Saxony considering site-characteristics, N-surpluses, water balance and denitrification in the soil. The tolerable N-surpluses needed to meet the environmental target are quantified as averages for each of the hydrogeological subdivisions by "backward" calculation using this model system. The required reduction of N-surpluses can then be estimated by comparing the tolerable N-surpluses to the actual state of nitrogen emission. For the evaluation of the amount and efficiency of water protection measures, the required reduction of N-surpluses to accomplish the environmental target is quantified, using the current status as a reference.

Keywords: Catchment management, diffuse source pollution, mitigation methods, river basin management, Water Framework Directive

1. INTRODUCTION

The fundamental objective of the Water Framework Directive (WFD) of the European Union [European Parliament and Council of the European Union, 2000] is to attain a good status of water resources in the member states of the EU by 2015. For groundwater bodies, whose good status can not be guaranteed by 2015, catchment wide operational plans and measurement programs have to be drafted and implemented until 2009. In the framework of the EU-LIFE project WAgriCo (Water Resources Management in Cooperation with Agriculture), a combined agroeconomic-hydrologic emission model is used

- to predict nitrate concentrations in the leachate at the scale of large catchments,
- to identify the maximal permissible diffuse nitrogen loads to guarantee nitrate concentrations in groundwater below 50 mg/l,
- to delineate priority areas for implementing nitrogen reduction measures at the scale of large catchments.

For macroscale areas, i.e. large river basins or Federal States, the achievement of good qualitative status of groundwater bodies entails a particular challenge as the complex ecological, hydrological, hydrogeological and agro-economic relationships in a catchment area have to be considered simultaneously. In this framework combined agroeconomic-hydrologic models, that can be applied for macroscale areas are powerful tools to analyse the actual pollution loads and "hot spot" areas and to predict the impact of reduction measures. The use of hydrological models to support the implementation of the WFD in larger river basins has been promoted for several years, and has lead to a clear improvement of the models being used [Lindenschmidt *et al.*, 2007]. Parallel to this, economic approaches for the support of the WFD have also been developed. These consider mainly the macro-economic effects of the improvement of water quality under the conditions of the WFD [Brouwer *et al.*, 2005; Mysiak and Siegel, 2004; Pulido *et al.*, 2005].

The development of model tools aiming at the derivation of measures to support the implementation of WFD is still the weakest area of the modelling [Lindenschmidt *et al.*, 2007]. First steps in this direction have been done already by the authors in a project conducted in the river Ems basin and sub catchments of the river Rhine basin. In this project a combined emission model was developed by coupling results from N-balancing using data from agricultural statistics with the hydrological model GROWA [Kunkel and Wendland, 2002], the DENUZ model approach [Kunkel *et al.*, 2004] for assessing denitrification rates in the soil and the reactive nutrient transport model WEKU for groundwater [Kunkel and Wendland, 1997].

2. STUDY REGION

Three pilot areas, Lager Hase, Große Aue and Ilmenau/Jeetzel, have been selected (see figure 1). The pilot areas are located in the North German Lowland within different large European river basins and cover about approximately 300000 hectares of agricultural land. The aquifers consist mainly of Pleistocene sand and gravel deposits. According to the first review of the qualitative and quantitative status of the groundwater bodies in Lower Saxony, for all of the three pilot areas the achievement of the good status is unclear or rather unlikely. Table 1 gives an overview about the agricultural structure in the pilot areas.

The size of the pilot areas is in a range of 1500 to 2000 km^2. The land use structure, however, is very different. Lager Hase and Große Aue are dominated by animal husbandry. In the Lager Hase pilot area the average size of the area per farm and the area of the adjacent arable land is smaller than in the Große Aue pilot area. The Ilmenau/Jeetzel pilot area is different to the other two pilot areas, since it is dominated by arable land and not by animal husbandry. In all pilot areas 55 % or more of the land surface is used agriculturally. Therefore, the natural conditions in groundwater and surface waters are significantly influenced by anthropogenic interferences into N- and water balance and runoff regimes.

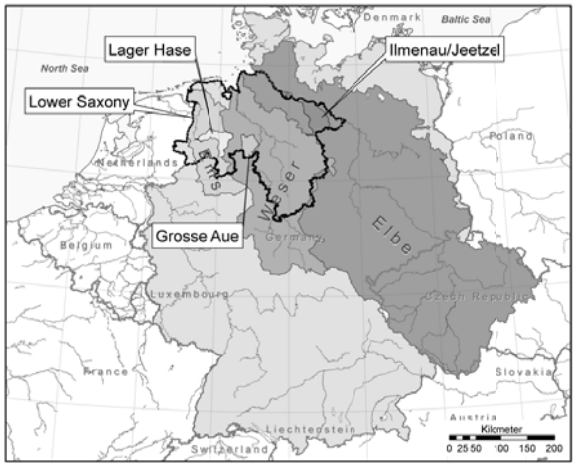


Figure 1. Location of the pilot areas of the WAgriCo project.

Table 1. Agricultural structure in the pilot areas in Lower Saxony.

Pilot area	Lager Hase	Große Aue	Ilmenau/Jeetzel
Area (km^2)	1420	1517	2052
Number of farms	3000	1620	1640
Animal husbandry	67 %	67 %	27 %
Arable	14 %	33 %	68 %
Other	19 %	0 %	5 %

3. THE EMISSION MODEL

The components of the combined emission model are using different regional resolutions: raster cells in the hydrological models and administrative units in the agro-economic model. This is due to the different data sources: while the hydrological models GROWA and DENUZ use (digital) maps to derive spatial inputs, the agroeconomic model employs agrarian statistical data [Schmidt *et al.*, 2007]. For this reason, regional nitrogen balances calculated by the N balance model as averages for the agricultural areas on a community level cannot be directly used as input variables in the hydrologic models.

For this purpose GIS supported model interfaces has been developed which enable not only the disaggregation and geographical referencing of N surpluses, but also the exchange of data, parameters and results between the models [Gömann *et al.*, 2005]. The process of adjusting the different spatial resolution of the models is based on commonly used land cover data, which enables a land cover classification into arable land, pasture, forests and urban areas. A lateral resolution of the model was set to 50x50 m^2, thus reflecting on the spatial differentiation of the land use data set. For the climate data, the time period of 1961-1990 have been used as a temporal reference period.

3.1 Regionalised differentiated agricultural economic modelling

With respect to diffuse water pollution the indicator “nitrogen surplus” is of particular importance. Agricultural statistics with data, e.g. on crop yields, livestock farming and land use, were used to balance the nitrogen supplies and extractions for the agricultural area. Nitrogen supply from manure is derived from nitrogen contents of the excrements of farm animals. The N balancing model differentiates between several processes of manure and its application, e.g. dung and liquid manure from cattle, hogs and poultry. Coefficients representing nutrient contents in manure, as well as utilization factors of plants, are taken from the literature and are also provided by experts of Federal Ministry of Food, Agriculture and Consumer Protection. As a rule, the difference between nitrogen supplies, primarily by mineral fertilizers and farm manure, and nitrogen extractions, primarily by field crops, leads to a positive N-balance [Gömann *et al.*, 2003]. Thus, nitrogen surpluses represent a risk potential since they indicate the amount of nitrogen potentially leaching into groundwater and surface water. Starting from these agricultural nitrogen surpluses, hydrogeological modelling is required in order to get closer to the problem of diffuse nitrate pollution of surface waters and groundwater.

3.2 Modelling nitrate flow through the soils

Denitrification losses in the root zone are calculated using a Michaelis-Menten kinentics by the DENUZ model [Kunkel and Wendland, 2006] as a function of diffuse N-surplus, denitrification conditions and related maximal denitrification rates per year and the residence time of percolation water in the soil. The maximal denitrification rates have been assessed on the basis of observed denitrification rates in German soils [NLfB, 2005] according to their geological substrate, the influence of groundwater and perching water and the average residence time of perching water in the soil as differentiation criteria. In this way the soil types were assigned into four classes ranging from D_{max} = 13.5 $kg \cdot ha^{-1} \cdot a^{-1}$ to D_{max} = 250 $kg \cdot ha^{-1} \cdot a^{-1}$.

The coupling of the N-surpluses occurring in the soil after denitrification to hydrological input pathways is carried out based on the grid-based water balance model GROWA [Kunkel and Wendland, 2002], which has been developed to support practical water resources management issues of large river basins and has already been applied in different regions of different sizes with different perspectives [Bogena *et al.*, 2005; Kunkel *et al.*,

2005; Tetzlaff *et al.*, 2007; Wendland *et al.*, 2005; Wendland *et al.*, 2007; Wendland *et al.*, 2003]. It employs an empirical approach with a temporal resolution of one or more years. Annual averages of the main water balance components in mm/a have been quantified as a function of climate, soil, geology, topography and land use conditions. The calculated total runoff can be separated into direct runoff (i.e. surface runoff, interflow and drainage runoff) and groundwater runoff [Wendland *et al.*, 2007]. Percolation water rates are modelled as mean long-term averages by subtracting mean long-term surface runoff rates from total runoff. Surface runoff is modelled using an empirical formula [US Soil Conservation Service, 1972]. Based on the calculated percolation water rates, N surpluses and denitrification rates in the soil, the nitrate concentration in the percolation water can be calculated directly.

3.3 Input data

A prerequisite for the integrated modelling is the compilation, updating and harmonizing of a digital data basis for the study region. The agrarian statistical data necessary to run the N-balancing model were provided by the Federal Ministry of Food, Agriculture and Consumer Protection or taken from the literature. The input data for the GROWA and DENUZ model, i.e. data on climate, topography, soil cover, soil parameters, hydrogeological parameters, water quality and groundwater bodies, have been made available for the whole Federal State of Lower Saxony and the German Meteorological Service. Many of these parameters were derived from digital maps, whose scale ranged from 1:50,000-1:200,000 (see table 2).

Table 2. Input data for modelling

Data set		Scale
Land cover	Landcover categories	1 : 25.000
Agricultural production	animals, cultivation harvest mineral fertilizer	Agrarian statistical data
Climate	Summer precipitation levels winter precipitation levels Potential Evaporation	Interpolated point data
Topography	slope exposition	50 x 50 m² raster
Soil parameters	Plant-available water denitrification capacity of soils groundwater influence	1 : 50.000
Hydrogeology	hydrogeological units geological profiles hydraulic conductivity	1 : 200.000
Hydrodynamics	Depth to groundwater runoff in rivers river network, drainage systems	1 : 200.000 or point data
Hydrochemistry	Groundwater monitoring data	Point data

4. MODEL RESULTS

The Emission model has been set up and used to characterize the present status of the N-emissions to groundwater and surface waters for the three pilot areas of the WAgriCo project. All model results have been presented and discussed with the local stakeholders in the pilot areas and have been approved. Here, selected model results on nitrate leaching will be presented for the Große Aue pilot area, located in the Weser basin (see figure 1).

4.1 Potential nitrate concentrations in the leachate

Percolation water rates, nitrogen surpluses and nitrate degradation rates in the soil are used for the calculation of nitrate concentrations in the leachate. The results for the Große Aue pilot area are shown in figure 2. For the major part of the Pilot area high concentrations of 150 mg NO_3/l and more have been calculated. This is due to the intensive agricultural use, i.e. high N-surpluses of the soils. Especially in the sandy areas in the northern and southern parts of the pilot area significant denitrification in the soil is not possible, thus leading to high nitrate concentrations.

On the other hand, low nitrate concentrations in the leachate (below 50 mg NO_3/l) are calculated for the lowland and the peaty areas. In that case the N surpluses are smaller, on one hand, and denitrification in the soil is much more effective due to high groundwater tables and high organic N-content in the soil, on the other hand.

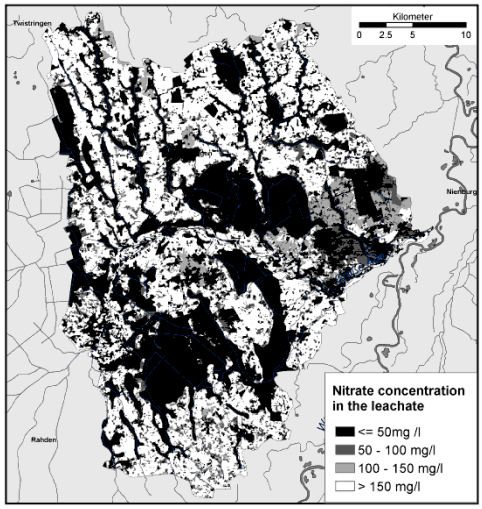


Figure 2. Calculated nitrate concentrations in the leachate for the Große Aue pilot area.

4.2 Assessing the required amount and efficiency of water protection measures

In the WAgriCo-project a mean long term nitrate concentration in percolation water of 50 mg/l was defined as a suitable environmental target for protecting groundwater against an exceeding of the EU quality standard for nitrate. This value is without any doubt appropriate for oxidized, i.e. not nitrate degrading, aquifers as it guarantees a nitrate concentration in groundwater below the EU quality standard for drinking water.

In reduced aquifers often low nitrate concentrations in groundwater have been observed, even in case of high inputs by percolation water. This is due to denitrification processes in the aquifer, taking place in the absence of oxygen and the presence of pyrite and/or organic carbon material. Most of the aquifers of the North German Lowland, where the pilot areas are located in, show a high denitrification potential [Wendland *et al.*, 2005].

This fact, however, should not be deceived into thinking that a possible denitrification capacity of groundwater is an argument to allow higher nitrate inputs into groundwater Denitrification in groundwater is associated with the irreversible consumption of substances in the groundwater, such as pyrite and organic carbon. Once these substances are exhausted nitrate can not be denitrified in groundwater any more. As a consequence, nitrate concentrations would start to rise, like it is described for many sites since many years [Rohmann and Sontheimer, 1985]. Consequently, the denitrification buffer of groundwater systems has to be prevented from damage, which implies a reduction of N-intakes into groundwater. A capable environmental target may be the nitrate concentration in the leachate. A limit of 50 mg NO_3/l would ensure a “good groundwater quality status” with respect to general quality standards even in the case of missing or exhausted nitrate degradation capacities in the aquifer.

Assuming the quantified percolation water levels and the nitrate degradation capacities of the soils to be constant, the nitrate concentration in percolation water depends exclusively on the nitrogen surplus level. Hence, by means of a backward calculation (inverse modelling), the maximal permissible N surplus for guaranteeing a mean nitrate concentration in percolation water of 50 mg NO_3/l has been calculated.

The according N-reduction level to reach 50 mg NO_3/l in the percolation water for every raster cell in the Große Aue pilot area is shown in the left part of figure 3. It can be seen

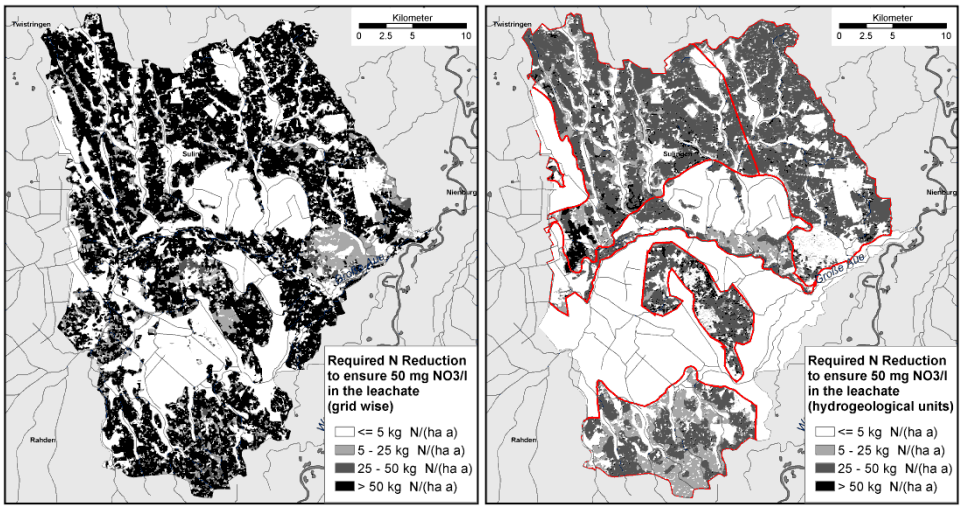


Figure 3. Required Reduction of N-surpluses to ensure a nitrate concentration in the leachate of 50 mg/l for the Große Aue pilot area with respect to individual grids (left part) or to on average for a subdivision (red lines).

directly that in many parts of the pilot area the amount of water protection measures needs to around 50 kg·ha^{-1}·a^{-1} or even more. This high reduction need, however, is a typical value for almost all intensively used agricultural areas in Northern Germany. Especially in regions with area-independent animal production the required N surplus reduction would be significant.

The Water framework directive is related to groundwater bodies and not to individual raster cells. Therefore, the influence of dilution areas with smaller nitrate concentrations in the leachate was considered by defining the individual target (50 mg NO_3/l in the leachate) not for individual raster cells, but for areas characterized by the groundwater bodies and the different hydrogeological settings within each groundwater body. In figure 3 (right part), the required N surplus reduction level is shown for the case that the average nitrate concentration in the leachate for each of the subdivisions needs to be 50 mg/l or less. It becomes clear directly that the overall reduction amount reduces significantly to roughly 25 kg·ha^{-1}·a^{-1} due to the effect of the dilution areas. Secondly, even in the case of local high inputs, there may be no reduction necessary to meet the environmental target since the average nitrate concentration in the leachate in the subdivision is below 50 mg/l.

Compared to the grid wise quantified N reduction levels, the required N reduction levels for subdivisions show the same hot spots areas, e.g. the Geest part in the northern part of the pilot area. Under deduction of compensation areas between the agriculturally used land within a groundwater body, the required degree of reduction is smaller.

5. APPLICATION TO LOWER SAXONY AND VALIDITY CHECK

Figure 4 shows the calculated reduction to ensure 50 mg NO_3/l in the leachate for the subdivisions in the groundwater bodies at risk for all three pilot areas of the WAgriCo project as well as for the whole Federal State of Lower Saxony. It can been seen that the required reduction of N surpluses to meet the environmental target varies significantly between the different regions of Lower Saxony. First of all, for the two pilot areas Lager Hase and Große Aue, both characterized by a high portion of animal husbandry, the required N-reduction still amounts to around 50 kg·ha^{-1}·a^{-1}. It can be doubted that a reduction of N surpluses in this extent can be achieved realistically by measures which still allow a cost-effective land-cultivation. In the Ilmenau-Jeetzel pilot area as well as for the whole north-eastern part of Lower Saxony, the necessary N reduction to meet the environmental target is much smaller. This can be attributed to the different type of land use, which is more coupled to arable land cultivation.

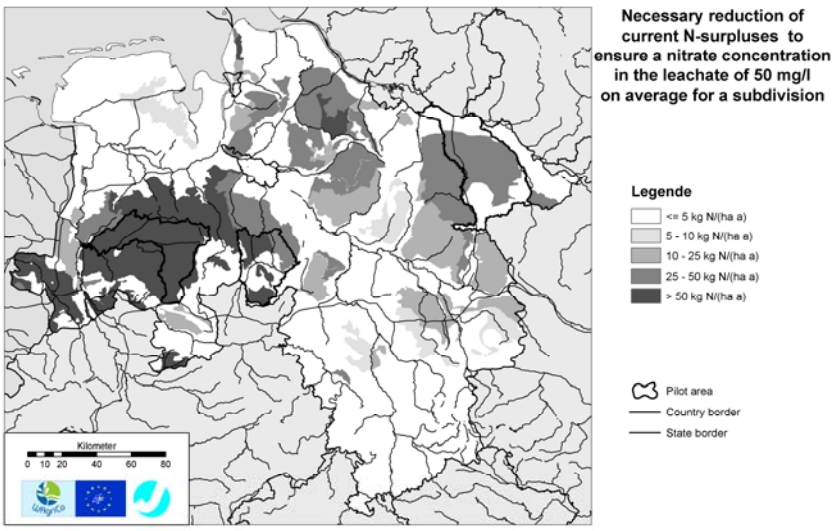


Figure 4. Required Reduction of N-surpluses to ensure a nitrate concentration in the leachate of 50 mg/l as an average for the subdivisions in the Federal State of Lower Saxony compared to the current status.

In order to use the emission model for water and nutrient management purposes, a validation of the model results is required. This is necessary both for the modelled percolation water rates and for the calculated nitrate concentrations in the leachate. The pilot areas have been selected on the basis of delineated groundwater bodies rather than for hydrological catchments. Therefore, a validation of the model results for the pilot areas itself is not possible. Instead, the model results have been checked for a number hydrological catchments located in the Federal State of Lower Saxony, which include the pilot areas.

For the nitrate concentrations in soil profiles, data from permanent soil observation sites (BDF sites), established in 1991 in the Federal State of Lower Saxony [Bartels *et al.*, 1991], could be used. In this monitoring network 14 long-term soil observation sites under agricultural land use have been established, in which the nitrate concentration in percolation water is recorded up to 24 times per year. Because the observed nitrate concentrations in the leachate vary not only with time but also with the sampling depth, a temporal and spatial averaging of the observed data in the range of depth 1.0-1.4 m below surface was necessary. The comparison of the calculated and observed values, shows a quite good agreement, which is indicated by a correlation coefficient of 0.79. Differences between individual values at some sites can be addressed to several reasons, e.g. the strong depth dependency of the nitrate concentrations, regionalization effects of the input data, errors attributed to the temporal variability and model uncertainties. The agreement of the calculated runoff rates and nitrate concentrations in the leachate allows the conclusion that the model, which is designed to be a small-scale model applicable to large areas represents the real situation sufficiently.

6. SUMMARY

A combined agroeconomic-hydrologic emission model has been used to predict nitrate concentrations in the leachate at the meso- and macroscale with the example of three mesoscale pilot areas and the Federal State of Lower Saxony. Model results showed considerable regional differences. Consistently, low nitrate concentrations in the leachate were modelled for areas displaying good denitrification capacities and long residence times of percolation water in the soil. Especially in the sandy Geest areas, which are used area-independent by animal husbandry especially in the western part of Lower Saxony, the nitrate concentrations are consistently more than 100 mg l^{-1} over the entire area. The high nitrate concentrations are a consequence above all of the high N surpluses from agriculture and the relatively low denitrification capacity of the soil.

Due to the manifold of possible factors, which influence the nitrate concentration in the leachate directly or indirectly, a comparison of modelled mean long-term nitrate concentrations in percolation water against averaged random nitrate concentrations observed in monitoring stations at a fixed sampling depth and at a fixed time is only possible under consideration of several constraints. Thus, the mean deviation of r^2= 0.79 determined for the BDF-sites can be regarded as a good correlation between modelled and measured concentrations. In any case, the modelled values should not be regarded as fixed values, which trace the specific nitrate concentrations at certain sites and at certain times, but as reference values, which represent mean long-term conditions and comprise regional blurring.

Modelled nitrate concentrations in the percolation water have been used to identify the maximal permissible diffuse nitrogen loads to guarantee nitrate concentrations in groundwater below 50 mg/l, i.e. the drinking water standard for nitrate in groundwater. In case nitrate concentration in percolation water exceeds 50 mg/l, the maximal permissible nitrogen surplus levels in agriculture to guarantee a mean long term nitrate concentration in percolation water below 50 mg/l has been identified based on a backward modelling approach. This has been done for different geographical reference areas, i.e. for individual grids and on the level of groundwater bodies.

Especially in the western part of Lower Saxony which is dominated by intensive animal husbandry, the necessary N-reduction to guarantee 50 mg NO_3/l in the leachate would be significant (> 50 $kg \cdot ha^{-1} \cdot a^{-1}$). In the eastern part of Lower Saxony, a reduction of up to 25 $kg \cdot ha^{-1} \cdot a^{-1}$ would be sufficient to fulfil the environmental target. Consideration of dilution areas reduces the required reduction levels significantly.

As the discussions on environmental targets proceed, it may be necessary to modify the trigger value of 50 mg NO_3/l in percolation water depending on the agricultural and hydrological site conditions in the groundwater bodies. In general however, this procedure is perceived to be particularly innovative since the political relevance of conclusions of this type of approach is gaining importance and the accuracy of recommended environmental policy instruments is improving.

7. REFERENCES

Bartels, R., E. Pluquet and H. Sponagl, Boden-Dauerbeobachtungsflächen im Niedersächsischen Bodeninformationssystem. *Geologisches Jahrbuch, A126*, 107-116, 1991.

Bogena, H., R. Kunkel, T. Schobel, H. P. Schrey and E. Wendland, Distributed modeling of groundwater recharge at the macroscale. *Ecological Modelling, 187* (1), 15-26, 2005.

Brouwer, R., M. Hofkes and V. Linderhof, A computable general equilibrium approach to the estimation of the direct and indirect economic costs of water quality improvement in the WFD. Proceedings of the International workshop Integrated River Basin Modelling and the EU Water Framework Directive,Amsterdam, The Netherlands, November 17-18, 2005, Vrije Universiteit, 2005.

European Parliament and Council of the European Union (2000): Directive 2000/60/EC of the European Parliament and the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. , edited, Official Journal of the European Communities.

Gömann, H., P. Kreins, R. Kunkel and F. Wendland, Koppelung agrarökonomischer und hydrologischer Modelle. *Agrarwirtschaft, 52* (4), 195-203, 2003.

Gömann, H., P. Kreins, R. Kunkel and F. Wendland, Model based impact analysis of policy options aiming at reducing diffuse pollution by agriculture - a case study for the river Ems and a sub-catchment of the Rhine. *Environmental Modelling & Software, 20* (2), 261-271, 2005.

Kunkel, R., M. Bach, H. Behrendt and F. Wendland, Groundwater-borne nitrate intakes into surface waters in Germany. *Water Science and Technology, 49* (3), 11-19, 2004.

Kunkel, R., H. Bogena, H. Goemann, P. Kreins and F. Wendland, Management of regional German river catchments (REGFLUD)impact of nitrogen reduction measures on the

nitrogen load in the River Ems and the River Rhine. *Water Science and Technology*, *51* (3-4), 291-299, 2005.

Kunkel, R. and F. Wendland, WEKU - A GIS-supported stochastic model of groundwater residence times in upper aquifers for the supraregional groundwater management. *Environmental Geology*, *30* (1-2), 1-9, 1997.

Kunkel, R. and F. Wendland, The GROWA98 model for water balance analysis in large river basins - the river Elbe case study. *Journal of Hydrology*, *259* (1-4), 152-162, 2002.

Kunkel, R. and F. Wendland, *Diffuse Nitrateinträge in die Grund- und Oberflächengewässer von Rhein und Ems*. Schriften des Forschungszentrums Jülich, Reihe Umwelt/Environment, Vol 62. Forschungszentrum Jülich GmbH, Jülich, Germany, 2006.

Lindenschmidt, K.-E., F. Hattermann, V. Mohaupt, B. Merz, Z. W. Kundzewicz and A. Bronstert, Large-scale hydrological modelling and the Water Framework Directive and Floods Directive of the European Union. Proceedings of the 10th Workshop on Large-Scale Hydrological Modelling, Advances in Geosciences, 11, pp. 1-6, 2007.

Mysiak, J. and K. Siegel (2004): Uncertainty in the WFD: Implications for Economic Analysis., DOI: http://en.scientificcommons.org/727520.

NLfB, Grundwasser: Methodenbeschreibung, EG-WRRL Bericht, Niedersächsisches Landesamt für Bodenforschung, Niedersächsisches Landesamt für Ökologie, Bezirksregierung Hannover, Hannover, Germany, 2005.

Pulido, M., J. Andreu, A. Solara and M. A. Pérez, Economic valuation and optimization of water systems management. From microeconomic data to basin scale results. A computable general equilibrium approach to the estimation of the direct and indirect economic costs of water quality improvement in the WFD. Proceedings of the International workshop Integrated River Basin Modelling and the EU Water Framework Directive,Amsterdam, The Netherlands, November 17-18, 2005, Vrije Universiteit Amsterdam. , 2005.

Rohmann, U. and H. Sontheimer, *Nitrat im Grundwasser: Ursachen, Bedeutung, Lösungswege*. DVGW-Forschungsstelle am Engler-Bunte-Institut der Universität Karlsruhe (TH), Karlsruhe, Germany, 1985.

Schmidt, T. G., B. Osterburg and T. Runge, A new strategy for agri-environmental measures to reduce N-losses more effectively. Proceedings of the 35. IAH Congress Groundwater and Ecosystems

Lisbon, Portugal, 17.-21. September 2007, Internatinal Association of Hydrogeologists, 2007.

Tetzlaff, B., R. Kunkel, W. Schäfer and F. Wendland, Determination and ranking of target areas in catchments for the implementation of nitrogen reduction measures. *Advances in Geosciences*, *11*, 69-75, 2007.

US Soil Conservation Service, *National Engineering Handbook: Chapter 4: Hydrology*. 2nd Reprint ed. U.S. Department of Agriculture, Washington, D.C., USA, 1972.

Wendland, F., H. Bogena, H. Goemann, J. F. Hake, P. Kreins and R. Kunkel, Impact of nitrogen reduction measures on the nitrogen loads of the river Ems and Rhine (Germany). *Physics and Chemistry of the Earth*, *30* (8-10), 527-541, 2005.

Wendland, F., R. Kunkel, H. Gömann and P. Kreins, Water fluxes and diffuse nitrate pollution at the river basin scale: Interfaces for the coupling of agroeconomical models with hydrological approaches. *Water Science and Technology*, *55* (3), 133-142, 2007.

Wendland, F., R. Kunkel, B. Tetzlaff and G. Dorhofer, GIS-based determination of the mean long-term groundwater recharge in Lower Saxony. *Environmental Geology*, *45* (2), 273-278, 2003.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

The Role of Groundwater Resources Models in Integrated Catchment Management

Farrell. R[1] & Whiteman. M[1]
[1]Environment Agency for England and Wales. rolf.farrell@environment-agency.gov.uk, mark.whiteman@environment-agency.gov.uk

Abstract: Innovative work has been undertaken by the Environment Agency for England and Wales, and will continue to be required to successfully meet the challenges presented by recent domestic legislation (Water Act 2003) and the EU Water Framework Directive. This has required a more integrated approach to catchment management of groundwater and surface water quantity and quality to protect ecological receptors. We have made a significant investment in tools, methods and approaches to ensure proper integrated management of groundwater. This paper describes the water resources aspects, however, there are also equivalent methods for contaminant hydrogeology.
The Environment Agency for England and Wales has made a significant investment in catchment-scale groundwater modelling for the purpose of assessing the condition and available resources of the countries' aquifer systems. It is important that the benefits of this substantial investment are maximised. Principal drivers for groundwater resource modelling include assessment of available groundwater resources for abstraction, assessment of baseflow to protect surface waters, determination of water balances for wetlands and the modelling of river augmentation schemes. Hydrogeologists are directly concerned with the impacts of new abstractions on existing sources; Water resources managers need to know how the groundwater regime impacts on surface waters; hydrologists need baseflow inputs to surface water models; and hydroecologists need knowledge of how the ecology might respond to stresses from the groundwater system during critical periods (e.g. droughts).
Keywords: groundwater; regulation; tools; guidance; benefits

1. INTRODUCTION

The Environment Agency for England and Wales has a statutory duty to manage the sustainable development of groundwater resources. Our role is to identify the sustainable yield of groundwater units, and to regulate abstraction to ensure that the impacts of abstraction on springs, rivers and wetlands are limited to a point which is acceptable [Burgess, 2002]. The Environment Agency also has a duty to protect groundwater quality. It is in line with Environment Agency policy that the management and regulation of groundwater should be based on a good scientific understanding of groundwater systems, and on good technical practice. Integrated catchment management, involving consideration of all aspects of the system rather than artificial focus on specific aspects of it such as surface water hydrology and water quality, is an important part of this. This paper describes the tools, methods and approaches used to deal with the groundwater resources component of integrated catchment management – equivalent methods are available for contaminant hydrogeology, although the latter is outside the scope of this paper . Integrated management of land and water requires that we consider the integration of data, understanding and decision making so that adopted remedial measures in one sector (such as risk) do not have negative effects in other sectors (such as fisheries). The UK Government is moving towards an ecosystems approach [Defra 2007] to conserving, managing and enhancing the natural environment in England. This is about adopting a new way of thinking and working, by:

• Shifting the focus of our policy-making and delivery away from looking at natural environment policies in separate 'silos' – e.g. air, water, soil, biodiversity – and towards a more holistic or integrated approach based on whole ecosystems; and

• Seeking to ensure that the value of ecosystem services is fully reflected in policy- and decision-making across Government at all levels.

The modelling of the major aquifers of England and Wales represents a significant investment for the Environment Agency. Catchment scale models, as they attempt to model a large part of the water cycle in some detail, tend to have potential benefits for a much broader cross section of the organisation's work than might at first be considered.

This paper summarises how the Environment Agency is attempting to ensure that these benefits are realised by means of both practical tools (such as the National Groundwater Modelling System (NGMS)) and Guidance documentation (such as guidance on Catchment Abstraction Management Strategies) and how we are using them to support integrated catchment management.

2. TOOLS

The Environment Agency has a number of different tools that can be used for assessing groundwater resources and contaminant hydrogeology. This paper will focus on the groundwater resources tools. These range from simple risk screening methods and GIS-based tools, through simple analytical models to more complex numerical models (Table 1).

Complexity/cost	**Water Resources**	**Groundwater Contamination**
Simple ↓	Conceptual modelling Guidance (Environment Agency, 2002)	Conceptual modelling (Environment Agency, 2001b)
	Simple risk screening tools e.g. Water Resources GIS risk engine	Groundwater vulnerability maps
	Simple/analytical (deterministic) models e.g. IGARF (Environment Agency, 2001a)	Analytical (deterministic) models e.g. Remedial Targets Method
	Analytical (probabilistic) code (rarely used)	Analytical (probabilistic) code e.g. ConSim
Complex	Numerical models e.g. MODFLOW	Numerical models (rarely used) e.g. MODFLOW+MT3D

Table 1: Parallel structures/guidance for groundwater resources and contaminant modelling components of integrated catchment management

2.1 GIS-based risk engine

To support implementation of the Water Framework Directive, the Environment Agency has developed a GIS-based risk engine for screening the level of risk from new water abstraction licences (Figure 1). The equivalent for groundwater contamination is the groundwater vulnerability map.

The GIS contains data sets of the different environmental classifications for rivers, lakes, groundwater etc along with supporting information (e.g. geology maps) that may help in assessing a site. The different layers allow a particular location to be assessed in terms of impact of, for example, a new abstraction on those features. It will be possible to export output from numerical groundwater models directly into the GIS allowing, for example, sample transient zones of influence to be viewed. Similarly, output from simple analytical models can also be viewed through the GIS. Also, it is intended that spatial plots from the numerical groundwater reports indicating zones of calibration accuracy of the models will be included to allow decisions to be made as to the best means to assess a catchments' water resources.

Operationally, the Water Resources GIS will be mostly used by water resource regulatory managers though support and datasets will also come from specialist hydrogeologists and hydrologists.

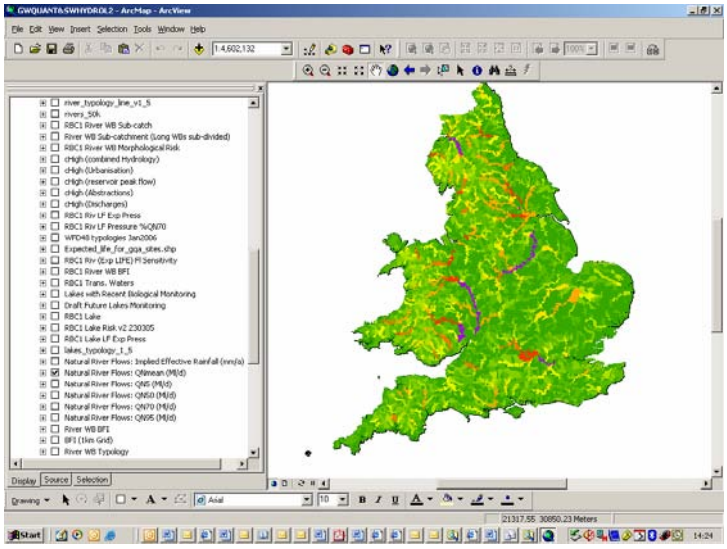


Figure 1 Water Resources GIS tool.

2.2 Simple Groundwater Modelling Tools

There is no automatic presumption that a numerical assessment of a catchment will require a full distributed groundwater model; various analytical tools have been used as alternatives. The decision as to whether a distributed model is required depends on assessment of what it is needed to know and why; this decision must be made carefully if benefits are to be maximised. The degree of risk to the environment and/or existing abstractors from the decision, as well as cost, available timescales and conceptual understanding (e.g. is the aquifer in an area of complex geology where development of a numerical model will be inappropriate?) all need to be taken into account. Generally, the analytical tools are used to answer specific, relatively localised issues.

The existing analytical tools, which include IGARF (Impacts of Groundwater Abstraction on River Flows)[Environment Agency, 2001a], a simple spreadsheet-based tool, and also traditional groundwater investigation formulae, are thought only to be adequate for more straightforward abstraction licence assessments.

Another commonly used approach is to use Catchmod, a simple rainfall-runoff model (with groundwater module). This was successfully used by the Environment Agency to provide timely information in response to the 2005-06 drought in England by assessing baseflow in rivers at critical surface water locations, and groundwater levels at key locations. The numerical groundwater models were also used to model the impact of the drought on the resource balance as a whole, something that Catchmod cannot do.

2.3 Numerical Distributed Groundwater Models

The key focus for groundwater resource assessment are the catchment scale groundwater models usually based on the Modflow code. Other, steady state models have been created using Flowpath for determining capture zones for groundwater protection. To maximise the achievable benefits from regional scale models, consideration of groundwater quality modelling is also necessary. A similar hierarchy of tools is available to staff focussing on Groundwater Quality issues (Table 1).

Recent IT developments have provided an exciting opportunity for the Environment Agency to develop new, user-friendly, graphical and GIS-based methods for presenting and using the catchment-scale groundwater models. The National Groundwater Modelling System (NGMS) [Farrell *et a*l, 2007] is a software front end (Figure 1), a development of the Delft Flood Early Warning System (FEWS) software (by Deltares, formerly Delft Hydraulics), already successfully proven by the Environment Agency's National Flood Forecasting System. NGMS allows configured models to be run either according to pre-defined, standard scenarios; historic (i.e. modelled observed), naturalised, fully licensed and recent actual as well as user defined what-if scenarios involving changes or additions to abstractions as well as predictive scenarios for climate change or drought prediction. From an end user perspective, the purpose of NGMS is to provide a means of running and examining model output (and input) enabling water management decisions to be made. Tools that support this include:

- Hydrographs for all time variant data types.
- Pre determined river accretion profiles for main rivers and watercourses
- Flow duration curves and supporting statistics
- Spatial displays for model input and output data plus aquifer parameters

The system is designed to enable easy comparison of data sets - either between different types of data or between different scenarios (for example, comparison of a historic model run and a naturalised run for a spatial plot effectively produces a drawdown plot). Whilst NGMS itself has varied functionality for post processing and display of data, export to other applications such as the MS Office suite and GIS programmes is straightforward.

As well as the ability to run the models, the NGMS allows archiving of model versions, model runs and supporting data.

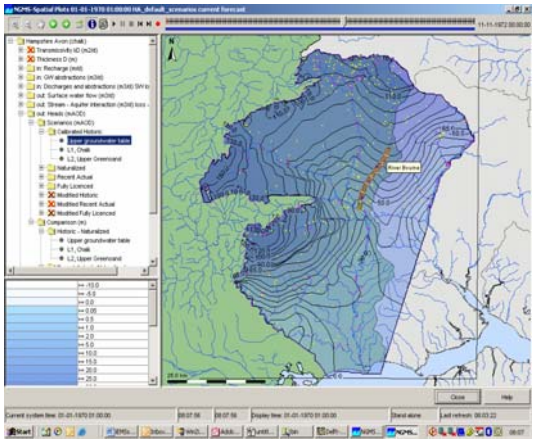


Figure 1 NGMS Groundwater Head Spatial Display

2.4 Use of Groundwater Models

The development of groundwater conceptual and numerical models is an important part of helping to meet the Environment Agency's regulatory responsibilities, including strategic planning and operational management of groundwater resources in England and Wales. Strategic planning comprises successful implementation of European Directives including the Habitats Directive (92/43 EEC) and Water Framework Directive (2000/60/EC), restoring sustainable abstraction by addressing hydrogeological impacts upon rivers and wetlands of conservation value, and preparation of Catchment Abstraction Management Strategies (CAMS). Operational management includes water resource abstraction licensing and water availability forecasts. Groundwater resource assessment also provides a

framework for groundwater quality investigations, including Source Protection Zones (SPZ), and for contaminated land work.

An example of operational use is assessment of impacts upon groundwater-dependent wetlands. Understanding the links between the ecological functioning of wetlands and changes in groundwater quantitative and chemical status remains a significant challenge to successful catchment management [Lerner et al, 2007]. Groundwater models provide a decision support tool to help answer questions faced by regulators and wetland managers, such as:

- Whether reduced groundwater discharge to a wetland as a result of abstraction has an adverse impact on the ecology (e.g. EU Habitats Directive).
- If there is an adverse impact, what options are effective to mitigate or remove the impact to ensure that favourable condition can be met (e.g. can reduced groundwater flow be compensated for by increased surface water supply, without significant damage to the ecosystem?).

The Environment Agency's eco-hydrological guidelines are used to determine the hydrological, chemical and management needs of the ecology [Wheeler et al, 2004]. The groundwater model is used to check if site-specific criteria are maintained including upflow to the topmost model layer, groundwater levels at the water table, spring flows and soil moisture characteristics [Whiteman et al, 2007]. A series of options for removing adverse impacts are then tested using the groundwater models, resulting in a Site Action Plan to restore the site to favourable condition.

At Foulden Common, in Norfolk (Eastern England), several wetland water supply mechanisms (WETMECS) [Wheeler and Shaw, in press] were identified [Whiteman *et al* 2007] and an eco-hydrological conceptual model for the site was developed to include these water supply mechanisms [Entec, 2007]. This helped guide the specification of local improvements within a regional scale groundwater model including, albeit relatively crudely, representation of 'top-layer' effects.

The *Schoenus nigricans–Juncus subnodulosus* mire (M13) vegetation community (National Vegetation Classification, Rodwell 1991 – 1995) represents the ecological feature of the site which is most sensitive to changes in the hydrology (e.g. to near surface groundwater levels and "flushing"). The criteria above were developed for this feature to help determine acceptable levels of abstraction on the basis that, if M13 were adequately protected, less sensitive communities would not be adversely affected.

Whiteman et al [2007] discuss the limitations of applying numerical groundwater models, which can only represent appropriate conceptual detail for the scale at which they are constructed. They conclude that existing relatively coarse scale models may be used, particularly to assess potential changes in hydrogeological behaviour, but that their application needs considerable care. Results from such models must not be used 'blind', and in particular should be interpreted with due regard to small scale variability.

3. GUIDANCE AND DOCUMENTATION

To ensure that the above tools are used by the right people at the right time, clear guidance and documentation is as necessary as the models and tools themselves. Examples of important Environment Agency guidance which are used to underpin management of groundwater resources include:

- Guidance for Catchment Abstraction Management Strategies (CAMS) - the principal aim of CAMS is to provide a framework for resource availability assessment and produce a licensing strategy which aids the sustainable management of water resources on a catchment scale, and each strategy is reviewed on a six-yearly cycle. CAMS is the first consistent methodology that has given us resource availability for the whole of England and Wales.

Groundwater modelling was used in the first cycle of CAMS in groundwater dominant areas, to assess the available groundwater resources. The second cycle of CAMS will see each catchment revisited and updated with the aim of improving further our confidence in the resource assessments.

The new, second cycle CAMS guidance and the Groundwater Modelling Guidance Notes will cross reference each other to ensure that the most appropriate groundwater modelling tools are used for assessing groundwater resources at any point of interest within a modelled catchment. This is in relation to both riverine and wetland ecological needs. In the future there will be the presumption to always use a groundwater model if there is one available and appropriate to the task.

- Groundwater Modelling Guidance Notes [Environment Agency 2002] have been developed to ensure that project managers are able to produce the right product at the right time. These notes are currently being updated to describe the use of models to address catchment management issues, based on recent experience. Supporting technical notes provide background aided by examples of good practice. .
- Regional Groundwater Modelling Strategies provide basic descriptions of the Region's major aquifers, how they will be modelled and what the drivers for those models are [Hulme *et al*, 2002]. They describe the principle water resource management issues that affect the main aquifer units. They then describe the current and proposed groundwater models and how they are expected to be used to address these issues. Supporting documents are updated annually to report back on progress against the strategy.
- Hosting of hyperlinked model documentation within the National Groundwater Modelling System allows people using the system to have immediate access to conceptual and technical information whilst reviewing the output of the models. Documents hosted on the system include conceptual reports, project reports, technical and non technical summaries and supporting plots and diagrams that may help in interpreting model output. Hyperlinks from NGMS into the model reports and within the documentation mean that the meaning of a model output can assessed quickly. For example, a scenario flow at a gauging station can be related back to documentation that describes the calibration accuracy and purpose of that gauging station.

4. FUTURE DEVELOPMENTS

4.1 Forecasting

Work is ongoing to ensure that groundwater models are kept sufficiently up to date for operational use (for example in drought scenario modelling):

- Planned data update programmes ensure that the models should never need more than 2 years worth of environmental data to be updated at any one time. This ensures that the task of updating shouldn't be too onerous. For models configured to NGMS, the update process will be largely standardised and the process supported by guidance documentation.
- Forecast scenarios have been developed. In 2006, scenarios based on specific months' rainfall patterns that represented 60, 80, 90, 120 and 140% of LTA were used to give a feel for the sensitivity of the system to certain quantities of rainfall. These scenarios were used with simple Catchmod model as well as the catchment-scale numerical groundwater models (figure 3)
- A rapid method of providing more immediate forecasts is needed; ideally such that a feel for how a drought might develop can be given within a day. One approach would be to make use of rainfall radar data that is gathered by the Agency National Flood Forecasting System (also based on Deltares FEWS). This can be configured to provide geostatistically corrected daily rainfall datasets which could be used to update the rainfall sequence, possibly calibrated against the telemetered rainfall observed data also used by NFFS. Combined with a simple utility to update abstraction records (perhaps by using the same rates as were abstracted in a previous year) it should be possible to provide a preliminary set of predictive scenarios within a few hours.
- Use of continuous monitoring (i.e data loggers) rather than periodic manual dipping, in observation boreholes ensures that we have real-time data that highlights the need for a

model or corrective action to be taken, and provides sufficient data resolution for short term event modelling as well as more longer-term resource assessments.

As well as providing the physical tools to enable forecasting to be carried out quickly, the appropriate guidance is being developed to support it as part of the guidance notes update. Examples of best practice, from the drought, as well as lessons learned will be incorporated so that we should be ready to provide timely answers when the climate next demands it.

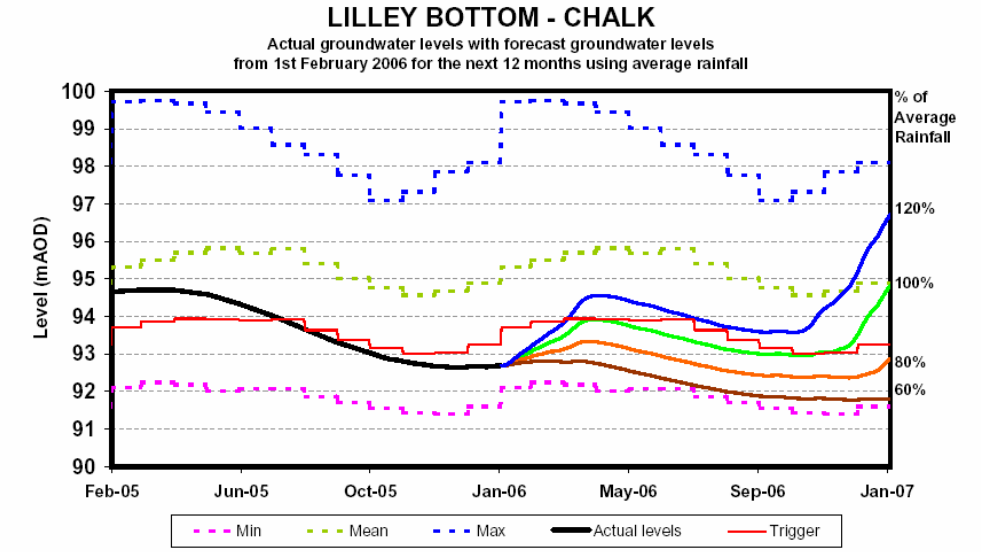


Figure 3 Forecast groundwater levels in EA Thames Region, 2006

4.2 'Intermediate' models

Jackson *et al [2007*] caution that the current analytical models do not produce results with sufficient confidence where, for example, a groundwater abstraction has a significant impact on more than two rivers, or a more accurate spatial or seasonal distribution of the impacts on both flow and groundwater level is important. For example, when the effects of several abstractions must be considered, and the regional context is important for estimating the local impact, which is often true for wetland sites.

Traditional catchment-scale Modflow models are inappropriate for the rapid assessment of these impacts since they often take two to three years to produce at a cost of £200k to £300k each. To address this need for "intermediate" models, the Environment Agency recently completed a proof of concept project for an intermediate tool which removes some of the limiting assumptions associated with the current low cost modelling tools (e.g. aquifer homogeneity or geometric simplicity) but which can be set up in a few hours [Jackson *et al, 2007*]. In future, this object-oriented tool should allow operational hydrogeologists to test many alternative scenarios quickly to get a better understanding of system behaviour, and to test the uncertainty associated with the key hydrogeological conditions which prevail for a particular site.

5. CONCLUSIONS

The Environment Agency for England and Wales has responded proactively to the challenges presented by recent legislation by taking a much more integrated approach to management of groundwater, where the primary driver for a new groundwater model is as likely to be to determine baseflow targets to meet in-river ecological needs as specifically groundwater resource related. It could also equally likely require velocity fields, not just bulk water flow, to address groundwater quality/contaminant hydrogeology issues.

This integrated approach also requires practices to be agreed between different disciplines which might have either input into models or have a requirement for their output.

Considerable progress has been made towards effective, integrated catchment management, by making both numerical tools and guidance easily available. The guidance must take into account the different needs and benefits that the models may provide answers for. Further work is still required to meet the challenges of integrated catchment management, with a need for not only integrated models but inclusion of interdisciplinary processes (e.g. the hyporheic zone).

REFERENCES

Burgess, D.B., Groundwater resource management in eastern England: a quest for environmentally sustainable development. In: Hiscock, K. M., Rivett, M. O. & Davison, R. M. (eds) *Sustainable Groundwater Development*. Geological Society, London, Special Publications, 193, 53-62, 2002.

Defra, Securing a healthy natural environment: An Action Plan for Embedding an Ecosystems Approach, Defra 2007.

Entec, Foulden Common site options plan – report to Environment Agency, Anglian region, Peterborough, United Kingdom, 2007.

Environment Agency, *Impact of groundwater abstractions on river flows*. Environment Agency Project Report NC/00/28, 2001a.

Environment Agency, *Guide to good practice conceptual models for contaminant transport in the subsurface.* Environment Agency Report NC/99/38/2, 2001b.

Environment Agency, *Groundwater resources modelling: Guidance notes and template project brief (Version 1).* Environment Agency R&D Guidance Notes W213, 2002.

Farrell, R., Whiteman, M., Gijsbers. P., Credibility in groundwater modelling: A national groundwater modelling system for England and Wales, *ModelCARE2007 Sixth International Conference on Calibration and Reliability in Groundwater Modelling Pre-published Proceedings* 1, 166-171, 9-13 September 2007.

Hulme, P., Fletcher, S., and Brown, L., Incorporation of groundwater modelling in the sustainable management of groundwater resources. In: Hiscock, K.M., Rivett, M.O., & Davison, R.M. (eds) *Sustainable Groundwater Development*. Geological Society, London, Special Publications, 193, 83-90, 2002.

C R Jackson, C.R., Mansour, M.M., Hughes, A.G. and Hulme, P.J. *Numerical modelling of the impact of groundwater abstraction on river flows* Environment Agency R&D Project: SC-30233 Model for Investigating the Impacts of Groundwater Abstraction, 2007.

Lerner, D.N., Harris, B. and Surridge, B., The role of groundwater in catchment management, *Groundwater and Ecosystems (Eds. Ribeiro. L., Chambel. A. and Condesso de Melo. M.T.). Proceedings of XXXV IAH Congress, Lisbon,* 17-21 September 2007.

Rodwell, J. S. (ed)(1991 - 1995) *British Plant Communities.* Volumes 1 – 5. Cambridge University Press, Cambridge.

Whiteman. M.I., Wheeler, B.D., Shaw. S.C., Lewis. R.T., Grout. M. and Charles. K., Eco-hydrological requirements of groundwater dependant ecosystems in England and Wales, *Groundwater and Ecosystems (Eds. Ribeiro. L., Chambel. A. and Condesso de Melo. M.T.). Proceedings of XXXV IAH Congress, Lisbon,* 17-21 September 2007.

Wheeler, B. D. and Shaw, S. C. *A Wetland Framework for Impact Assessment at Statutory sites in Eastern England.* Environment Agency R&D Technical Report (in press)

Wheeler. B.D., Gowing. D.J.G., Shaw. S.C., Mountford. J.O. and Money. R. *Eco-hydrological Guidelines for Lowland Wetland Plant Communities* (Eds. Brooks. A.W., Jose. P.V. and Whiteman. M.I.). Environment Agency (Anglian Region), 2004

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Model based sediment quality management on river basin scale

Bernhard Westrich

Institute of Hydraulic Engineering, University of Stuttgart, Pfaffenwaldring 61, 70550 Stuttgart, Germany (bernhard.westrich@iws.uni-stuttgart.de)

Abstract:

Contaminated sediment mobility plays an important role in river engineering and water resources management. Numerical models are powerful tools to be used for analysis, prediction and remediation issues in sediment management. Erosion stability plays a key role for contaminated sediment mobility assessment. Experimental investigations on erosion stability of undisturbed sediments reveal a large spatial variability of sediment properties. A 1-d and a 2-d numerical transport model is applied to analyse erosion and sedimentation processes of contaminated sediments in the Upper River Rhine reservoirs and in the river Elbe focussing on the effect of the near bank low flowing water zones.

Keywords: sediment, erosion, contaminants, samples, modelling.

1. INTRODUCTION

Sediments play an important role in river engineering, water quality management and river rehabilitation. In the past, rivers have been regulated for navigation, hydropower and flood protection by building different structures such as weirs, dams and river training works which have great influence on the sediment transport behaviour. In industrialized countries, many river sediments are contaminated by heavy metals and organic substances due to municipal and industrial waste water discharge [Förstner et al., 2005]. Contaminated sediment deposits can be found in low flowing or stagnant adjacent water bodies such as near bank so called groyne fields, harbours, dead arms, flood plains and retention reservoirs as well. Most deposits are subject to resuspension and long distance transport by erosive flood events with wide spread impact on the aquatic ecosystem. Therefore, a risk based sediment management is required which accounts for the probability of contaminant mobilization by erosive discharge events and the impact on the environment.

Because of the complex interaction of hydrological, hydrodynamic, sedimentological and biochemical processes an interdisciplinary approach on a river basin scale is required to provide basic information for a sustainable sediment management including remediation measures and decision support [Westrich and Förstner, 2007].

2. INTEGRATED RISK ASSESSMENT

Referring to the EU-Water Frame Work Directive a sustainable sediment management is required to improve the ecological status of surface water bodies by reducing the risk of adverse impact and ecological damage due to toxicants. A comprehensive risk assessment should be based on an interdisciplinary approach to cope with various processes on different space and time scale. Management strategies must provide solutions on local,

regional and river basin scale. Since each catchment has its characteristic pattern in terms of water use and pollution the remediation concept must be site and catchment specific.

Floods play a dominant role in sediment erosion risk assessment because of their extreme erosion capacity followed by large scale dispersion and uncontrolled immission. Hence, older most highly contaminated sediments in deeper layers can be resuspended and redeposited in yet uncontaminated water bodies or on flood plains. In this context the sediment erosion stability turns out a key factor as it controls the mobility and contaminant mass flux and hence, allows to estimate the total amount of released contaminants by flood events. Erosion probability can be determined by combining the statistics of the hydrodynamic bed shear stress based on discharge hydrology and the sediment specific erosion resistance.

Resuspended fine sediments are often transported over a long distance through the river network while simultaneously different interacting processes are going on such as: mixing with bed material; dilution or additional loading by tributaries, point sources or erosion; fractional sedimentation; ad-/ desorption of pollutants; repartitioning of pollutants and, finally biochemical transformation and degradation.

3. TRANSPORT AND EXPOSURE MODELLING

Numerical models allow to integrate processes of different scale in space and time. Physically based numerical models are useful tools to describe the pathway and fate of contaminants in rivers and to quantify the relationship between emission and immission. Moreover, predictive numerical models can also be used for a risk reduction analysis by investigating the effect of alternative remediation measures and providing basic information for a cost-benefit analysis. Hydrological scenario modelling produces data on intensity and duration of immission which can be used for further statistical modelling of a dose-effect relationship. Consistent field data of extreme events are very poor and hence, contaminant transport model calibration and validation are subject to high uncertainty However, neglecting biochemical transformation and degradation processes an engineering approach with conservative assumptions can be made for a preliminary study focussing on hydrological and hydraulic issues related to erosion, transport and deposition of particulate contaminants

To describe long-term and large scale transport processes one-dimensional models have been proven useful. A 1-d multi-strip model was developed aiming to describe the erosion, dispersion and deposition dynamics of fine suspended sediments for regulated rivers with typical training works such as near bank groyne fields. The total river cross section is subdivided into three compartments: main channel, adjacent groyne fields on the left and right bank and flood plains, respectively [Prohaska and Westrich, 2006]. The set of 1-d transport equations for the strips are coupled by dispersive and advective exchange terms. For water levels below the crest of the structure the water body is considered a dead water zone with homogeneous suspended sediment concentration. Because of the complex flow and transport processes inside the groyne area the exchange coefficients are measured by laboratory experiments, confirmed by a 2-d flow and transport model [Jacoub, 2004] and implemented into the 1-d model strip model as an engineering approach.

4. SEDIMENT DATA

Sediment input data have different quality in terms of accuracy and density in space and time. To reduce the model uncertainty due to great variability of sediment properties several samples must be taken at the site of interest The application of a contaminated sediment transport model requires a comprehensive data base comprising physical, chemical and biological parameters such as: critical shear stress for erosion and deposition [Gerbersdorf et al., 2004], erosive and depositional mass flux for sediments and contaminants, particle fall velocities, sorption parameters, kinetic parameters for

transformation and degradation processes. The spatial distribution of erodability and contamination including depth profiles plays a key role for predicting the total mass of eroded sediments and contaminants. In Fig. 1 representative depth profiles of some samples of typical groyne fields of the river Elbe are depicted showing the spatial and temporal variability of the critical erosion shear stress. Samples from another site reveal a typical pattern, meaning that deeper layers are highly polluted whereas the upper layers are less polluted due to reduction of river pollution in recent years (Fig. 2). Numerical Monte Carlo simulations based on the statistics of critical erosion shear stress and erosion rate coefficients have shown the influence on the expected value and the variance of the total mass of sediments eroded by a flood [Li, 2004].

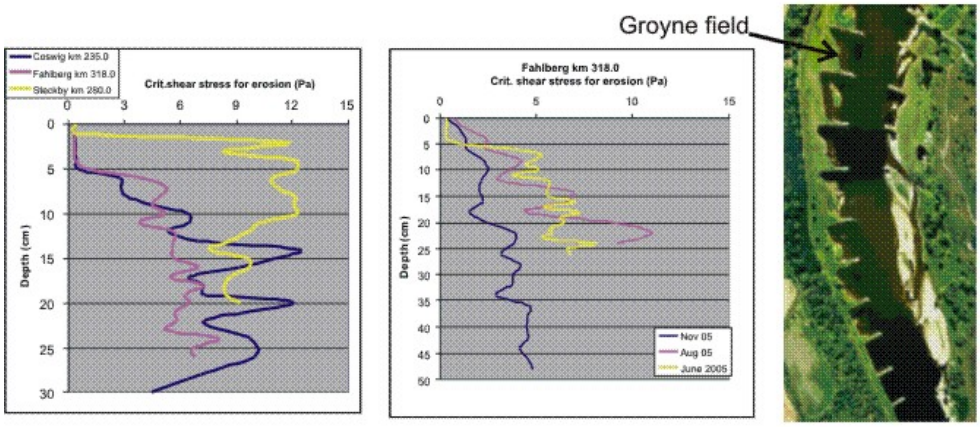


Figure 1. Spatial and seasonal variation of critical erosion shear stress: groyne fields of the river Elbe at Coswig, Fahlberg and Steckby [Jancke et al., 2006].

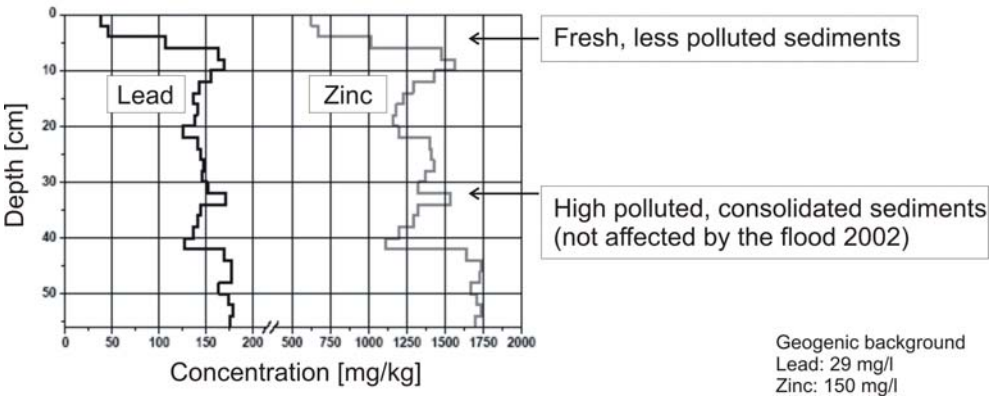


Figure 2. Depth profile of lead and zinc concentration, Elbe km 420,9 [Schwartz and Kozerski, 2004].

5. MODEL APPLICATION

5.1 Hot spots in the headwaters at the Upper River Rhine hydropower stations

The headwaters of the hydropower stations at the Upper River Rhine (built in the years 1961 to 1977) are considered to be one of the most important depots of contaminants in terms of toxicity and total amount in the Rhine catchment (Fig. 3). The headwater area shows characteristic sedimentation pattern as depending on the arrangement of the hydropower, the spillway section and the ship lock. The objective was to estimate the total mass of eroded particulate HCB (Hexachlorobenzene) during the flood in 1999 in a retrospective manner. Each of the six reservoirs was investigated to estimate the mass of

eroded HCB and to quantify the cumulative contribution of each reservoir to the total particulate HCB load released to the Lower River Rhine and monitored at the German/Dutch border. Unfortunately, prior to the flood no data on deposited sediments were available. The river discharge hydrograph, the suspended sediment outflow concentration and the associated daily HCB load through the turbine section were given and used as boundary conditions.

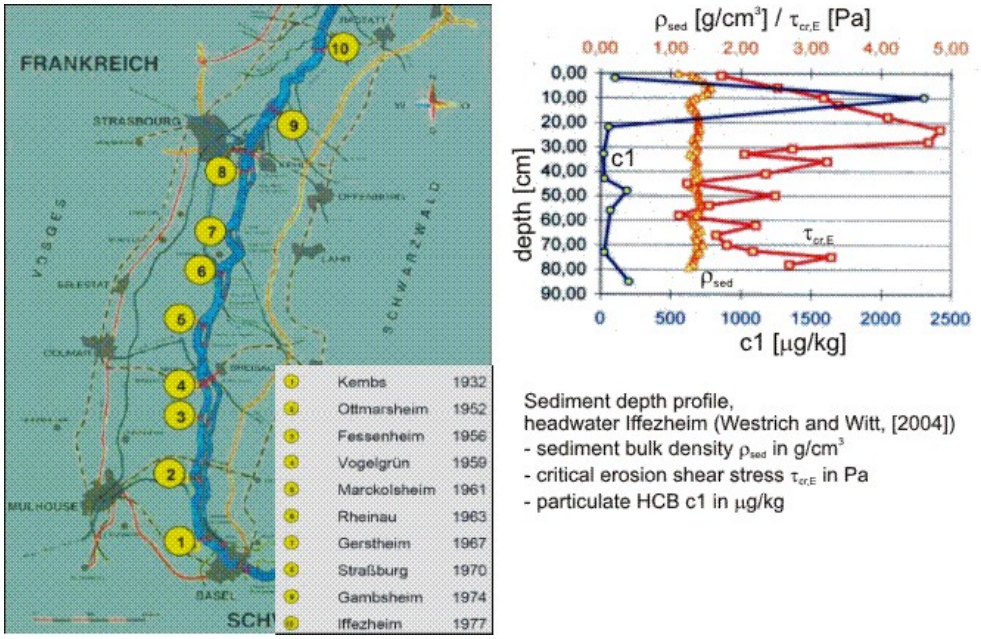


Figure 3. Layout of the hydropower stations at the Upper River Rhine, with a representative sediment core analysis.

Despite the extensive post-flood field investigation at the upstream station Marckolsheim the uncertainty of the calculated eroded particulate HCB was unsurprisingly high. The main reason was the small number of sediment samples, and the high spatial variability of sediment erodability and contaminant content of the post-flood samples (Fig. 3). The estimated eroded mass of HCB during the flood ranges from 2.4 to 17 kg with an average of 5 kg (Table 1). The computed values are beyond the ICPR (International Commission for the Protection of river Rhine) target value of 1.3 kg referring to a maximum permissible sediment contamination of 40 µg/kg for HCB [Jacoub, 2004].

Table 1. Particulate contaminant mass eroded during the flood in 1999 in the headwater of the lower six hydropower stations at the Upper River Rhine [Jacoub, 2004].

hydropower station reservoir	contaminated area (10^3 m^2)	particulate HCB (kg)	specific HCB mass content (g/m^2)	uncertainty assessment
Marckolsheim	54	5	0.11	underestimated
Rheinau	13	10	0.08	estimated
Gerstheim	178	14	0.08	plausible
Strasbourg	230	23	0.10	plausible
Gambsheim	68	6	0.09	highly underestimated
Iffezheim	36	3	0.08	highly underestimated
Total		61 (145 measured)		highly underestimated

According to the numerical results of the investigated reservoirs the total mass of HCB mobilized during the flood in all six reservoirs mounts up to some 61 kg which must be considered an underestimation because of the fact that the computation was performed

based on contamination data which were measured after the flood and averaged over the erosion depth of about 0,1m. Beside the numerical results, it is evident that the measured value of 145 kg HCB must be considered too low because the samples were taken in front of the turbines (right side river branch with discharge capacity of 1050 m^3/s) instead in the main stream through the weir branch (with discharge of 2900 m^3/s) and hence, the measurements can only capture a fraction of the total load.

4.2 Sediment dynamics in the near bank water zones of the River Elbe

The study aims at modelling the effect of the near bank river training works, so called groyne fields, on the transport dynamics of suspended sediments and particulate pollutants for different hydrological scenarios of the Elbe catchment. The interaction between the groyne fields and the main channel was investigated over a 112 km long stretch of the river Elbe to describe sedimentation during low flow periods and erosion during high discharge periods. The study includes the contribution of two major tributaries to the potential pollution of the main channel sediments (Fig. 4). To provide the lateral exchange coefficient and the overall sedimentation coefficient of the groyne field for the 1-d modelling laboratory experiments were performed. In addition, to provide high spatial resolution a depth averaged 2-d model for detailed local investigation of sedimentation, erosion processes in the groyne fields of a 50 km long inner section was applied and used to support and underline the 1-d model results. The 2-d model was partly calibrated on and applied to the extreme flood event of August 2002 as shown by the following results.

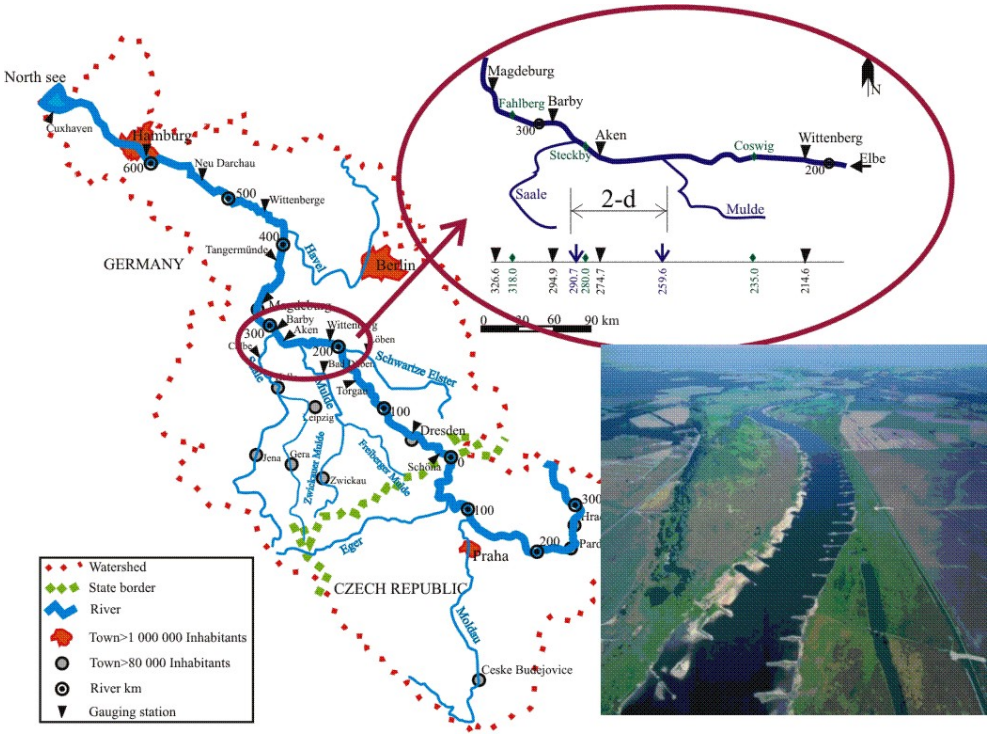


Figure 4. River Elbe catchment with a 112 km long section for numerical model investigation on the effect of groyne fields on the transport dynamics.

During the low discharge periods, from 1-st to 10-th August 2002, the river discharge was about 220 m^3/s with small flow velocities up to 0.8 m/s. Suspended sediments are in part deposited in the groyne fields (Fig. 5). The local sedimentation rates vary between about 5 x 10^{-5} – 5 x 10^{-6} kg/m^2s. High sedimentation occurs in the groyne fields and at the mouths of the tributaries Mulde and Saale as compared to the main channel. At the flood peak with Q=3965 m^3/s at 19-th August 2002, the flow velocities mounted up to 2.2 m/s and erosion mainly occurred in the groyne fields. The water depth on the flood plains is about 2.5 m and the velocity about 1.4 m/s. Due to high flow resistance of the river training works the maximum flow velocity near the banks is in the range of 0.8-0.9 m/s. The inflowing suspended sediment concentration at the upstream boundary increases to its maximum

value of 0.096 kg/m^3. The maximum concentration reaches a value of 0.11 kg/m^3 in the groyne fields due to erosion (Fig. 5).

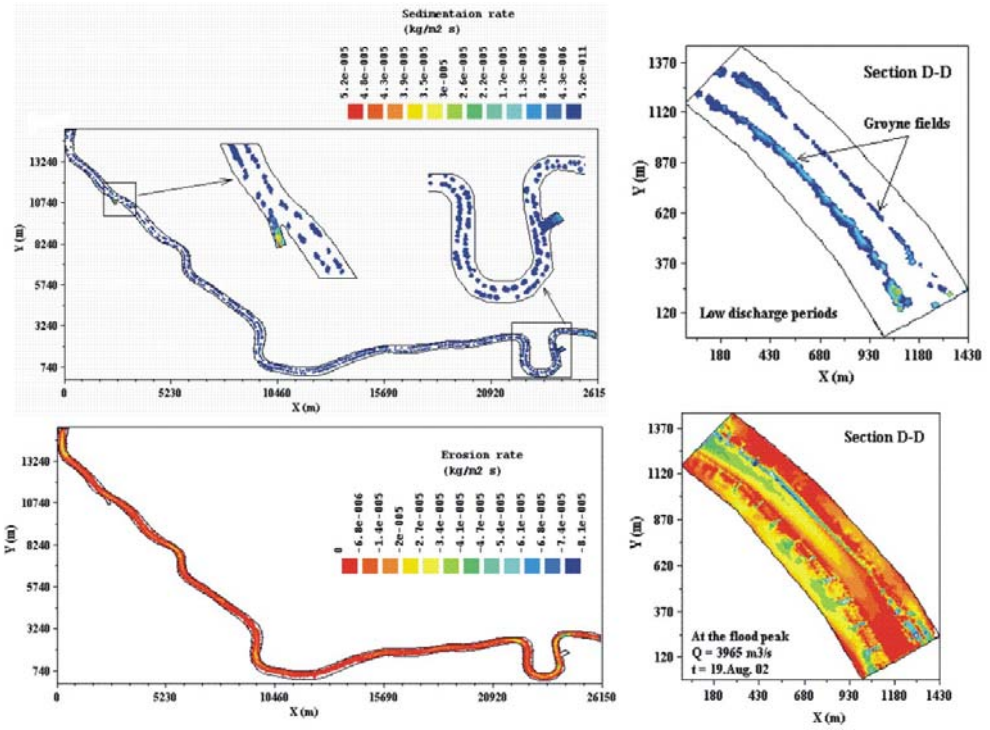


Figure 5. Numerical results: sedimentation and erosion in the groyne fields of the 50 km section embedded in the 112 km section of the 1-d model [Jacoub and Westrich, 2006].

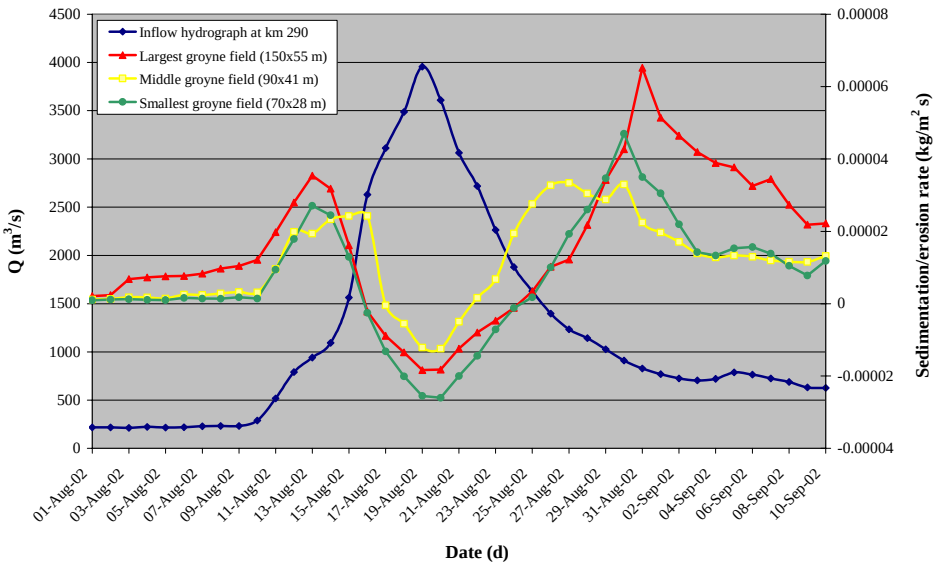


Figure 6. Spatially averaged sedimentation and erosion rates in three different types of groyne fields during the flood in August 2002 [Jacoub and Westrich, 2006].

Erosion rates also vary by an order of magnitude, i.e. from 7 x 10^{-6} – 7 x 10^{-5} kg/m^{2}s. The average erosion depth in the groyne fields varies from 0.004-0.02 m depending on time as well as size, shape and inclination angle of the groyne structure to the flow direction. The transport dynamics is clearly illustrated by the discharge controlled alternating sedimentation and erosion processes as shown by the model results in Fig. 6.

The 1-d strip model was used to investigate the long range sedimentation and erosion dynamics for selected hydrological scenarios. The results show the influence of a typical flood event in the catchment of the major polluter Mulde depicted in Fig. 7 as: (a) the longitudinal development of the spatially averaged deposition rate D in kg/m^{2}d in the left groyne fields, (c) the cumulative deposition in the groyne field approaching about 45% of the total input, and (d) the nearly exponential decrease of suspended matter concentration in the main channel with the dilution effect of the lower tributary Saale. The suspended sediment concentrations both in the main channel C_{mainch} and the adjacent groyne fields C_{gf} are larger but still proportional

$$c_{gf} = \alpha \cdot c_{main\ ch}$$

$$\alpha = \frac{1}{1 + \frac{\xi\, v_s\, b}{\varepsilon\, u\, h}}$$

(ξ = sedimentation coefficient, ε = exchange coefficient, b = groyne width, h = water depth, u = main channel flow velocity) however, the depositional adaptation length is significantly shorter in nature compared to the 1-d results. However, the total amount of deposited sediments in not affected by the 1-d approach. Despite the fact that the 1-d strip model is a simplified approach it shows the function of the groyne fields as sinks and sources of deposits and the influence of tributaries. The scenario of a 100 years flood event shows that about 33 % of the total inflowing suspended sediments are deposited on the flood plains, which is a reasonable figure compared to 27 % resulting from the mass balance of the spring flood in 2006 over a river section of 69 km as evaluated by Krüger and Jancke (not published).

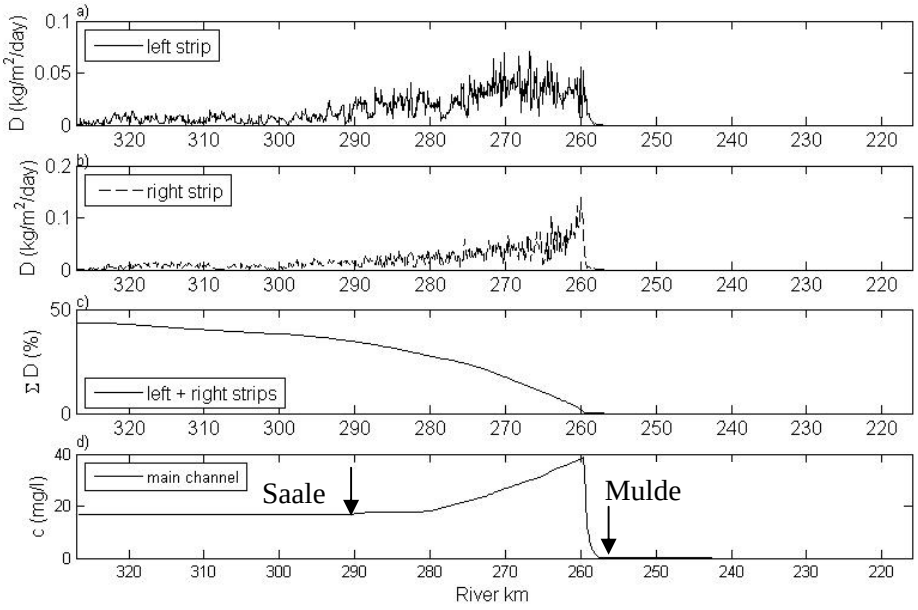


Figure 7. Longitudinal development of sediment deposition in the groyne fields due to input from the tributary Mulde and the dilution due to Saale [Westrich and Prohaska, 2008].

5. CONCLUSIONS

Numerical models are useful tools to investigate sediment associated pollutant transport processes in natural rivers, navigational channels and backwaters. Erosion, dispersion and deposition of mobilized sediments can be predicted on river basin scale as basic information for an environmental risk assessment. Field measurements are required to

capture the large spatial variability of both sediment properties and contaminant inventory and to improve the quality of the model predictions towards risk assessment and efficiency of remediation measures. Contaminated sediment transport models allow to bridge the gap between emission and immission issues as required by the EU Water Framework Directive WFD. Numerical models provide basic information for a site specific risk assessment by tracing mobilized hot spots and localizing pollutant sources. They can also be used by stakeholders and decision makers, e.g. water authorities, water and environmental engineers, to establish a risk based sediment management at the catchment scale.

ACKNOWLEDGEMENTS

The experimental and numerical investigation on the river Elbe were supported by the European Commission; project MODELKEY (Contract No 511237-GOCE).

REFERENCES

Förstner, U., S. Heise, R. Schwartz, B. Westrich and W. Ahlf, Historical Contaminated Sediments and soils at the river Rhine basin scale, *Journal of Soils and Sediments* 4(4) 247-260, 2005.

Gerbersdorf, S., T. Jancke, and B. Westrich, Physico-chemical and biological sediment properties determining erosion resistance of contaminated sediments - Temporal and vertical pattern at the Lauffen reservoir river Neckar, Germany, *LIMNOLOGIA*, 35, 132-144, 2004.

Jacoub, G., Development of a 2-D Numerical Module for Particulate Contaminant Transport in Flood Retention Reservoirs and Impounded Rivers, Ph.D. thesis, University of Stuttgart, Heft 133, ISBN 3-933761-36-0, 2004.

Jacoub, G., and B. Westrich, Modelling transport dynamics of contaminated sediments in the headwater of a hydropower plant at the Upper river Rhine, *Acta hydrochimica and hydrobiologica*, 3 (3), 2006.

Jancke, T., S. Prohaska, and B. Westrich, Analysis and evaluation of data on river morphology, to identify sedimentation and erosion areas, MODELKEY scientific report, 2006.

Li, C.C., Deterministisch-stochastisches Berechnungskonzept zur Beurteilung der Auswirkungen erosiver Hochwasserereignisse in Flussstauhaltungen, Ph.D. thesis (in German), Universität Stuttgart Mitteilungen, Heft 129, ISBN 3-933761-32-8, 2004.

Prohaska S., and B. Westrich, One-dimensional multi-strip transport model for long distance sediment transport in navigation channels, case study the River Elbe, paper presented at 23. Conference of the Danubian Countries on the Hydrological Forecasting and Hydrological Bases of Water Management, Belgrade, Serbia, August, 2006.

Schwartz, R., and H.P. Kozerski, Bestimmung des Gefahrenpotenzials feinkörniger Buhnenfeldsedimente für die Wasser- und Schwebstoffqualität der Elbe sowie den Stoffeintrag in Auen, In: Geller et al. (Eds): *Schadstoffbelastung nach dem Elbe-Hochwasser 2002*, UFZ, Magdeburg, 2004.

Westrich B., and U. Förstner, *Sediment Dynamics and Pollutant Mobility in Rivers – an interdisciplinary approach*, Springer, Heidelberg, 2007.

Westrich, B., and S. Prohaska, Exposure modelling aiming at the prediction of spatial distribution of deposited particulate key toxicants for different hydrological scenarios of the River Elbe and its tributaries, MODELKEY scientific report, 2008.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Large-scale modelling of P-inputs from point and diffuse sources in German river basins

Björn Tetzlaff and Frank Wendland
Research Centre Jülich, Institute of Chemistry and Dynamics of the Geosphere (Agrosphere, ICG-4), D-52425 Jülich, Germany
E-mail: b.tetzlaff@fz-juelich.de

***Keywords**:* P-inputs, eutrophication, river basin management, EU-WFD

Mean annual P-inputs from diffuse and point sources in large river basins can be quantified with the MEPhos model. Its emission approach is described, which distinguishes between both eight different input pathways (municipal waste water treatment plants, industrial plants, combined sewer overflows, rainwater sewers, artificial drainage, erosion, wash-off, groundwater-borne runoff) and various types of critical source areas named phosphotopes. Then, the results of a model application to two large-scale river basins are presented and discussed. The river basins (ca. 12,500 km²) each show contrasting natural conditions, land use patterns as well as population and industry densities. Total P-emissions from each river basin range around 1,600 tonnes per year but the importance of sources varies significantly. While in the Ems river basin 87 % of all P-inputs originate from diffuse sources, their portion achieves only 32 % in the Rhine river sub-basin. Due to the fact that diffuse entries are mainly made up by erosion (11 % of all P-inputs), measures for tackling P-inputs will aim at erosion prevention and sediment retention.

1. INTRODUCTION

As the inventory of the EU-water framework directive (EU-WFD) revealed in 2004, high nutrient inputs are a major concern for most inland and coastal waters in Germany. As a result the "good status" will not be achieved for 62 % of rivers and 86 % of coastal waters (BMU 2005). To improve water quality until 2015 the EU-WFD demands the drawing-up of detailed river basin management plans and programmes of measures until 2009 (EC 2000). This provokes a demand for instruments and models applicable at the river basin scale for the quantification of nutrient inputs (actual state) and for the prediction of potential future states. Model applications should aim at the localization of critical source areas within river basins as a basis for proposing reduction measures which are adapted to site properties. This requires modelling approaches differentiating between sub-areas and input pathways.

The authors were involved in an interdisciplinary research project named 'REGFLUD' commissioned by the Federal German Research Ministry (BMBF) and finished in 2005. REGFLUD was part of the BMBF-research programme "river basin management". The aim of the project was to develop policy reduction measures for tackling eutrophication in German river basins. In this context the pathway- and area-differentiated phosphate model MEPhos was developed (Tetzlaff 2006, Tetzlaff et al. 2008a). The model was applied to two study areas, the catchments of the River Ems (12,900 km²) and parts of the River Rhine (12,160 km²), located in the north western part of the Federal Republic of Germany. Both study areas are of similar size but show contrasting natural conditions. The Ems catchment is mainly characterized by its lowland situation with sandy and boggy soils under intensive agricultural use. In accordance with the relatively low soil fertility intensive animal

husbandry is conducted with partly more than three livestock units per hectare. The investigation area Rhine is dominated by upland conditions with forest and grassland use. Arable land use is restricted to the flat Rhenish bight with fertile loess soils. Furthermore the area comprises sub-regions with high population and industrial densities.

2. GENERAL DESCRIPTION OF THE P-MODEL MEPHOS

The empirical model MEPhos was developed to calculate mean annual total-P inputs from point and diffuse sources to surface waters (Tetzlaff 2006). The model distinguishes between eight different pathways (artificial drainage via moles, pipes or channels), groundwater-borne runoff, erosion, wash-off, rainwater sewers, combined sewer overflows, municipal waste water treatment plants, industrial effluents) as shown in figure 1.

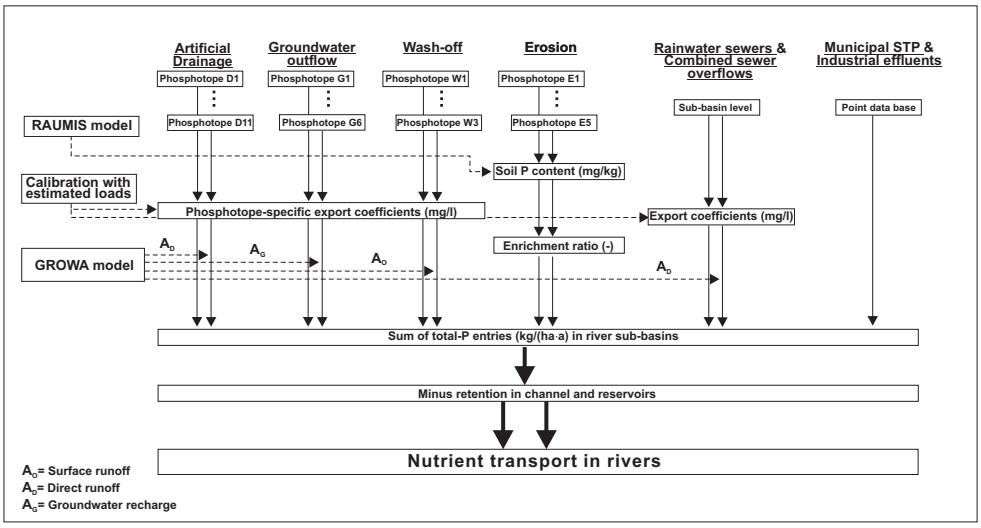


Figure 1. Overview of the MEPhos model

The model approach for each pathway depends on the resolution of input data. Modelling diffuse P-inputs via artificial drainage, groundwater-borne runoff, erosion and wash-off is performed area-differentiated, based on phosphotopes as discontinuous source areas. Phosphotopes are regarded as homogeneous types of sub-areas, defined by a set of different parameters. The coincidence of several parameters on the same area controls the emission potential. For each pathway several phosphotopes are defined (figure 1). The parameters controlling the mean annual P-input from diffuse sources are described in Tetzlaff et al. (2008a) and can be quantified for macroscale investigation areas with already available data sets from federal or state authorities (table 1). Phosphotopes are processed in GIS by classification and overlaying of data sets. This process is explained in greater detail in Tetzlaff et al. (2007b). For modelling diffuse P-emissions via artificial drainage, groundwater-borne runoff, erosion and wash-off 25 different phosphotopes are derived in total (figure 1). In order to model P-inputs to surface waters calibrated export coefficients (unit mg/l) are assigned to all phosphotopes instead of using standard literature values. For determining the pathway-related mean annual runoff components the MEPhos model is coupled to the water balance model GROWA (Kunkel and Wendland 2002). This model provides data about the mean annual sums of surface runoff, drain flow and groundwater runoff needed as transport media for the pathway-differentiated P-modelling (figure 1).

Table 1: Data sets for modelling and validating P-inputs to surface waters

Data sets	**Spatial resolution and origin**
Mean annual sums of surface runoff, drain flow and groundwater recharge	Results of water balance model GROWA (Grid data, 50·50 m²)
Soil map and parameters (depth of ground-water table, perching water influence, soil water content, clay content, USLE factors)	Soil map 1:50,000 (Geological Surveys)
Land use	CORINE (Federal Statistical Office)
Digital Terrain Model	DTM 10·10 m² and 50·50 m² (State agencies for Land Survey)
Livestock density	Statistical data for municipalities (State statistical offices)
Nutrient balance P	Results of the agro-economic model RAUMIS
Agricultural land with artifical drainage	Derivation procedure described in Tetzlaff et al. (2008b)
River network	ATKIS-DLM 25 (State agencies for Land Survey)
River discharge data	Gauging station network run by state agencies
Surface water and groundwater quality data	Sampling station network run by state agencies
Point data on effluents from municipal sewage treatment plants and industry	Data base from state agencies
- Data about stormwater retention ponds and grade of connection to waste water systems - Reservoir data	Data base from state agencies, literature

Due to the availability of coarser data for modelling mean annual P-inputs via rainwater sewers and combined sewer overflows, no phosphotopes can be derived. Therefore, these emissions are modelled integratively for river sub-basins with sizes between approx. 50 and 500 km² (figure 1). P-inputs from municipal waste water treatment plants and from industry are modelled separately for each site using point data from monitoring data bases from environmental state agencies (figure 1, table 1, equation 1).

$$F_{KA} = \frac{1}{n} \cdot \sum_{i=1}^{n} (Ci_{Pges} \cdot AWM) \quad (1)$$

FKA = Mean annual P-input from MWWTP [t/a]; AWM = mean annual plant discharge [m³/a]; CiPges = mean concentration of total-P in plant discharge [mg/l] for measurement i; n = number of measurements i per year

After calculating mean annual P-inputs via all eight pathways the emissions are summed up for river sub-basins related to water quality gauges (figure 1). Then the mean annual load of an upstream sub-basin is added and the P-retention is subtracted (equation 2). In MEPhos P-retention in both running and standing waters is considered. This enables a validation of the modelled P-loads by comparison with mean annual P-loads estimated from measured water quality and discharge data.

$$F_{bas} = \left[\left(\sum F_{dr} + \sum F_{gw} + \sum F_{wash} + \sum F_{eros} + \sum F_{rws}\right) + \sum F_{mwwtp} + \sum F_{id} + \sum F_{cso}\right] + F_{upbas} - R_r - R_s \quad (2)$$

F_{bas} =estimated load for river sub-basin, F_{dr} = P-inputs via artificial drainage, F_{gw} = P-inputs via groundwater-borne runoff, F_{wash} = P-inputs via wash-off, F_{eros} = P-inputs via erosion, F_{rws} = P-inputs via rainwater sewers, F_{mwwtp} = P-inputs from municipal waste water treatment plants, F_{id} = P-inputs from industrial effluents, F_{cso} = P-inputs via combined sewer overflows, F_{upbas} = P-load of an upstream river sub-basin, Rr = P-retention in running waters, Rs = P-retention in standing waters

The retention of P during the transport phase in surface waters is modelled separately for running and standing waters. Behrendt and Opitz (2000) have developed an approach to describe the retention in rivers integratively, which is applied in the MEPhos model. The increased P-retention occuring e.g. in reservoirs due to reduced flow velocity, increased travel time as well as higher sedimentation rate is modelled according to Molot and Dillon (1993).
In the following chapter an example will be given about modelling erosion-related P-inputs with MEPhos.

3. EXAMPLE: MODELLING DIFFUSE P-INPUTS VIA EROSION

The controlling factors for P-inputs to surface waters via erosion are soil loss from arable land within river catchments, sediment delivery ratio, P-content in the top soil as well as by P-enrichment during the erosion process. Diffuse P-inputs to surface waters via erosion are modelled after equation 3. This equation follows in general the universal soil loss equation adapted to German conditions by Schwertmann et al. (1990), but implies several modifications in order to take into account the high sensitivity of relief features.

$$E = SSF \bullet [(R \bullet K \bullet LS \bullet C)] \bullet SDR \bullet PC \bullet \frac{P_S}{P_{OA}} \quad (3)$$

E = mean annual P-inputs via erosion [kg/(ha·a)]; SSF=slope shape factor after Prasuhn and Grünig (2001); R = rain erosivity factor [N/(h·a)]; K = soil erodibility factor [t·h/(ha·N)]; LS = combined slope steepness and slope length factor after Moore and Wilson (1992) [-]; C = crop and management factor [-]; SDR = sediment delivery ratio [%/100]; PC = content of total-P in the top soil [mg/kg]; ER = enrichment ratio [-]; P_S = P-content of suspended matter given at high discharge levels [mg/kg]; P_{OA}= P-content in the top soil of sediment delivering arable land [mg/kg]

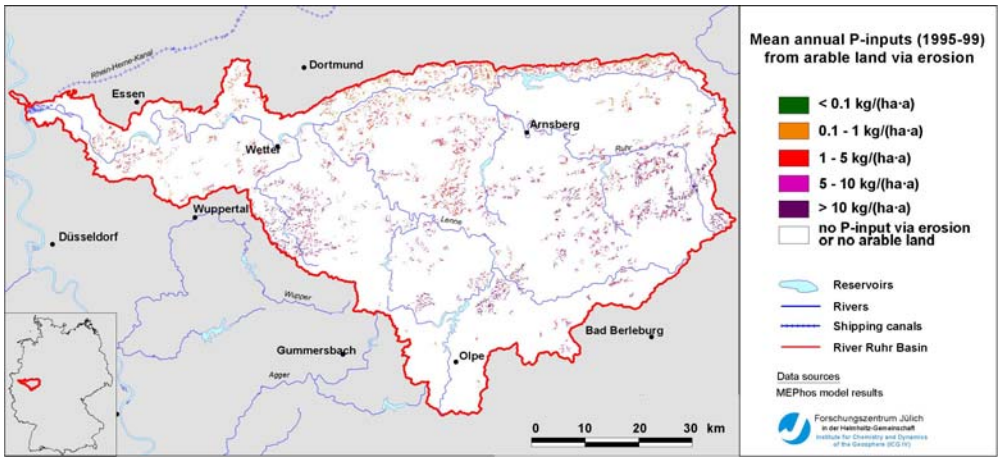


Figure 2. Example for MEPhos results for modelling mean annual P-inputs (1995-1999) via erosion: the River Ruhr catchment

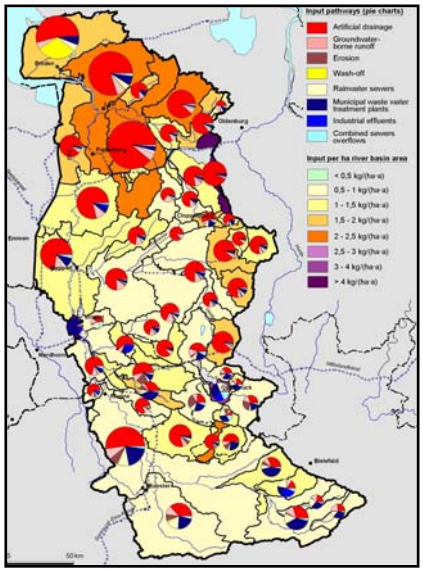
With regard to proposals for efficient eutrophication reduction measures modelling of diffuse P-inputs by MEPhos has to be performed area-differentiated, i.e. the eroding arable land area has to be disaggregated. Therefore, phosphotopes for modelling erosion inputs are delineated, they represent source areas for the release of sediment and particulate-P and are made up by erodible arable land which is at the same time hydraulically connected to the network of rivers and flow paths (Tetzlaff and Wendland 2007). These connections are provided by flowpaths determined by the relief. Such morphological flow paths are modelled using a highly-resolved digital terrain model and the algorithm “D infinity” (D∞, Tarboton 1997). Then, those flow paths, which are connected to the river network, are

selected and buffered with stripes of 2·30 m width, according to findings of Fried et al. (2000). Only those sub-areas contained in the 60 m stripes are regarded as hydraulically connected. Then, phosphotopes for modelling P-inputs via erosion are derived by intersecting the data sets of erodible arable land with the buffered network of flow paths and rivers. Depending on their erosion potential five different phosphotopes are distinguished (Hydraulically connected arable land with erosion potential of <1 t/(ha·a), 1-5 t/(ha·a), 5-10 t/(ha·a), 10-15 t/(ha·a) and >15 t/(ha·a)).
According to equation 3, the level of particulate P-inputs results not only from the sediment delivery to surface waters, but also from the P-content of the top soil and the enrichment ratio. The calculation of the total-P content in the top soil is based on P-surpluses and clay content following Behrendt et al. (1999). P-surplus is modelled on a county level employing the agro-economic model RAUMIS developed and run by the German Federal Agricultural Research Centre (IAP and FAL 1996), the clay content is given by soil maps on a scale of 1:50,000 (figure 1, table 1). An enrichment ratio has to be determined in order to consider the selective effect of water erosion. For this a method is employed in the MEPhos model requiring measured water quality data, above all, about the P-content of suspended matter (Behrendt et al. 1999). The enrichment ratio for river sub-catchments is given by the quotient of the P-content of suspended matter given at high discharge levels (PS) and the P-content in the top soil of sediment delivering arable land (POA), as shown in equation 3. More details about the erosion modelling routine of MEPhos can be found in Tetzlaff and Wendland (2007a).
An example for the model results achievd by using equation 3 is given in figure 2. Because of the high spatial differentiation a clipping of the map for the River Ems and Rhine sub-catchments is presented, concentrating on the River Ruhr catchment (4,485 km²), being a tributary of the River Rhine. Inputs of total-P vary between less than 0.1 and more than 10 kg/(ha·a). Erosion-related P-inputs are characterised by a small spatial extent of sediment delivery areas with highly differing input levels (figure 2). The small spatial extent of the phosphotopes results from the disaggregation of erodible arable land, its intersection with buffered flow paths and the application of highly-resolved data sets, above all the digital terrain model with a resolution of 10·10 m². High erosion-related P-inputs are to be found mainly in sub-regions with steep slopes and widespread arable land use (figure 2). For the entire River Ems catchment mean annual P-inputs of 34 t/a or 2.94 kg/(km²·a) are modelled, the results for the entire River Rhine sub-catchment are significantly higher with 234 t/a or 14.8 kg/(km²·a). The difference reflects the contrasting natural and land use conditions between the two areas.

4. TOTAL P-INPUTS FROM DIFFUSE AND POINT SOURCES AND VALIDATION OF MODEL RESULTS

After modelling mean annual P-inputs via all eight pathways diffuse and point source emissions can be summed up. This leads to mean annual total loads for the period 1995-1999, which amount to 1,666 t/a in the River Ems catchment and to 1,574 t/a for the River Rhine sub-catchment. Figures 3 and 4 show the relevance of all eight pathways for the mean P-input. In the entire River Ems catchment diffuse P-emissions account for 87 % of all P-inputs to surface waters. It can be stated for almost all of the 56 sub-catchments that diffuse inputs dominate and that the pathway “artificial drainage” is of highest importance in this lowland river catchment. The percentage of inputs via artificial drainage differs between 16 and 89 % in total, it is above 50 % in most sub-catchments (figure 3). The highest loads from non-point sources are modelled for the lower reaches of the River Ems and its tributaries, which correspond with the decreasing population density and the increasing intensity of agricultural activities from south to north. Due to the origin of these high loads from artificially drained agricultural land, i.e. as mainly soluble reactive and therefore highly algae-available P, the agricultural activities on raised bog soils result in a tremendous local and regional eutrophication potential for surface waters. A significant reduction of P-fluxes to the receiving coastal waters due to P-retention during fluvial transport cannot be expected due to the short distance between the source areas and the mouth of the River Ems. As a consequence, also the wadden sea receives a high phosphate

input from diffuse sources causing increased eutrophication risks in coastal waters at least, where no limitations due to the nitrogen content exist.



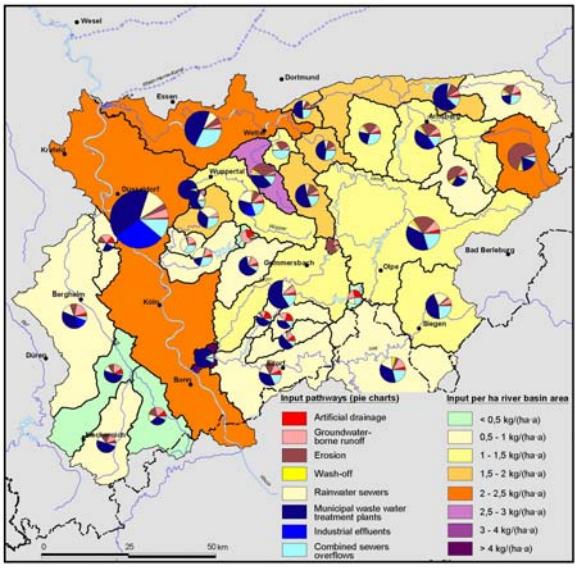


Figure 3. Mean annual P-inputs (1995-1999) of sub-catchments and relevance of pathways for the River Ems catchment (left)

Figure 4. Mean annual P-inputs (1995-1999) of sub-basins and relevance of pathways for the River Rhine sub-basin (right)

While the mean total P-inputs (1995-1999) in the investigation area Rhine equal those in the Ems area, the mean relation between diffuse and point source inputs of 32:68 differs significantly (figure 4). The main reasons are the high densities of population and industry, above all at the River Rhine and along the lower reaches of the River Ruhr. High emissions via combined sewer overflows play an important role, too. When the MEPhos modelling results are examined for river sub-catchments, areas where diffuse emissions are a major concern, are revealed also for the investigation area Rhine (figure 4). They are located in the upper reaches of the Rivers Erft and Ruhr, i.e. in the south-west and east of the investigation area. As a result of arable land use on steeper slopes, partly on loess soils, soil erosion is responsible for these increased diffuse P-loads.

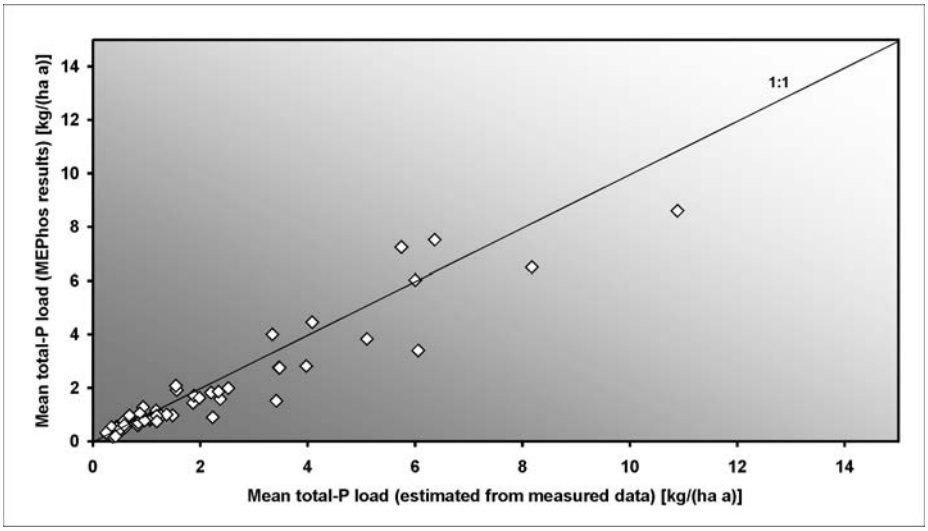


Figure 5. Validation of MEPhos modelling results against mean annual loads (1995-1999)

The validity of the MEPhos modelling results (1995-99) is checked by a comparison with mean annual P-loads estimated from measurements of daily discharge and monthly total-P-

concentration following the OSPAR method (OSPAR Commission 1998). When selecting the gauging stations, attention was paid to achieve a great variability with respect to basin size, natural conditions and population density. Furthermore, the extent of gaps in the time series of the measured data should be as small as possible. These requirements are fulfilled for 32 gauge-related sub-catchments in the River Ems basin and 24 sub-catchments in the River Rhine sub-basin. The selected sub-catchments are sized between approximately 50 and 500 km². The general validation process is explained in chapter 2. Figure 5 shows the validation results.

The mean annual P-loads differ between approx. 0.1 kg/(ha·a) and approx. 11 kg/(ha·a) (figure 5). In general, the diagram shows a good correlation between loads estimated from measurements and modelling results. The mean deviation amounts to 7.3 %, the coefficient of determination reaches 89 %. For eight sub-catchments the deviation is below 10 %, for 20 of 56 sub-catchments below 20 %. Errors in this order of magnitude are within the usual variation range of empirical models. Although the most recent data sets with the highest spatial resolution and information content were used in this project, unavoidable measuring and interpolation errors are undoubtedly involved. Systematic errors causing the deviations could not be identified. Because of the restricted data availability in the field of urban drainage and sanitation, the model routines for the quantification of mean annual P-inputs via rainwater sewers and combined sewer overflows have to get by on a limited number of parameters. With respect to these restrictions the model results can be regarded as valid.

5. CONCLUSIONS AND MANAGEMENT OPTIONS FOR TACKLING EUTROPHICATION

As outlined in chapter 1, an overall objective of the EU-WFD is to achieve the good status of surface waters. Important tools in this context are cost-effective measures for the reduction of nutrient inputs (UBA 2004). Based on valid model results, management options for the reduction of eutrophication can be proposed and scenarios calculated. With respect to cost-effective measures the focus for reduction has to be on “hot spots” as priority areas, i.e. critical source areas with small spatial extent and high emission. From the area-differentiated MEPhos model results it can be concluded that in the River Rhine sub-catchment P-inputs via erosion from arable land with soil loss of >15 t/(ha·a) sum up to approx. 24 % of all diffuse inputs, originating from less than 1 % of the catchment area. Published results from field studies were used to assess the level of sediment input reduction to be achieved by various protection measures. The MEPhos results were then reduced due to these findings. On-site and off-site erosion protection measures like mulching combined with riparian buffer stripes would lead to a reduction from 87 t/a to 39 t/a. Converting these highly eroding arable lands into grasslands would reduce the P-inputs by 98 %. But due to the fact that in the River Ruhr catchment a series of reservoirs act as sediment sink and due to the dominance of point sources, tackling diffuse sources in the River Rhine sub-catchment has only little effect on the general water quality situation, especially in the lower reaches, where water quality problems due to eutrophication have been observed (LUA 2002). A far larger improvement would be achieved by taking technical measures for small and medium sewage treatment plants, e.g. P-elimination techniques, and by extension of measures to reduce combined sewer overflows.

By providing area-differentiated model results about diffuse and point source P-inputs to surface waters via eight different pathways, applications of the MEPhos model can contribute to the tasks of the EU-WFD and therefore to a sustainable use of water resources. In the meantime, MEPhos has been applied to the German state of Hesse (21,115 km²) to identify surface water bodies with eutrophication risks due to high inputs of ortho-P. Furthermore, policy advice for the administrative client is provided by quantifying the effects of state-wide mitigation measures for the surface water bodies.

REFERENCES

Behrendt, H., P. Huber, M. Kornmilch, D. Opitz, O. Schmoll, G. Scholz, and R. Uebe, *Nutrient emissions into river basins in Germany*. – Federal Environmental Agency Germany, Berlin, 1999.

Behrendt, H. and D. Opitz, Retention of nutrients in river systems dependence on specific runoff and hydraulic load, *Hydrobiol*. 410, 111–122, 2000.

BMU, Federal German Ministry for Environment, *Environmental Policy - Water Framework Directive - Summary of River Basin District Analysis 2004 in Germany*. 2nd edn., Berlin, 2005.

Fried, J.S., D.G Brown, M.O. Zweifler, and M.A Gold, *Mapping Contributing Areas for Stormwater Discharge to Streams Using Terrain Analysis*, in: Wilson, J.P. and Gallant, J.C. (eds): Terrain Analysis - Principles and Applications. New York, 183-203, 2000.

IAP and FAL (eds), *Entwicklung des gesamtdeutschen Agrarsektormodells RAUMIS96 - Endbericht zum Kooperationsprojekt*. Bonn and Braunschweig-Völkenrode, 1996.

Kunkel, R. and F. Wendland, The GROWA98 model for Water balance analysis in large river basins - the River Elbe Case Study, *J. Hydrol.* 259, 152-162, 2002.

LUA, Environment Agency State North Rhine Westphalia (eds), *Gewässergütebericht 2001 Nordrhein-Westfalen, Berichtszeitraum 1995-2000*. Essen, 2002.

Molot, L.A. and P.J. Dillon, Nitrogen mass balances and denitrification rates in central Ontario Lakes, *Biogeochem* 20, 195–212, 1993.

Moore, I.D. and J.P. Wilson, *Length-slope factors for the revised universal soil loss equation: Simplified method of estimation*, J Soil Water Cons 49, 174-180, 1992.

OSPAR Commission (ed.), *Principles of the comprehensive study on riverine inputs and direct discharges (RID)*. Reference 1998-05, 1998.

Prasuhn, V. and K. Grünig, *Evaluation der Ökomaßnahmen - Phosphorbelastung der Oberflächengewässer durch Bodenerosion*. SchrR FAL 37, Zürich-Reckenholz, 2001.

Schwertmann, U., W. Vogl and M. Kainz, *Bodenerosion durch Wasser - Vorhersage des Abtrags und Bewertung von Gegenmaßnahmen*. 2nd edn, Stuttgart, 1990.

Tarboton, D.G., A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resources Res* 33, 309-319, 1997.

Tetzlaff, B., *P-loads from diffuse and point sources in large river basins* (in German). Ph.-D. thesis, Research Centre Jülich, Umwelt/Environment, 65, 2006.

Tetzlaff, B. and F. Wendland, *Modelling P-fluxes from diffuse and point sources in heterogeneous macroscale river basins using MEPhos*. In: Förstner U. and B. Westrich (eds.), Sediment Dynamics and Pollutant Mobility in Rivers. Series Environmental Science, Springer, 206-216, 2007a.

Tetzlaff, B., P. Kreins, R. Kunkel and F. Wendland, Area-differentiated model-ling of P-fluxes in heterogeneous macroscale river basins. *Water Science and Technology* 55, 123-131, 2007b.

Tetzlaff, B., J.-F. Hake, H. Vereecken and F. Wendland, Sustainable use of water resources in Europe and the role of integrated modelling of phosphate fluxes, *International Journal of Global Environmental Issues* (in press), 2008a.

Tetzlaff, B., P. Kuhr and F. Wendland, A new method for creating maps of artificially drained areas in large river basins based on aerial photographs and geodata, *Int. J. Irrig. Drain*. (in press), 2008b.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A System Dynamics Model for the Assessment of Risks and Risk Mitigation Options at Catchment Scale

M. Finkel[a] **and U. S. McKnight**[b]

[a] *Center for Applied Geoscience, University of Tuebingen, Sigwartstr. 10; 72076 Tuebingen, Germany (michael.finkel@uni-tuebingen.de)*

[b] *Center for Applied Geoscience, University of Tuebingen, Sigwartstr. 10; 72076 Tuebingen, Germany (ursula.mcknight@uni-tuebingen.de)*

Abstract: In Europe and world-wide, adequate water quality is one of the most eminent concerns for future generations. Diffuse pollution inputs such as agriculture, as well as local sources in terms of industrial, military and mining sites pose considerable stress on water quality. For the design of sustainable management strategies, decision makers need appropriate planning tools to assist them in the assessment, selection and optimisation of possible alternatives. Recent policies set a focus on the management of existing risks to human health and ecosystems within a tiered assessment and decision making approach giving increased emphasis to modelling. We propose a novel system dynamics modelling approach that quickly provides estimates of current and future risks originating from soil and groundwater contamination, as well as of the costs of possible risk reduction strategies. The use of analytical approaches for contaminant fate and transport modelling, as well as for risk assessment, enables a fast and effective "screening" method that is particularly qualified for large-scale applications. Results from a pilot case study in Germany illustrate the large potential of the proposed method for the preliminary assessment of management scenarios involving individual treatment technologies, as well as combinations of those.

Keywords: Risk-based land management, modelling, system dynamics, cost-effectiveness

1. INTRODUCTION

In Europe and world-wide, adequate water quality is one of the most eminent concerns for future generations. This is recognized by recent EU policies including the Water Framework Directive (WFD), the associated Groundwater Directive (GWD), and the EU communication on a Thematic Strategy for Soil Protection. Stress on water quality comes from a range of diffuse pollution inputs such as agriculture, as well as from local sources. The latter stem predominantly from industrial, military and mining activities during the past century, which have led to a vast contamination of soil, ground-water and surface waters. In many cases a complete restoration of the subsurface is both technically and economically infeasible, and remedial actions have to be focused on the management of existing risks to human health and ecosystems.

For the design of sustainable management strategies, decision makers need appropriate planning tools to assist them in the assessment, selection and optimisation of possible alternatives. Complexity and scale of the situation at these sites (in terms of contamination extent, boundary conditions/limitations, stakeholders, etc.) has led to the proposal of a tiered framework for site investigation, risk assessment and management [e.g. UK Environment Agency, 1999; ASTM, 2002]. Recent policies request an increased emphasis on modelling (e.g. EU Water Framework Directive).

A methodological gap exists particularly with respect to preliminary assessment methodologies to support early decisions on prioritising hot spots and preselection of mitigation options to streamline the further planning process. To fill this gap, we propose a novel system dynamics modelling approach utilising a multiple source-pathway-receptor concept. The model quickly provides reasonable estimates of current and future risks originating from soil and groundwater contamination, as well as of the costs of possible remediation (risk reduction) strategies for integration into the planning process.

2. THE MODEL CONCEPT

2.1 System Dynamics Approach

System dynamics (SD) is a methodology for representing complex systems and analysing their dynamic behaviour and dates to the work of Forrester [1961]. The focus of SD is to study and analyse the behaviour of complex system changes through time. The SD approach has been widely applied to water related issues, including water balance simulation [Khan et al., 2008] and world water dynamics modelling [Simonovic, 2002]. In this study, an SD model is used to assess the risks originating from contaminations in soil and groundwater, as well as possible risk mitigation options. The model describes the dynamic system of soil, water and contaminants in the subsurface and is implemented in the decision support tool (DSS) CARO-plus [McKnight & Finkel, 2008a and 2008b]. The dynamics of the system are dominated by contaminant release from long-term source zones and contaminant flux along the pathway to the receptor(s). The contaminant mass flux model, the risk assessment module and all other parts of the model framework, including technological performance and cost models for various remediation options, were developed using the off-the-shelf software platform Vensim [Ventana Systems, 2007]. The Vensim platform and other similar SD tools, e.g. STELLA [isee systems, 2007], particularly simplify the conceptualisation and set-up of models of dynamic systems. The modelling process starts with sketching the relevant variables and input constants of a model and its relationships (causal connections). The variables are then further distinguished into stocks (e.g., ‘Compound Mass i at Source’; see Figure 2), flows (e.g., ‘Mass Release Rate Compound i’) and auxiliaries (e.g., ‘Mole Fraction Compound i’). Model building is completed by specifying equations for each of the relationships and numerical quantities for input parameters. This kind of “visual” icon-based model building does not only ease model building but also facilitates the communication of the model to the stakeholders. The computation of a model run includes an automatic ordering of the network of equations and time step-wise calculation of auxiliaries, flows, and finally, stocks by integration of previously calculated flows going in and out of the particular stock.

2.2 Assessment Tool CARO-plus

The utilisation of a source-pathway-receptor (SPR) concept implies a considerable simplification of reality in the model. Mass transport is conceptualised via two SPR sequences for source zones in soil and groundwater (Figure 1). This simplification is, on the one hand, an adequate measure since the data that is available for early preliminary assessment is typically scarce and so does not allow for sophisticated modelling. On the other hand, the simplification is a necessary means to use analytical models and to establish a computationally efficient model-based assessment tool. The undisturbed natural system, as well as the effects of possible remediation options may be analysed with CARO-plus in terms of contaminant mass flux and concentration, and risk indices (carcinogenic/non-carcinogenic). Using either of these three metrics, compliance with environmental targets can be controlled at various lines or points of compliance. Following the MEPAS approach developed by Strenge & Smith [2006], the risk assessment considers various transfer paths that are specific to the particular land use type. Simulation of the effects of potential risk mitigation actions includes partial source removal and plume treatment options. To give an example of the model structure, Figure 2 sketches the major structural elements of the source release model. Contaminant release from NAPL is modelled according to Raoult’s

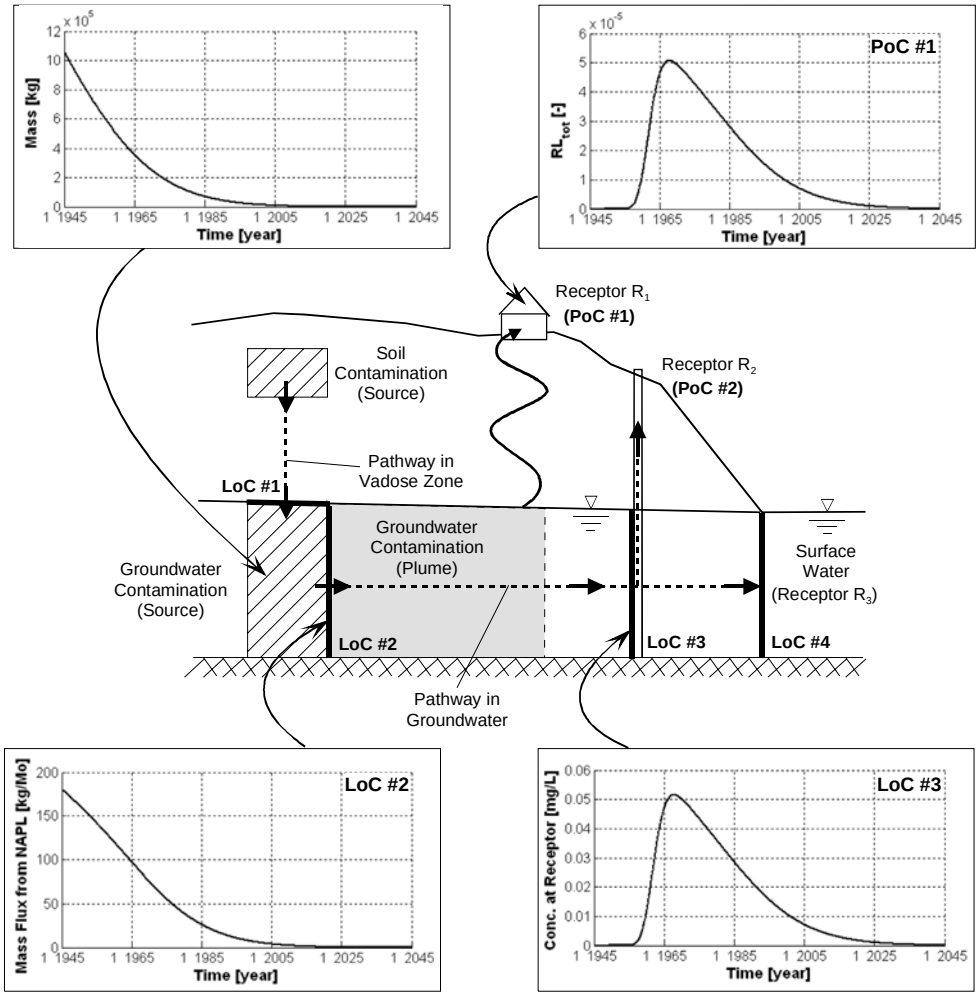


Figure 1. Source-pathway-receptor scheme with receptors R_i and options to set lines and points of compliance (LoC and PoC). Data plots show exemplary results for compound mass at source and selected LoCs or PoCs.

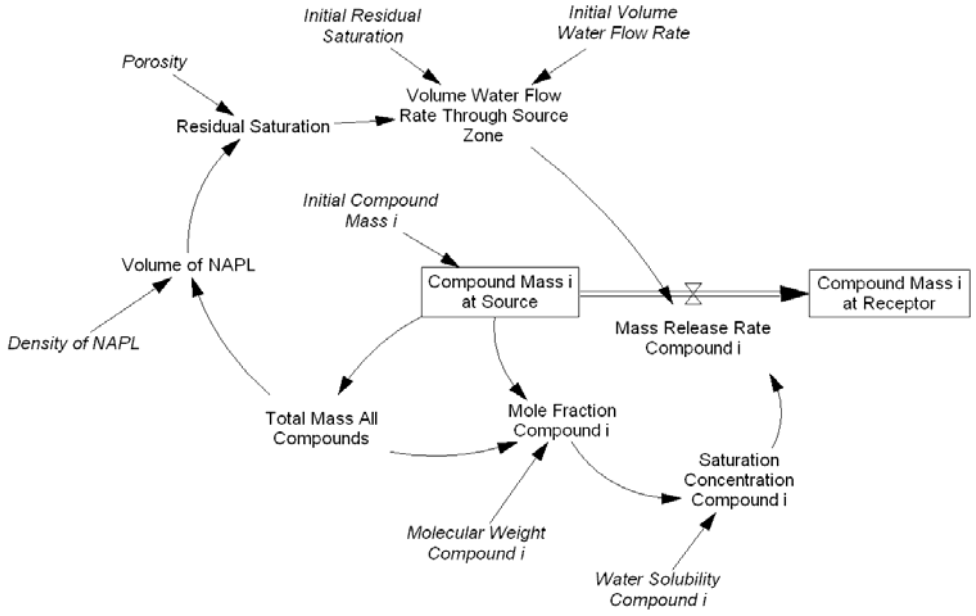


Figure 2. Major structure of contaminant release model according to Raoult's law.

law and features two feedback loops. A positive feedback loop reinforces contaminant release via a reduction of residual saturation, a negative feedback loop counteracts contaminant release via a lesser saturation concentration due to a reduced mole fraction. Several extensions to this "natural" contaminant release model quantify the effect of source remediation technologies, e.g. through an enhancement of the saturation concentration by surfactant application or thermal treatment.

2.3 Linking to Existing Models

The core module of CARO-plus, as shown in Figure 1, may be linked to an existing ordinary flow model in two ways. A straightforward option is to transfer key descriptive effective parameters of hydrogeology and source zone characteristics, as well as travel times between source zone(s) and receptor(s), from the flow model (incl. particle tracking) to CARO-plus. Figure 3 illustrates how larger areas can be structured into smaller units in the form of SPR sequences. Output data (e.g. contaminant flux, risk and cost profiles over time) may be communicated back to the flow model (or to a controlling GIS system if available). A second option is to integrate the core module as a real part of the existing model, i.e. to replicate it in each cell of the host model [e.g. Voinov et al., 1999].

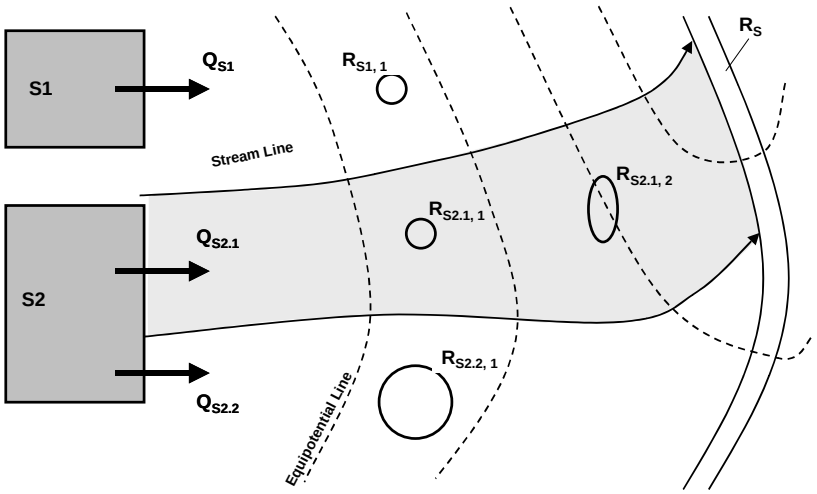


Figure 3. Structuring of a large-scale area into source-pathway-receptor sequences (one sequence shaded in grey). Concerned receptors may correspond to individual sequences (e.g. a domestic well $R_{S2.1,1}$) or to multiple sequences (e.g. surface water R_S).

3. CASE STUDY

The case study features the site of a former coking plant, located in Germany. The site exhibits an extensive contamination of both soil and groundwater, dominated by BTEX and PAH. The contaminations are mainly due to destruction during world war II. Figure 4 illustrates the situation at the site. The contamination is also vertically spread within the source zone, which is an LNAPL pool dominated by benzene, and a DNAPL pool, dominated by naphthalene. Starting from this source zone, one SPR sequence, considering the downgradient groundwater pathway to the residential area, was examined. Best guess parameter values and exposure pathways considered are listed in Table 1. The given best guess estimates are believed to best describe prevailing conditions on average or in terms of an effective parameter value. The estimates were derived based on expert knowledge rather than on statistics, since the limited available site data did not allow otherwise. This also applies to the type and parameters of the input parameter distributions used in the probabilistic analysis, which are listed in Table 2. These kind of rather rudimentary (e.g. uniform, triangular, etc) distribution assumptions are adequate if the analysis is primarily of an exploratory nature [e.g. Helton, 1993], as is the case in preliminary assessment. Note that uncertainty may also refer to performance or cost parameters of risk reduction

technologies, but were not considered in the scenarios presented here. We used Monte Carlo (MC) simulation to analyse the combined effect of variations in the inputs (Table 2). The Latin Hypercube sampling technique was selected to reduce computational time, and to produce more stable analysis outcomes than pure random sampling [Helton & Davis, 2003].

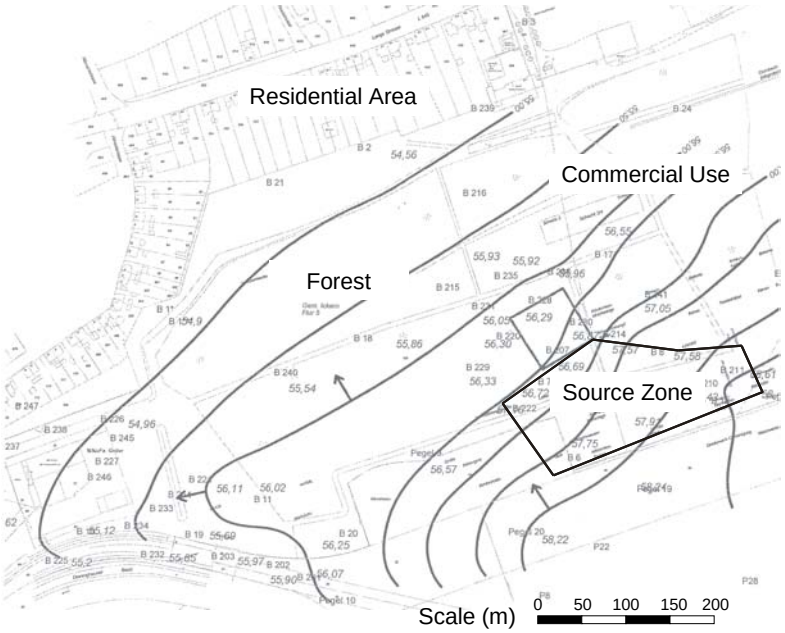


Figure 4. Map of the case study area with locations of potential receptors, labelled residential area, industrial forest and commercial use.

Table 1. Best guess values ("Estimate") of major model input parameters, considered pathways and target values for scenario "residential area".

Parameters	Estimate	Exposure pathways	Targets
Plume aquifer thickness [m]	5	Drinking water ingestion	Risk:
kf [m month^{-1}]	129.6	Leafy vegetable ingestion	$HQ_{tot} < 1$
i [-]	0.00625	Soil ingestion	$RL_{tot} < 1E\text{-}5$
NA rate GW [month^{-1}]	0.058	Inhalation of resuspended soil	Concentration:
Retardation [-]	4.3	Outdoor air inhalation	C < 0.01 mg/l
Today's average concentration in source zone [mg L^{-1}]	79	Indoor air inhalation	
Plume width receptor [m]	250		
Distance source/receptor [m]	300		

Table 2. Probabilistic model input parameters.

Parameters	Estimate	Distribution	Min	Max	Start	Peak	Stop
Plume aquifer thickness [m]	5	Uniform	4	6			
Conductivity K [m month^{-1}]	129.6	Uniform	100	160			
Hydr. Gradient i [-]	0.00625	Uniform	0.004	0.008			
NA rate GW [month^{-1}]	0.058	Triangular	0.0024	0.114	0.0024	0.058	0.114
Retardation factor R [-]	4.3	Uniform	1.8	6			
Today's average concentration in source zone [mg L^{-1}]	79	Triangular	14.32	272	14	142	670
Distance source/receptor [m]	300	uniform	200	300			

4. RESULTS

The results from a first deterministic screening analysis based on best guess values of all model input parameters indicate that the risk peak has already passed. The predicted values of both concentration and risk level do slightly exceed the target values (Figure 5). The preliminary assessment therefore suggests to conduct further and more detailed investigations to obtain more data and information. Depending on what the new data will reveal, a second screening-level assessment or a more sophisticated modelling should then be performed. The planning of new investigation campaigns may be supported by an early assessment of risk mitigation options in order to get a first idea of feasible options. This may help to set the focus of new investigations on gaining knowledge specifically regarding the most feasible technological options. As an example, Figure 5 shows the effects we may expect from e.g. a pump-and-treat measure. Here, an operation time of 21 years (from 1/2000 until 12/2020) is assumed. Since the pump-and-treat system is hypothetically considered to be a protection measure, the system's location is set to 50 m upgradient of the receptor. As an addition to the deterministic analysis, risk and risk mitigation options were also assessed using a probabilistic method, as described above. Some example results are shown in Figure 6 for the "natural" system (no action scenario) as well as with respect to the effects of the abovementioned pump-and-treat measure.

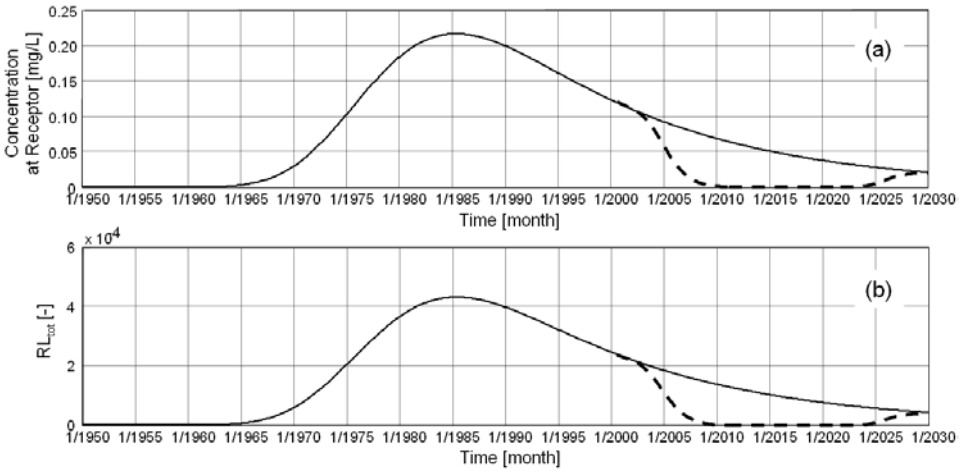


Figure 5. Deterministic results for receptor "residential area": (a) concentration in groundwater at the receptor, (b) risk level at the receptor. Dashed lines indicate the effect of a pump-and-treat measure between 1/2000 and 12/2020.

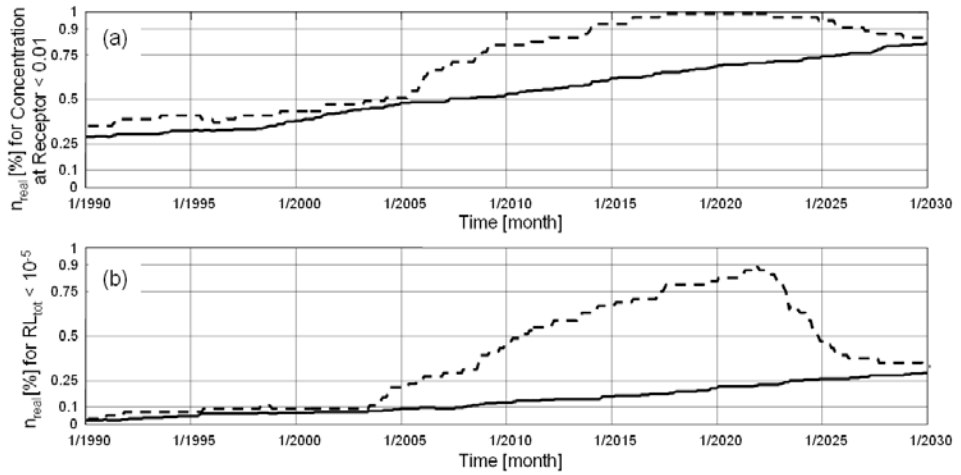


Figure 6. Probabilistic results for receptor "residential area": (a) probability to reach the groundwater concentration target at the receptor, (b) probability to reach the target level of risk at the receptor. Dashed lines indicate the effect of a pump-and-treat measure between 1/2000 and 12/2020.

5. CONCLUSIONS

It is has been shown for a pilot case study in Germany that there is a large potential of preliminary assessment for contaminated land management. The presented assessment tool CARO-plus is capable of analysing the existing and future situation at a site with respect to risks, as well as regarding the effects of possible risk mitigation options in terms of contaminant mass flux, concentration and risk indices. Costs of mitigation options are also estimated, but haven't been shown here. A special emphasis is given to the uncertainty of predictions using Monte Carlo simulations.

ACKNOWLEDGEMENTS

The authors wish to thank the German Ministry for Education and Research (BMBF) for the financial support as part of the research priority program KORA. Funding has also been provided by the Helmholtz Centre for Environmental Research – UFZ in Leipzig.

REFERENCES

ASTM – American Society of Testing and Materials, Standard Guide for Risk-Based Corrective Action Applied at Petroleum Release Sites. Report E1739-95, 2002.

Forrester, J. W. Industrial Dynamics, Cambridge, Massachusetts: MIT Press., pp. 340., 1961.

isee systems, STELLA version 9, isee systems, Inc., Lebanon (NH), U.S.A., 2007.

Helton, J.C., Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, *Reliability Engineering and System Safety*, 42, 327-367, 1993.

Helton, J.C., Davis, F.J., Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability Engineering and System Safety*, 81, 23-69, 2003.

Khan, S., Yufeng L., Ahmad, A., Analysing complex behaviour of hydrological systems through a system dynamics approach, *Environmental Modelling & Software*, in press, 2008.

McKnight, U.S., Finkel, M., Model-Based Preliminary Assessment of Mega-Site Management Options: A New Approach to Improve Tiered Decision-Making, paper presented at Groundwater Quality 2007: Securing Groundwater Quality in Urban and Industrial Environments, to be published in IAHS Red Books series, 2008a.

McKnight, U.S., Finkel, M., Development of a screening-level model and DSS to facilitate the assessment of contaminated land management options, in preparation, *Environmental Modelling & Software*, 2008b.

Simonovic, S., World water dynamics: global modeling of water resources, Journal of Environmental Management, 66, 249-267, 2002.

Strenge, D.L. and Smith, M.A., Multimedia Environmental Pollutant Assessment System (MEPAS): Exposure Pathway Module Description. Prepared for the U.S. Department of Energy, Contract DE-AC06-76RLO 1830, 2006.

UK Environment Agency, Methodology for the Derivation of Remedial targets for Soil and Groundwater to Protect Water Resources, R&D Technical Report P20, 1999.

Ventana Systems, Vensim Simulation Environment – User's Guide, Revision date: July 4, 2007, Ventana Systems, Inc., Harvard (MA), U.S.A., 2007.

Voinov, A., Costanza, R., Wainger, L., Boumans, R., Villa, F., Maxwell, T., Voinov, H., Patuxent landscape model: integrated ecological economic modeling of a watershed, *Environmental Modelling & Software*, 14, 473–491, 1999.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Multiple-criteria decision making for integrated watershed management in the Ramian watershed, Golestan, Iran

Amir Sadoddin[a, *], Vahedberdi Sheikh[a], Raoof Mostafazade[b], Mohammad Ghassem Halili[b]

[a] *Academic staff of Gorgan University of Agricultural Sciences and Natural Resources, amir.sadoddin@gmail.com; v.sheikh@yahoo.com*

[b] *MSc student at Gorgan University of Agricultural Sciences and Natural Resources*
[*] *Corresponding author*

Abstract: This research has concentrated on the physical and socio-economic impacts of vegetation-based management scenarios targeting on flooding and soil erosion issues in the Ramian sub-watersheds. The Ramian watershed with an approximate area of 24000 ha is located in Golestan, Iran. For each sub-watershed, four biological actions (activities) and 16 management scenarios have been considered. Physical impacts have been studied using the SCS (Soil Conservation Service) hydrologic model and the EPM (Erosion Potential Method) model. Economic and social impacts have been evaluated using the cost/benefit analysis as well as the examination of the results of a social survey. Some physical, social, and economic indices were chosen and quantified. The indices then were standardised using the interval standardisation technique. Best scenarios were determined using a multiple-criteria decision making technique. To weight the indices, four perspectives were used. For the first three perspectives one of the physical, social and economic criteria were assigned higher weights while for the fourth perspective all criteria were given the same weight. Trade-off analysis of the results indicates that for most sub-watersheds more than one single management solution can be recommended on the basis of the different perspectives. The results showed that the multiple-criteria decision making serves as a valuable tool to represent the watershed system as a whole, to incorporate output from models and expert-judgments to examine the trade-offs among outcomes necessary to decision making and to communicate uncertainty of the parameters.

Keywords: integrated watershed management; multiple-criteria decision making; vegetation-based scenarios; the Ramian watershed

1. INTRODUCTION

Globally and particularly in developing countries, population growth and demands for food security resulted in improper use of soil and water resources. Conversions of upland forest areas into croplands have lead to accelerated soil degradation and the depletion of soil productivity. Flooding, water pollution, and household socio-economic welfare problems are

the other negative consequences of inappropriate use of soil and water resources. Integrated watershed management is globally accepted as a sound approach for water, land, and related resources which takes care of equilibrium between socio-economic demands of watershed inhabitants and ecosystem sustainability. The main aim of this study is to develop and demonstrate an approach which is used for integrated management of watersheds. Synthesis of problems, driving factors, biophysical and socio-economic perspectives, watershed-dependent communities, data and models of different scales are elements of integrated watershed assessment and management [Jakeman et al., 2005]. Detailed understanding and trade-off analysis of results from implementation of different management scenarios at various spatial and temporal scales will improve decision making. Sadoddin et al. [2003] using the Bayesian decision networks studied the socio-economic and biophysical impacts of biological scenarios for salinity management in Australia and presented the best management scenarios. Their results showed that the Bayesian decision network serves as a valuable tool to represent the watershed system as whole, to incorporate output from models and expert judgments to examine the trade-offs among outcomes necessary for decision making and to communicate uncertainty of the parameters. Alweshah et al. [1999] conducted a regional flood analysis using WMS (Watershed Modelling System) model in Jordan and compared three scenarios of forestation, terracing and check-dams construction. They concluded that the forestation is the best scenario which leads to peak discharge reduction as well as total volume discharge reduction up to 70 percent.

2. STUDY AREA DESCRIPTION

The Ramian watershed, a sub-basin of the Gorganrood River basin, is located in the eastern part of the Golestan Province, Iran. It has an area of about 240 km^2. The geographic position lies between 55^o 02' – 55^o 18' E longitude and 36^o 49' – 37^o 02' N latitude. Elevation ranges from 220 to 2890 m MSL. Forest is the dominant land cover type in the watershed. Mean annual precipitation is 898 mm and mean daily temperature is 16.5 ^oC. According to the de Martonne and Emberger classifications the climate of the study area is humid and cold humid, respectively [Tajiki, 2007]. The Ramian town with an approximate population of 80,000 is located in the immediate downstream of the watershed's outlet. Considering the high flooding susceptibility of the watershed and socio-economic importance of areas exposed to inundation, flood mitigation measures are necessary to be planned and implemented. Furthermore, the susceptible geological formations, water erosion, intense road construction and land use conversion from forests to croplands in some steep areas have lead to on-site soil degradation and off-site sedimentation in downstream [Shomal Consulting Engineering Inc., 2007]. In order to cease the depletion of watershed resources, in this study a scenario-based approach has been used to predict the impacts of different management activities within an integrated watershed management framework. In addition, a trade-off analysis of the results has been carried out to provide a proper basis for decision making.

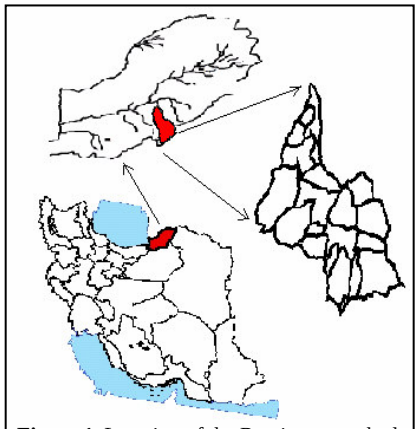


Figure 1. Location of the Ramian watershed

3. METHODOLOGY

3.1 Watershed management with scenario analysis approach

Contrary to mathematical optimization models, scenario-based approaches increase the insight of watershed inhabitants regarding watershed system [Cain, 2001]. Scenario-based approaches allow the users to choose different management scenarios and evaluate their possible positive

and/or negative outcomes. Using a scenario-based approach in this study is consistent with the intention of supporting decision makers rather than making decision for them [Cain, 2001]. This procedure improves our knowledge about watershed system and its behavior and helps to identify the best management scenario.

3.2 Conceptual model for integrated management of the Ramian watershed

In this study a conceptual model representing the cause and effect relationships between variables and comparing the impacts of management scenarios on biophysical and socio-economic characteristics of the watershed system have been integrated (Figure 2).

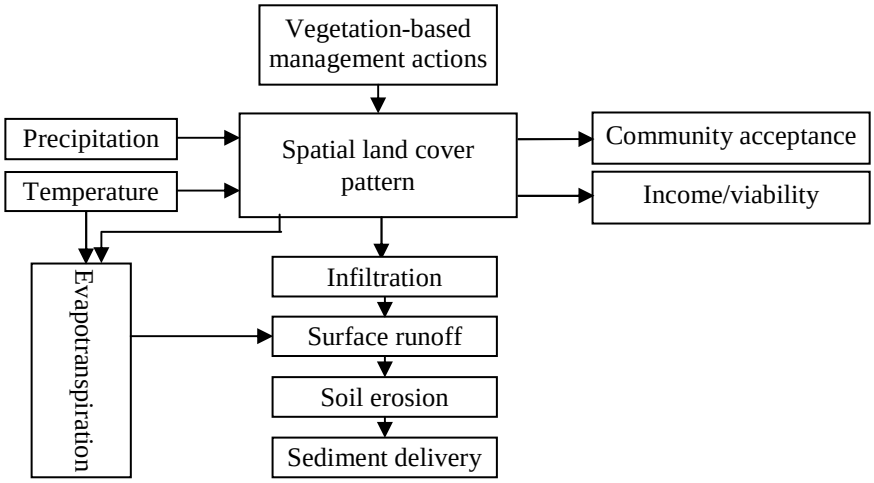


Figure 2. The conceptual model framework for vegetation-based management of the Ramian watershed

3.3 Developing of a mutually exclusive vegetation-based management scenarios

First the sources of surface runoff and sediment problems over the Ramian watershed system and their relative importance are identified. Then a list of all feasible solutions for mitigation or elimination of these problems is prepared. Maintaining the current condition can sometimes be a solution for a watershed to recover itself through natural evolution particularly once the watershed disturbance is not extended in a large scale [Heathcote, 1998]. Furthermore it can be used as a base case scenario to evaluate the other scenarios. After identifying the contemporary management activities in the study area, the possible management actions were defined considering the existing constraints in the watershed. For the Ramian watershed four biological actions were considered including: agro-forestry, tree plantation, seeding, and sowing. Combination of these four actions leads to 16 (2^n) different management scenarios (Table 1).

Table 1. Vegetation-based management scenarios for the Ramian watershed

Scenario	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
Agro-forestry	-	+	-	-	-	+	-	-	+	+	-	+	-	+	+	+
Seeding	-	-	+	-	-	+	+	-	-	-	+	+	+	+	-	+
Sowing	-	-	-	+	-	-	+	+	+	-	-	+	+	-	+	+
Tree planting	-	-	-	-	+	-	-	+	-	+	+	-	+	+	+	+

+ Sign indicates the presence, and - Sign indicates the absence.

It should be mentioned that the management scenarios should be exclusive. In other words, admission of one scenario leads to refusal of other scenarios. Regarding the fact that most croplands of the study area are cultivated with wheat and there is a tendency among the watershed community for establishing fruit trees, a walnut-wheat agro-forestry system has been proposed as an action.
Input map layers including sub-watersheds boundary, hypsometry, slope, vegetation cover type and density, soil depth, and hydrologic soil groups were prepared and superimposed within the ArcGIS environment in order to specify the spatial distribution of various management activities considering the scenario development rules shown in Table 3. In developing the scenarios, 100 % of suitable areas for each vegetation-based activity have been allocated for it.

Table 2. Scenario rules for vegetation-based management of the Ramian watershed

Biological	Suitable areas
Agro-forestry	Croplands with slope less than 40% and semi-deep to deep soils
Tree planting	1) barren lands, shallow soils, low density vegetation, slopes less than 60%, elevations up to 1600 m above MSL 2) forests, semi-deep soils, moderate vegetation density, slope less than 60% elevation up to 1600 m above MSL
Seeding	Forests with semi-deep soils and moderate vegetation density
Sowing	Forests with deep soils and moderate vegetation density, slope more than 60%

4. ANALYSIS OF INTEGRATED MANAGEMENT OF THE RAMIAN WATERSHED

4.1 Modelling the physical impacts of vegetation-based scenarios

The SCS model was used to simulate the effects of vegetation cover changes on hydrological characteristics. This method can relate the watershed characteristics to the flow parameters. The Curve Number (CN) is calculated based on hydrologic soil group, antecedent soil moisture condition, and the land use type. The outputs of the model are rainfall excess depth, peak discharge, and time to peak over the watershed with spatial resolution of sub-watershed unit. The initial abstraction and surface runoff depth are calculated with equations 1 and 2, respectively.

$$S = \frac{25400}{CN} - 254 \qquad (1) \qquad R = \frac{(P - 0.2S)^2}{P + 0.8S} \qquad (2)$$

Where, S is initial abstraction in mm; CN is curve number; P is rainfall in mm and R is rainfall excess in mm.
Peak discharge is calculated with equations 3 and 4.

$$Q_p = \frac{2.083AR}{t_p} \qquad (3) \qquad t_p = \frac{D}{2} + t_l \qquad (4)$$

Where, Q_p is peak discharge in m^3/sec; A is watershed area in ha; R is rainfall excess in mm; t_p is time to peak in hour; D is duration of rainfall excess in hour and t_l is watershed lag time in hour. The 50 years discharge volume for each sub-watershed was calculated by multiplying the 50 years peak discharge and the respective time to peak.
The EPM model has been used to predict the effects of vegetation cover changes on soil erosion rate. The coefficient of erosion (see Equation 5) of the model has been used as a surrogate index of soil erosion rate.

$$Z = Y.X.a.(\Phi + \sqrt{I}) \qquad (5)$$

Where, Y is coefficient of soil resistance to erosion; X is land use coefficient; 'a' is conservation coefficient; Φ is coefficient of the observed erosion process; and I is mean slope of the surface. In order to predict the soil erosion changes in response to each scenario the product of land use coefficient (X) and protection coefficient (a) was calculated and compared with its value at the present condition.

4.2 Modelling the economic impacts of vegetation-based management scenarios

Gross margin and variable cost have been used as indices to predict the effects of vegetation cover changes on economic conditions. For each scenario total gross margin is calculated with Equation 6.

$$G = \sum_{i=1}^{n} [P_i Y_i - C_i] A_i \qquad (6)$$

Where, Y_i is yield of crop *i* (unit of production per unit of area); P is the price of crop (Iranian Rials per unit of production); C is variable costs of crop i (Iranian Rials per unit area); A is the area devoted to crop i (unit area) and n is the number of activities.
To assess the economic impacts of vegetation-based management activities, the decision horizon was considered to be 80 years in order to reflect the time span required for the planted trees to be matured to the stage of industrial harvesting.
Cash flow during 80 years period was translated to the present value. In this study discount rate (the combination of inflation rate and interest rate) was assumed 5%. The documented values in the Biological Study Report of the Ramian watershed along with economic expert-knowledge were utilized to estimate the input parameters.

4.3 Modelling the social impacts of vegetation-based management scenarios

The number of stakeholders to be consulted will vary in relations to the scope and breadth of the management questions that will be addressed. A compromise is needed between costs, time, practicality, and diversity and dynamism of stakeholders [Baran and Jantunen, 2004]. In this study 22 farmers were consulted as a sample of the watershed community to conduct a social survey in order to evaluate the acceptance level of the management scenarios among the community. Social survey participants were enquired about their intentions to implement the vegetation scenarios in the 5 years ahead (starting from 2006). The results of the social survey were used to analyse the likely social outcomes of attempt to implement the management scenarios in the watershed. To this end, the binomial probability distribution was used. In the class of the binomial probability experiments, the trials are independent with the probability of acceptance (*p*) the same for y trial [Harshbarger and Reynolds, 1989]. The probability of *y* successes in *n* trials is calculated by Equation 7.

$$P(y_i) = \frac{n!}{y_i!(n-y_i)!} p_i^{y_i} q_i^{n-y_i} \qquad (y_i = 0,1,2,...,n) \qquad (7)$$

Where, *n* is the number of trials (22 participants); P_i is the probability of acceptance (positive answer) of the scenario *i* in each trial; q_i is probability of non-acceptance (negative answer) of the scenario *i* in each trial; y_i is the number of acceptance of the scenario *i* in *n* trials; $P(y_i)$ is the probability of y_i acceptance in *n* trials; and i is scenario number (1, 2, …, 16). To analyze the community acceptance towards vegetation-based scenarios, four levels of acceptance (no acceptance, low acceptance, moderate acceptance and high acceptance) were used. The following assumptions were used to assign the level of acceptance. If 2 out of 22 trials were positive, the management scenario was considered to be rejected by the community (no-

acceptance), if 3 – 8 trials were positive, it is considered as low acceptance level, if 9 – 16 trials were positive, it is considered moderate acceptance level, and if more than 17 trials were positive, it is considered as high acceptance level. Table 3 represents the probability of acceptance of the 16 scenarios. It should be mentioned that p_i is the probability of acceptance identified when financial supports such as subsidies, interest-free loans, and other incentives are provided for improvement of vegetation condition in the watershed.

Table 3. Probability of acceptance of the management scenarios

Scenario	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**
Pi	0.45	0.32	0.23	0.23	0.32	0.18	0.33	0.09	0.14	0.18	0.09	0.14	0.14	0.09	0.09	0.09

4.4 Multiple-criteria decision making for integrated watershed management

Some researchers attempt to translate all outcomes affiliated to different criteria even the environmental and social ones into the economic outcomes. However in some circumstances due to shortcomings of this approach, multi-dimensional approach is applied in which all criteria are evaluated separately. In other words, the best scenario is selected considering all criteria collectively. Since criteria are of different nature, they need to be standardised first. This can be done using the interval weighting technique, where each index value is standardised to a range between 0 and 1 and each standardised value is multiplied by its respective weight. The weighted sum then determines the best scenario [Sharifi, 2004]. In the interval weighing technique, the indices are categorized into two groups: benefits and costs. Accordingly, one of the equations 8 and/or 9 is used for standardisation purposes [Sharifi, 2004].

For benefit group:

$$S\tan dardised \quad value \quad of \quad an \quad index = \left(\frac{score - lowest \quad score}{highest \quad score - lowest \quad score} \right) \qquad (8)$$

For cost group:

$$S\tan dardised \quad value \quad of \quad an \quad index = 1 - \left(\frac{score - lowest \quad score}{highest \quad score - lowest \quad score} \right) \qquad (9)$$

5. RESULTS: TRADE OFF ANALYSIS OF THE RESULTS FROM IMPLEMENTING MANAGEMENT SCENARIOS

As mentioned earlier the indices value corresponding to each management scenario were standardised. Different weights were assigned to the indices based on four different perspectives that might be chosen by the stakeholder in the Ramian watershed (see Table 4).

Table 4. Weights assigned to the indices for different perspectives

Perspective number	**Perspective of weighing**	**Variable costs**	**Gross margin**	**Community acceptance**	**Soil erosion**	**Total flood discharge**
1	all equal	20	20	20	20	20
2	economic	30	30	20	7.5	12.5
3	social	10	10	60	7.5	12.5
4	physical	10	10	20	25	35

In the next step, the standardised values of indicators were multiplied by their weights and summed up to determine the best scenario(s). Prioritisation of scenarios in different weighing perspectives has been shown in tables 5 to 8.

Table 5. List of best management scenarios given equal weights for all criteria

Preference	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Sub11	Sub12	Sub13	Sub14	Sub15	Sub16
First	S1	S3	S2	S2	S3	S13	S1	S6	S7	S7	S5	S2	S2	S14	S16	S7
Second	S5	S2	S14	S16	S7	S3	S3	S2	S4	S4	S15	S12	S16	S16	S14	S4
Third	S2	S1	S16	S12	S14	S7	S7	S16	S9	S9	S9	S16	S15	S2	S2	S12

Table 6. List of best management scenarios based of the economic perspective

Preference	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Sub11	Sub12	Sub13	Sub14	Sub15	Sub16
First	S1	S2	S2	S2	S3	S3	S1	S2	S7	S7	S9	S2	S2	S2	S2	S1
Second	S2	S1	S3	S9	S7	S7	S3	S6	S6	S4	S2	S12	S12	S12	S12	S7
Third	S7	S13	S12	S12	S6	S6	S7	S12	S11	S9	S1	S5	S9	S6	S6	S4

Table 7. List of best management scenarios based of the social perspective

Preference	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Sub11	Sub12	Sub13	Sub14	Sub15	Sub16
First	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1
Second	S5	S2	S2	S2	S2	S2	S2	S2	S5	S5	S5	S2	S2	S2	S5	S2
Third	S2	S5	S5	S5	S5	S5	S3	S5	S7	S7	S2	S5	S5	S5	S2	S5

Table 8. List of best management scenarios based of the physical perspective

Preference	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Sub11	Sub12	Sub13	Sub14	Sub15	Sub16
First	S5	S3	S16	S16	S14	S13	S8	S16	S5	S5	S5	S2	S16	S14	S16	S12
Second	S10	S5	S14	S15	S11	S5	S1	S14	S7	S7	S16	S16	S15	S16	S14	S9
Third	S16	S2	S15	S5	S5	S9	S5	S15	S16	S4	S15	S14	S5	S5	S5	S7

6. DISCUSSION AND CONCLUSIONS

In using the scenario-based approach, it is necessary to use models which are able to predict the impacts of implementing different scenarios on watershed scale. The results of this study indicate that SCS and EPM models are capable of predicting the impacts of vegetation changes on total discharge and soil erosion rate, respectively. The application of the binomial distribution in the social impact analysis provides a probability distribution function for analysing the community attitudes towards each scenario. By putting emphasis on one of the economic, social, and physical criteria watershed managers and watershed communities choose best

scenario(s) in each sub-watershed. In case of encountering a difficulty in identifying a preferenced criterion, the best scenario is derived based on the equal weights for all criteria. Trade-off analysis of the results shows that when social criteria are emphasised, in most sub-watersheds scenario 1 gets the highest score followed by scenario 2 (agro-forestry). This indicates that the communities of the Ramian watershed are unwilling to extend forest areas. Choosing criteria depends directly on the national and regional strategies. Therefore, to use the findings of this study, these strategies should be considered. This warranties the feasibility and suitability of the results.

As shown in Tables 8 to 11, for most sub-watersheds more than one unique management solution is recommended given the different perspectives. The approach used in this study allows decision makers and/or stakeholders of the Ramian watershed to reach their own conclusions on the basis of their improved understanding of the system and of the trade-offs among various outcomes arising from implementing management scenarios. This approach is intended to support decision makers in their understanding of trade-offs rather than providing them with a single "optimal" decision. The approach also allows account to be taken of other possible factors not included in the modelling framework for the watershed. This approach occurs with the fact that watershed systems are dynamic and complex and it is a difficult task to capture all of the disciplinary components involved in management of natural resources watershed-wide. To further extend this study and to provide a comprehensive picture of the watershed system ecological criteria can be incorporated in the conceptual modelling framework of this study.

REFERENCES

Al-weshah, R.A., and F. El-Khoury, Flood analysis and mitigation for Petra area in Jordan, *Journal of Water Resources, Planning and Management*, Vol.125 (3):170-177, 1999.

Barran, E., and T. Jantunen, Stakeholder consultation for Bayesian decision support systems in environmental management. Regional Conference on Ecological and Environmental Modeling. (ECOMODE), Penang Malaysia, 2004.

Cain, J., Planning improvements in natural resources management. Wallingford, Oxon, UK, Centre for Ecology and Hydrology, pp. 124, 2001

Gavrilovic, Z., M. Stefanovic, M., Cotric and J. Milojevic, Erosion potential method: an important support for integrated water resource management. Institute Development of Water Resources, pp. 14, 2004.

Harshbarger, R.J., and J.J. Reynolds, Mathematical applications for the management, life and social sciences, Lexington, Massachusetts, D.C. Heath and Company, pp. 1184, 2003.

Heathcote, I.W., Integrated watershed management, John Wiley and Sons, pp. 414, 1998.

Jakeman, A.J., R.A. Letcher, S. Rojanasoonthon, S. Cuddy, and A. Scott, Integrating knowledge for river basin management: progress in Thailand. ACIAR Monograph No. 118, pp. 220, 2005.

Sadoddin, A., R.A. Letcher, and A. J. Jakeman, A Bayesian decision network approach for salinity management in the Little River catchment, MODSIM Congress, Australia, 2003.

Sadoddin, A., Bayesian network models for integrated-scale management of salinity. PhD thesis, Australian National University. pp. 264, 2006.

Sharifi, A., M. Herwijnen, and W. Torren, Spatial decision support systems. International-Institute of Geo-information Science and Earth Observation. (ITC), pp. 153, 2004.

Shomal Engineering Consulting Inc., Study Report of the Ramian watershed, Volume9 (Synthesis), pp. 206, 2007.

Tajiki, M, Evaluation of watershed management practices on flooding and sediment delivery characteristics (case study: the Ramian watershed), MSc. thesis, Gorgan University of Agricultural Sciences and Natural Resources, Iran, pp.138, 2007.

Wanielista, M.P., Hydrology and water quality control. University of Central Florida, pp. 565, 1990.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

A Toolbox for the Identification of Parsimonious Semi-Distributed Rainfall-Runoff Models: Application to the Upper Lee Catchment

B. Orellana[a], I.G. Pechlivanidis[a], N. McIntyre[a], H.S. Wheater[a] and T. Wagener[b]

[a] *Department of Civil and Environmental Engineering, Imperial College London, SW7 2AZ, London, UK (a.orellana02/i.pechlivanidis05/n.mcintyre/ h.wheater@imperial.ac.uk)*

[b] *Department of Civil and Environmental Engineering, The Pennsylvania State University, 226B Sackett Bldg., University Park, Pennsylvania 16802, USA (Thorsten@engr.psu.edu)*

Abstract: In operational hydrology identification of an appropriate model structure and suitable parameter sets for a specific catchment is a challenging task. This identification process is often based on data availability, catchment characteristics and modelling objectives, and will often result in a range of different model structures. This process of model identification becomes even more challenging when moving from lumped to distributed models as the potential number of model parameters increases proportionally to the number of spatial units considered, and due to the existence of ungauged spatial units. A Semi-Distributed Rainfall-Runoff Modelling Toolbox (RRMT-SD) has been developed to estimate continuous streamflow at points along the river system using conceptual and hybrid representations of the rainfall-runoff processes which vary from low to medium model complexity. The user can easily implement different model structures and calibration strategies, considering multiple objective functions and Monte Carlo analysis. To show the potential of the toolbox a case study on the Upper Lee catchment (1040 km^2), UK, using hourly time-steps is presented. The study area was divided into gauged subcatchments and each one of them represented through smaller spatial units of similar areas. Different model structures were applied on the spatial units using estimated a priori parameter values based on a simple regression method. The models were calibrated using spatial multipliers to adjust the a priori parameter values to the scale of the spatial units. Results showed that for different types of subcatchments (low and high base flow types) two soil moisture model structures (the Probability Distributed Moisture Model and the Catchment Wetness Index, respectively) were justified, and that parsimonious semi-distributed rainfall-runoff models on the Upper Lee catchment can perform reasonably well for a single criteria (e.g. average NSE values of 0.74).

Keywords: Rainfall-runoff models; RRMT; Semi-distributed; Parsimonious modelling; Calibration; Regionalisation.

1. INTRODUCTION

Although conceptual rainfall-runoff (RR) models are widely utilised in practice, they usually suffer from a lack of identifiability [Beven, 2001]. Research has showed that different model structures or different parameter sets within the feasible parameter space can produce virtually indistinguishable simulated streamflows [e.g. Duan et al., 1993, Spear, 1995]. Model identifiability can be improved by reducing model complexity in terms of number of model parameters considered so that all of them can be identified from the available data (parsimonious modelling) [Wheater, 2002]. There are many examples of using parsimonious, spatially lumped models, to provide reliable streamflow estimations

[e.g. Perrin et al., 2001, Wagener, 2002, among many others]. Spatially distributed models, compared to lumped, have the potential to estimate streamflow and other state variables at interior catchment points. The literature reports numerous distributed RR models and modelling approaches that address a wide variety of issues as prediction and operational forecasting environments [e.g. Turcotte et al., 2003]. Although the use of distributed models has been encouraged for many different reasons, e.g. availability of distributed rainfall data, and data on distributed catchment properties, there are issues that emerge such as scale and parameter estimation [Wooldridge et al., 2001], but also calibration strategy [Carpenter and Georgakakos, 2006]. Particularly, as the number of model parameters increases with the degree of spatial discretisation, distributed models can easily become overparameterised and subsequently ill-posed with respect to the available input-output data. Thus uncertainty in parameter estimates and hence uncertainty in model identification is a common problem [Madsen et al., 2002].

Semi-distributed RR models have been suggested to combine the advantages of lumped and distributed approaches. They can be considered as simpler distributed models that apply conceptual or hybrid metric-conceptual RR models to spatial units smaller than the catchment area. This paper presents a Semi-distributed Rainfall-Runoff Modelling Toolbox (RRMT-SD) that has been developed to estimate streamflow at points along the river system using parsimonious conceptual model structures. As an example, RRMT-SD is applied to 8 years of hourly historic data from the Upper Lee catchment, UK. Several model structures were applied and different calibration strategies investigated.

2. THE SEMI-DISTRIBUTED RAINFALL RUNOFF MODELLING TOOLBOX

The RRMT-SD has been developed to estimate continuous streamflow at points along the river system. It allows efficient building and evaluation of spatially semi-distributed rainfall-runoff models. The catchment is conceptualised through a network of subunits and a parsimonious model structure is applied to each subunit. RRMT-SD is an extension of the generic Rainfall-Runoff Modelling Toolbox (RRMT) [Wagener et al., 2002] developed previously to produce parsimonious lumped model structures estimating streamflow at the catchment outlet. RRMT has been applied for prediction and to investigate model identifiability in gauged and ungauged catchments [e.g. Wagener and McIntyre, 2005, Lee et al., 2006].

The user builds up a rainfall-runoff model on each subunit (Fig. 1(A)) by selecting from a list of soil moisture accounting (SMA) models, routing models and channel routing models. The user also has the ability to add new modules providing additional flexibility. The SMA module determines effective rainfall (ER), actual evapotranspiration (AET) and an estimation of soil moisture status; the runoff routing module considers a fast and slow runoff routing; and the channel routing module estimates discharge at the outlet of the subunits. The SMA and runoff routing models available correspond to the ones use within the established RRMT framework. Topologically, RRMT-SD simulates streamflow for the uppermost stream subunits first and accumulates the flow down the channel network.

Each subunit is considered as lumped, with the hydrological processes and climatologically forcing data homogeneous within each one of them. Then the degree of spatial distribution is represented mainly through the number of subunits, which can be subcatchments or hydrological response units.

Using the toolbox interfaces the user defines: (1) the catchment network specifying the number of subunits to consider and the corresponding identification number for each one of them; (2) the mode of the modelling environment, i.e. calibration, simulation or uncertainty, and the modelling timescale in seconds per time step which must be consistent with the temporal resolution of the input-output data; (3) the model structure for each subunit which can be specified individually for each subunit or considering homogeneity across subunits, i.e. the same model structure and parameter values. The toolbox allows for different optimisation methods for calibration: a uniform random search, the Shuffled Complex Evolution [Duan et al., 1992], and local nonlinear multi-constrained method based on simplex searching [Cormen et al., 2001].

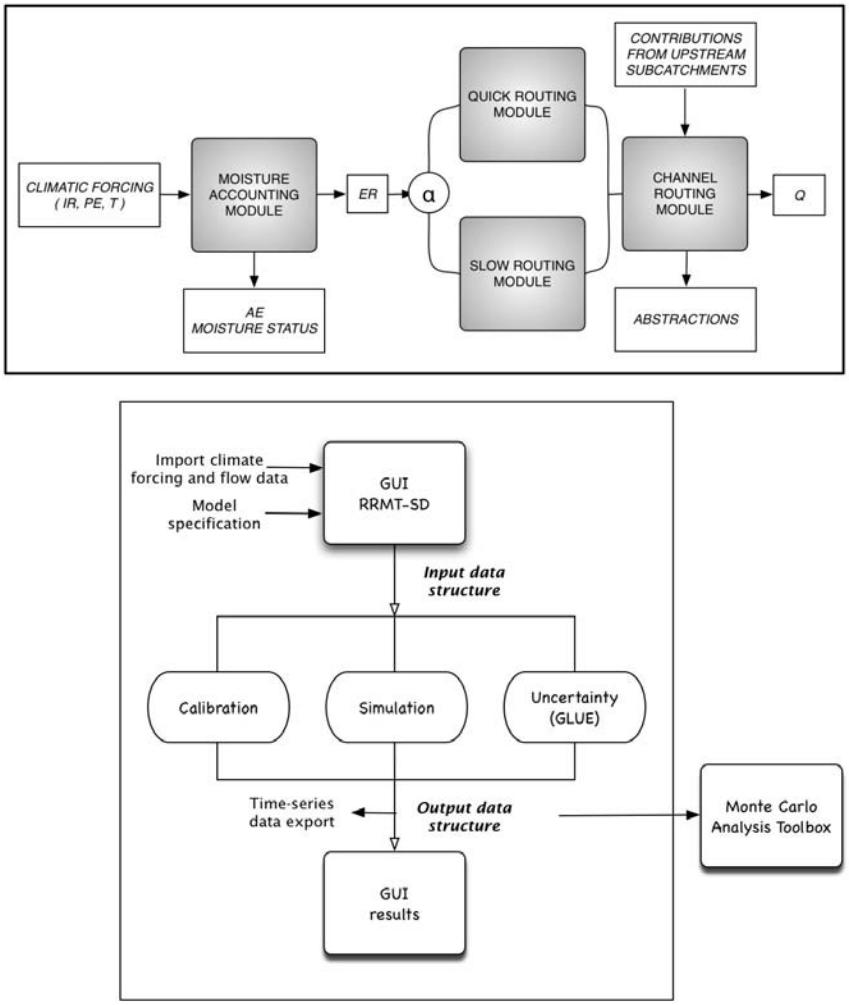


Figure 1. (A) Subcatchment module structure; (B) General Toolbox representation

The uniform random search can be used considering the same model structure and parameter sets among the subunits (URS-LP) or applying independent model structure and/or parameter sets (URS-SD). Also the user can use a priori parameter estimates (which may be non-uniform across subunits) and calibrate the model structures using calibration multipliers (URS-MP), which can be uniform or not across subunits. Multiple objective functions for measuring calibration and validation performance can be selected which are computed in all subunits where streamflow data are provided. Calibration of ungauged subunits is based on the closest downstream gauged outlet.

After finishing the modelling computations the input data and simulated variables in every subunit can be analyzed by selecting a variety of visualisation tools: input and output time series, and estimated state variables (i.e. rainfall, effective rainfall, streamflow, soil wetness, potential and actual evapotranspiration), water balances, flow duration curve, double mass plot, residual analysis to identify trends, normality and autocorrelation, and plots which show sensitivity of outputs over the parameter space.

RRMT-SD has so far been applied in a range of catchment types. A semi-distributed IHACRES model was applied to hourly data from the 734 km^2 Wadi Ahin catchment in Oman. This arid zone catchment was split into 20 topographically defined subunits in order to represent the strong spatial variability in input rainfall, and associated travel time distributions in the channel network, while parameters were assumed uniform across the catchment. The model was used for flash flood peak and volume estimation. RRMT-SD has also been applied to the small Rhosasflo catchment (1.3 km^2) in Wales, UK. Here, the 73 subunits were fields of average area 0.018 km^2 and the spatial distribution of parameter

values was defined according to land use. The model was applied to investigating the effects of land use change on flood flows. RRMT-SD is presently being applied to the large Illinois catchment, USA, as part of the international DMIP2 project, exploring calibration strategies for distributed models. Finally, RRMT-SD has been applied to the Upper Lee catchment, UK, and this application is described in some detail below.

3. THE UPPER LEE CASE STUDY – CATCHMENT DESCRIPTION

The RRMT-SD toolbox is demonstrated here in application to the Upper Lee catchment (1040 km^2) located in the Thames region, UK, see Fig. 3. The catchment is characterised as humid temperate with mean annual precipitation of 630 mm, while the elevation varies between 20 and 250 meters above UK ordnance datum (AOD). The study area corresponds to the whole Upper Lee catchment at Feildes Weir but also considers three inner subcatchments called Upper Lee at Luton Hoo, Mimram at Panshanger, and Stort at Glen Faba (see Fig. 3).

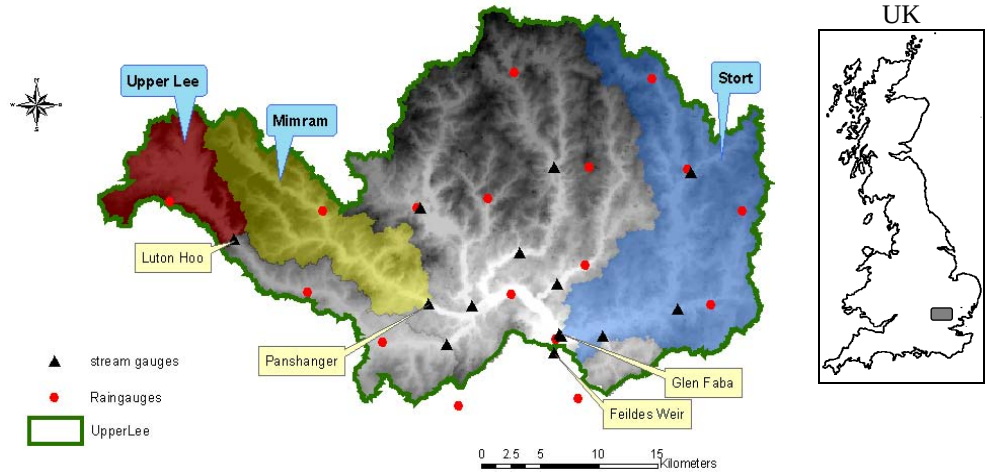


Figure 3. Map of the Upper Lee catchment, the subcatchments considered in this study are displayed in colour.

The Upper Lee subcatchment at Luton Hoo (70 km^2) is mainly chalk with high permeability. However, the subcatchment is highly urbanised with more than half covered by the city of Luton, while the land cover in the rest of the area is arable and horticulture. The subcatchment is dominated by high baseflow.

The Mimram subcatchment (130 km^2) is also mainly of chalk geology with high baseflow. Although the catchment is homogeneous and mainly rural, there is a small urban segment due to the city of Welwyn Garden. The upper part of the Stort subcatchment (278 km^2) is mainly rural, whereas the valley in the lower part is urban. Three cities of Bishop Stortford, Sawbridgeworth, and Harlow affect the runoff. The geology is chalk and clay. Overall, the Upper Lee at Feildes Weir (1040 km^2) is mainly rural, characterised by arable farming. The area has seen significant growth in housing, with urban areas covering 15 % of the total area. The Upper Lee is classified as medium baseflow dominated catchment.

A relatively fine subdivision of the subcatchments into subunits of about 10-25 km^2 was applied (Table 1), to investigate the significance of spatial rainfall on runoff generation (details can be found in Pechlivanidis et al. [2008]).

Historical hourly flow, rainfall and monthly mean potential evapotranspiration data were provided by the Environment Agency of England and Wales for the period 1991-2002. For each subcatchment, mean areal precipitation was estimated interpolating raingauge records (Fig. 3) based on the Inverse Distance Weighting method. The precipitation is assumed homogenously distributed over the area of each subcatchment. Only the stream gauges at the outlet of the study areas were considered.

Table 1. Catchment spatial discretisation into subunits

Sub Catchments	Upper Lee @ Luton Hoo	Mimram @ Panshanger	Stort @ Glen Faba	Upper Lee @ Feildes Weir
Area (km^2)	70	130	280	1040
No. Sub units	4	9	22	74
Range subunits area (km^2)	15 - 25	15 - 20	15 – 20	15 - 25

4. THE UPPER LEE CASE STUDY - MODEL DESCRIPTION AND METHODS

4.1 Semi-distributed model structure

Different RR models were built up within RRMT-SD and applied to each subunit of the Upper Lee catchment. Here only the models with better performance are included. The two SMA modules considered are (1) The Probability Distributed Moisture (PDM) model [Moore, 2007], and (2) The Catchment Wetness Index (CWI) model [Jakeman and Hornberger, 1993].

The PDM assumes that rainfall during each time step accumulates in the soil moisture store, where a specific function is used to describe the distribution of storage capacity over the catchment. The soil moisture storage is depleted by evaporation as a linear function of PE and the degree of saturation. The soil moisture storage capacity, in this study, is described by a 2-parameter Pareto distribution. In catchments with high base flow a bypass mechanism (BP) was introduced to represent groundwater runoff generation due to highly permeable soils, such as chalk. The CWI model uses an empirical model to simulate an index of wetness. The losses are proportional to the value of this index, while the gain in the index at each time-step is linearly related to the input rainfall. The effective rainfall is linearly related to the value of this index and rainfall.

The routing models applied were two linear reservoirs in parallel (2 par), 3 linear reservoirs in parallel (3par), and 2 linear reservoirs in parallel including macro-pores (2pmp).

A linear reservoir is used as channel routing for all the combinations of SMA and routing models.

4.2 Calibration strategies

Based on the Monte Carlo approach, a uniform random searching procedure was used to explore the feasible parameter space and to investigate parameter identifiability. Eight years of data were used for calibration (1991-1998) and four for validation (1999-2002). Two calibration strategies were considered. One strategy assumed that the rainfall-runoff process is homogeneous among the subcatchments thus the same model structure and parameter sets are applied on all subunits (URS-LP), and a single “optimal” parameter set was used.

The other calibration strategy estimates a priori parameters values at the subunit scale (SD-prior) by regionalisation and downscaling of calibrated parameters (see Pechlivanidis et al. [2008]). These a priori estimates were used to define the spatial variability of the parameters across subunits, and then the value of each parameter was adjusted uniformly across subunits using a multiplier which was calibrated to optimise the performance. This second calibration strategy was only applied using the PDM model because a priori parameter estimates have not yet been developed for the CWI model.

All catchments were calibrated using streamflow data at the catchment outlet using two objective functions (OFs) (Eq. 1): the Nash-Sutcliffe Efficiency (NSE) and a root mean square error based on low flows only (FSBM).

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Qobs_i - Qsim_i(\theta))^2}{\sum_{i=1}^{n}(Qobs_i - \overline{Qobs})^2} \qquad FSBM = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(Qobs - Qsim(\theta))^2}}{(1/n)\sum_{i=1}^{n} Qobs} \quad (1)$$

where *Qsim* is the calculated flow using the parameter set θ, *Qobs* is the observed flow, and *n* is the length of the time series in hours. The threshold considered for low flows corresponds to 25% of the mean observed flows.

5. THE UPPER LEE CASE STUDY - RESULTS

Table 3 summarises the model performance in terms of the NSE of the various models applied during calibration. Although the PDM-2par model is performing better than the other structures in the case of low-medium baseflow catchments (Stort at Glen Faba and Upper Lee at Feildes Weir), the structure does not seem to be able to describe the hydrological processes of the high baseflow areas (Upper Lee at Luton Hoo and Mimram at Pangshanger). The CWI-2par structure appears to be the most suitable for these areas. This result is difficult to explain by the representation of the hydrological processes and it is thought to be related to the ability of CWI model to compensate for mass balance errors. In all cases the "best" model structures, as measured by NSE, represent the runoff routing as two linear reservoirs in parallel.

Fig. 3(A) shows the variability of one of the PDM's calibrated parameter values and Fig. 3(B) the relationship between the model performance (based on NSE) and the calibration multipliers used in the PDM-2par model in the Upper Lee at Feildes Weir catchment. The results demonstrate that even though an adjustment factor is assumed homogeneous over the catchment, the application of a priori estimates which are consistent with observed catchment characteristics [Pechlivanidis et al., 2008] can represent the spatial variation of model parameters in a parsimonious manner. In the case of the Upper Lee catchment at Felides Weir, some refinement of the ranges of multipliers values seem to be needed (results for Kchannel and %q in Fig. 3B).

Fig. 4 shows observed streamflow in the Stort catchment at Glen Faba and the "best" simulated streamflow series, based on the NSE criteria, using calibration that considered the same parameter sets in each subunit (URS-LP), different parameter sets (URS-Dist), and based on prior parameter values (SD-prior).

Table 3. Model performance based on the NSE criterion, on grey the selected model structure for each subcatchment based on this criterion

Models	Upper Lee @ Luton Hoo 0.87*	Mimram @ Panshanger 0.72*	Stort @ Glen Faba 0.49*	Upper Lee @ Feildes Weir 0.56*
PDM-2par	-0.09	0.51	0.86	0.84
PDM-2par(BP)	0.1	0.52	0.80	0.81
PDM-3par	0.13	0.54	0.83	0.82
PDM-3par(BP)	0.16	0.57	0.78	0.81
CWI-2par	0.61	0.72	0.67	0.66
CWI-3par	0.56	0.71	0.7	0.67
CWI-2pmp	0.57	0.68	0.69	0.65

* represents BFIHOST number

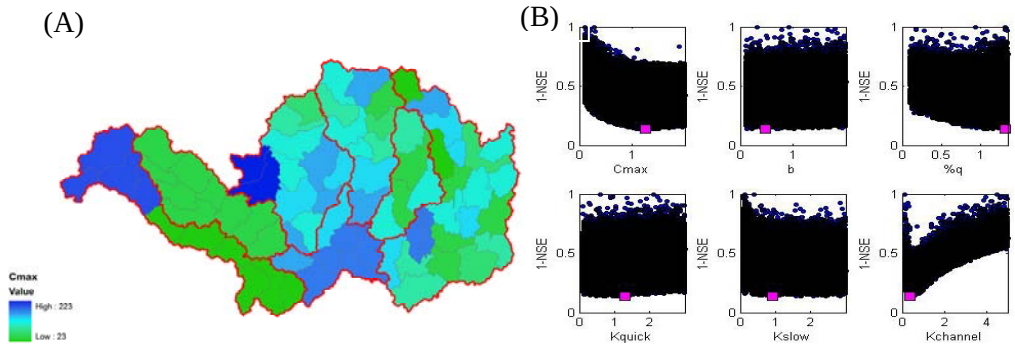


Figure 3. (A) Example of spatial parameter variability in the Upper Lee catchment; (B) Identification of calibration multipliers for PDM-2par in the Glen Faba subcatchment

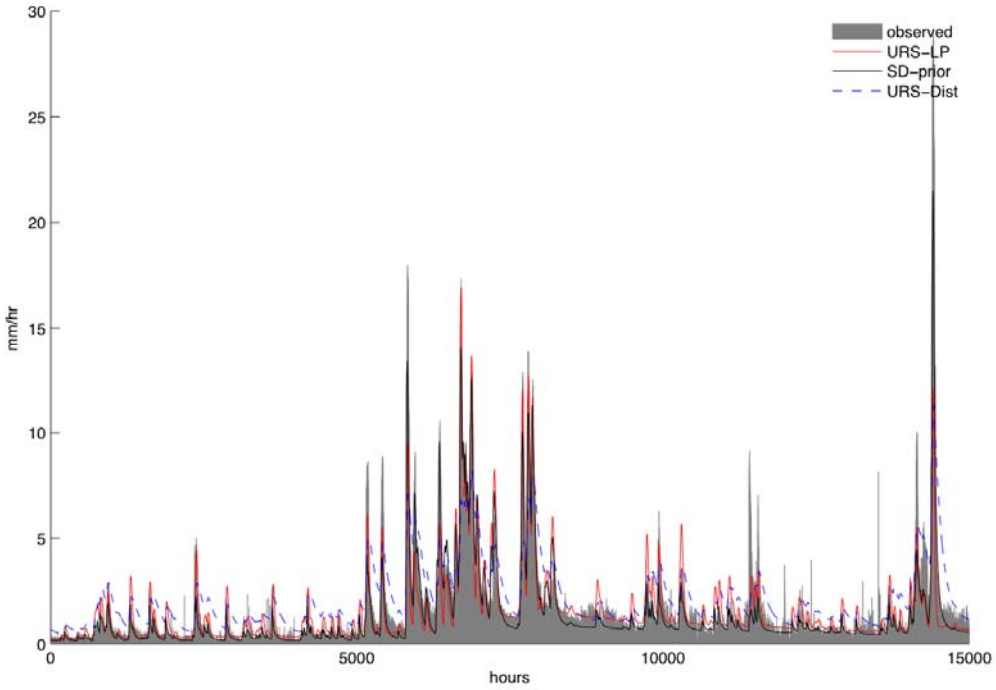


Figure 4. Simulated and observed streamflow in at Stort Glen Faba

6. CONCLUSIONS

The RRMT-SD toolkit presented here facilitates the development, analysis and application of semi-distributed parsimonious rainfall-runoff model structures. RRMT-SD has a high degree of model flexibility, which allows the implementation of established or new structures with low or medium level of complexity quickly and efficiently.

The case study of hourly data from the Upper Lee catchment showed some of the applicability of RRMT-SD to investigate calibration strategies using different criteria and inter-comparisons of conceptual model structures. Also, the case study has shown the potential of the toolbox for developing regional equations for a priori estimation of model parameters and subsequent optimisation using multipliers, hence maintaining spatial variations which are consistent with catchment characteristics while maintaining identifiability. Various other applications of RRMT-SD at Imperial College London have demonstrated the toolbox potential applicability to a range of hydrological problems, including flash flood simulation in arid regions and simulating effects of land use change. Further applications of the toolbox are in development.

REFERENCES

Ajami, N.K., H. Gupta, T. Wagener, and S. Sorooshian, Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, *Journal of Hydrology*, vol. 298, no. 1-4, pp. 112-135, 2004.

Beven, K., *Rainfall-Runoff modelling. The Primer* John Wiley and Sons, Chichester, UK, 2001.

Boyle, D.P., H.V. Gupta, S. Sorooshian, V. Koren, Z.Y. Zhang, and M. Smith, Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resources Research*, vol. 37, no. 11, pp. 2749-2759, 2001.

Carpenter, T.M., and K.P. Georgakakos, Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, *Journal of Hydrology*, vol. 329, no. 1-2, pp. 174-185, 2006.

Cormen T.H., C.E. Leiserson, R.L. Rivest, and C. Stein, *Introduction to Algorithms*, Second Edition. MIT Press and McGraw-Hill, Section 29.3: The simplex algorithm, pp.790–804, 2001.

Duan, Q., V.K. Gupta, and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimization, *Journal of Optimization Theory and Applications*, vol. 76, no. 3, pp. 501-521, 1993.

Gupta, H.V., S. Sorooshian, and P.O. Yapo, Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resources Research*, vol. 34, no. 4, pp. 751-763, 1998.

Koren, V., M. Smith, and Q. Duan, Use of a priori parameter estimates in the derivation of spatially consistent parameters sets of rainfall runoff models, in *Calibration of Watershed Models*, vol. 6 American Geophysical Union, pp. 239-254, 2003.

Lee, H., N. McIntyre, H.S. Wheater, and A.R. Young, Predicting runoff in ungauged UK catchments, *Water Management*, vol. 159, no. WM2, pp. 129-138, 2006.

Madsen, H., G. Wilson, and H.C. Ammentorp, Comparison of different automated strategies for calibration of rainfall-runoff models, *Journal of Hydrology*, vol. 261, no. 1-4, pp. 48-59, 2002.

Pechlivanidis, I.G., N. McIntyre, and H.S. Wheater, Relating conceptual rainfall-runoff model structure and parameters to catchment characteristics: A case study on the Upper Lee catchment, in *HYDRA Postgraduate Research Poster Meeting*, 17 December, University College London, London, UK, 2007.

Pechlivanidis, I.G., N. McIntyre, and H.S. Wheater, The significance of spatial variability of rainfall on runoff, in *International Congress on Environmental Modelling and Software* (iEMSs 2008), 7 July, Barcelona, Spain, 2008.

Perrin, C., C. Michel, and V. Andreassian, Does a large number of parameter enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, in *Journal of Hydrology*, vol. 242, pp. 275-301, 2001.

Wagener, T., D.P. Boyle, M.J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian, A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, vol. 5, no. 1, pp. 13-26, 2001.

Wagener, T., M.J. Lees, and H.S. Wheater, A toolkit for the development and application of parsimonious hydrological models, in *Mathematical models of large watershed hydrology - Volume 1*, 1 edn, V. P. Singh & D. Frevert, eds., USA, pp. 87-136, 2002.

Wagener, T., and N. Mcintyre, Identification of rainfall-runoff models for operational applications, in *Hydrological Sciences*, 50(5), pp. 735-751, 2005

Wheater, H.S., A.J. Jakeman, and K.J. Beven, Progress and directions in rainfall-runoff modelling, in *Modelling change in environmental systems*, A. J. Jakeman, M. B. Beck, and M. J. McAleer, eds., New York, pp. 101-132, 1993.

Wheater, H.S., Progress in and prospects for fluvial flood modelling, in *Philosophical Transactions of the Royal Society of London Series A - Mathematical Physical and Engineering Sciences* 360 (1796), 1409-1431, 2002.

Wooldridge, S., J. Kalma, and G. Kuczera, Parameterisation of a simple semi-distributed model for assessing the impact of land-use on hydrologic response, in *Journal of Hydrology*, vol. 524, pp. 16-32., 2001

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Development of matrices to build scenarios of cropping systems distribution for integrated catchment assessment. A proposition for an irrigated area in south-western France

Lucie Clavel[a] *and Delphine Leenhardt*[b]

[a] *INRA, UMR1248 - AGIR (Agrosystèmes et développement territorial) B.P. 52627, F-31326 Castanet Tolosan Cedex, France (lucie.clavel@toulouse.inra.fr)*

[b] *INRA, UMR1248 - AGIR (Agrosystèmes et développement territorial) B.P. 52627, F-31326 Castanet Tolosan Cedex, France (delphine.burger-leenhardt@toulouse.inra.fr*

Abstract: Scenarios are increasingly accepted as useful tools for assessing environmental and economic impacts of land use changes, for decision making and for involving different stakeholders around land use issues. In rural areas, where agriculture is an important user of land and water, land use scenarios are particularly relevant and should consider specifically cropping systems distribution. In such areas, cropping system distribution has direct impact on water availability in term of quantity and quality. This paper details the method that was used to organise spatially distributed data in order to describe a cropping system distribution intended by a stakeholder and to evaluate it.

For developing this method, we chose a south-western France catchment, where alternative cropping systems distributions were suggested for reducing water scarcity. We defined evaluation indicators relevant for this case study (water demand for irrigation and economic yields), and identified a model (MO_uSTICS) that could simulate these indicators for different cropping systems. The strategy was to distribute the model over the considered catchment in order to simulate its outputs and aggregate them at different scales. The work consisted in (i) defining the support units where evaluation indicators could be returned and (2) organising data within these support units for running the model.

We defined supports units relevant for water management and economical issues. A matrix approach was selected to specify all possible combinations of model input data in all support units. It separates the different factors that could be mobilised by stakeholders for describing cropping systems location within the catchment. The association of support units and matrices appeared to be a flexible solution for facilitating construction of scenarios of cropping systems distribution and for returning model outputs at different scales for scenario evaluation.

In the next future, we plan to test this method with a set of stakeholders in order to identify unexpected factors, which could lead to modify the matrices and the support units.

Keywords: Cropping systems; Scenario; Spatial variability; Water resources; Land planning

1. INTRODUCTION

Integrated catchment management (ICM) promotes the catchment as territorial unit for optimising the allocation of land and water resources to all users [Gorredale, 1992]. ICM is often supported by the use of models that represent complex and interacting environmental, social and economical processes occurring at the catchment scale. These models permit the estimation of various, often competing and conflicting demands in land and water and make them match the offer [Jakeman and Letcher, 2003].

In rural areas, agriculture is often the main user of land and water resources, and agricultural water use is directly related to agricultural land occupation (e.g. cropping systems, field pattern) and catchment physical characteristics (e.g. soil, slope, climate, etc.). In such areas, the distribution of cropping systems has a direct impact on water availability in term of quantity and quality. Thus, assessing the impact of cropping systems distributions within a catchment area is (or should be) a key part of integrated assessment modelling (IAM) for ICM.

As specified by Jakeman and Letcher [2003], IAM can provide good guidance for decision making, particularly through scenario simulation. The use of scenario facilitates interaction and understanding between experts, stakeholders and policymakers [Alcamo, 2001].

In rural catchments, stakeholders may thus be interested in evaluating alternative distributions of CSs, which could be influenced by controllable driving forces such as regulation or incentives. We then assume that a participative construction of scenarios of cropping systems distributions can be an appropriate way for helping catchment management and planning.

This paper details the methodology that was used in the Neste System, a south-western France catchment, to organise spatially distributed data in order to describe and evaluate alternative cropping systems distributions. It first focuses on the methodological issues to be considered. After a description of the particular context of the Neste System, it describes the way data was organised in order to provide adequate scenario evaluation.

2. PRELIMINARY DEFINITIONS

According to Sebillotte [1990], a cropping system (CS) is the set of technical modalities implemented on fields with the same treatment, i.e. the same succession of crops (rotation) and the same management system associated to each crop. An agricultural land can then be viewed as a set of land areas affected to CSs. By "cropping systems distribution" (CSD), we refer, in this paper, to a more or less detailed description of the location of CSs within the concerned area. Benoît and Papy [1998] insist on the fact that, at such scales, crop successions and crop management systems do not result from a random process but are the result of logic and decision processes. In a given catchment, CSD can then be explained by the soil-climatic properties of fields, but also by factors at greater levels, e.g. at farm or landscape levels [Jouve, 2007; Sebillotte, 1990]. For example it is frequent to observe intensive CSs close to the village and extensive CSs far away [Jouve, 2007]. These factors can be considered as invariant, at least in a short term horizon. Other factors depending more on politics and economics (e.g. prices, subsidies) can influence the evolution of CSD. They are usually called driving forces [Rounsevell et al., 2006].

We define "scenarios of cropping systems distribution" (SCSD) as alternative CSDs, resulting from some change in the underlying stakeholders' logic and decision processes. SCSDs can be formulated by stakeholders for reaching a given objective in an ICM context. Our study does not deal with the forces that drive the evolution of CSs and are controllable by some stakeholders. It focuses on invariant factors that may constraint CSD and its impact on the environment (i.e. economic, ecological or social).

Evaluating SCSDs consists in assessing the values of "indicators", i.e. meaningful variables for stakeholders with regard to the objectives they aim to reach. For example, if the objective is to reduce the pressure on water resources, the water demand for irrigation at the catchment scale is a good indicator (but not the only one) because it is a meaningful variable for the water manager. When the multi-criteria evaluation of a SCSD is assumed as relevant by stakeholders, they can attempt to reach this CSD by implementing a policy that modifies adequate driving forces. It is then crucial to provide adequate indicators for enabling scenario evaluation by each concerned stakeholder.

3. METHODOLOGICAL ISSUES

The purpose of this work is to provide stakeholders with a method that helps them to build and evaluate scenarios of cropping systems distribution. This paper details the method used for organising data in order to permit (i) any stakeholder to participate to the process of scenario construction and (ii) to evaluate a scenario by using a spatialised simulation model. We detail here three key issues that arised from these objectives.

3.1. Stakeholders, scales and criteria

In ICM, stakeholders are possibly numerous (land and water planners, economical actors, farmers, agricultural extension services, etc.). Their scales and criteria of analysis may differ from one to another. An objective of ICM and IAM is to be responsive of, and integrate, different (groups of) stakeholders [Parker et al, 2002]. The challenge is therefore to be able to account for a variety of scales and criteria in the whole process of construction and evaluation of scenarios, and, as far as possible, not to limit them. Indeed, it is recognised that such multi-criteria approaches, which may integrate diverse social, economic and ecological values, favour a collaborative planning and decision making environment [Mendoza, 2006; Sheppard and Meitner, 2005].

Consequently, the method to be developed should account for a large range of (i) modalities of CSs that may be modified in a scenario, (ii) factors that may orient the location of these modifications (i.e. invariant factors) and (iii) levels (i.e. management units) where these modifications can be relevant. For scenario evaluation, stakeholders should be able to consider different indicators at different levels or scales (e.g. economic impact of a SCSD at catchment level or its environmental evaluation at particular sensitive sites).

3.2. Spatialisation of models

Whereas individual oriented models (e.g. multi-agent systems, cellular automata) allow simulating the temporal evolution of a situation to another according to driving forces (e.g. economic, social or environmental pressures), a common way to evaluate impacts of a given and static agricultural situation at a catchment level is to distribute a model (e.g. crop model, soil water balance model) over the considered area. The so-called "model spatialisation", is a way to represent the spatial variability of the phenomenon simulated by the model for this extent [Klöcking and Haberlandt, 2002; Faivre et al., 2004]. Once the model chosen (i), spatialising the model implies (ii) to define elementary spatial Support Unit (SU) where data will be collected, and (iii) to collect model input data that characterise these SU.

The selection of the model should be done with regard to the agri-environmental problem considered. The selected model should also provide necessary outputs to calculate indicators of interest for the intended stakeholders.

The definition of SUs should take account of available data, of spatial variability of model inputs (and therefore of model outputs), and of stakeholders perceptions about scales and criteria for scenario construction.

The collection of input data on all SUs raises the questions of the source of information to use and of the way the information should be distributed to the SUs. Two kinds of data are to be collected: (i) model input data and

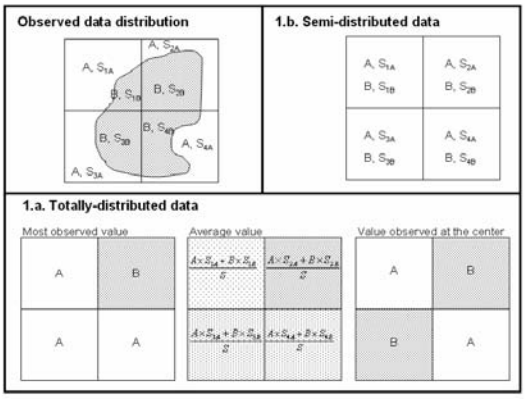


Figure 1 : Example of data distribution possibilities within four SUs (represented by the four squares). A and B are modalities of the variable to be distributed; S_{iA} is the area of modality A in SU_i

(ii) invariant factors of CSDs or CSs' technical modalities that can be used by stakeholders for building SCSD. These two groups are not necessarily completely distinct: some model input variables may be considered by stakeholders as factors of CS modalities.

Regarding one particular variable, the SU may be characterised by one value only (the variable is then totally distributed) (Figure 1.a) or characterised by two or more values, each value being associated with the proportion of the SU concerned (the variable is, in this case, considered as semi-distributed) (Figure1.b).

When more than one variable is semi-distributed, different combinations of modalities observed for each variable may coexist in each SU. The model can then be run for each combination of modalities of input variables. Each combination is called a "simulation unit". The result for the whole SU is the weighted sum (or average) of the simulation units outputs, the weight being the area of each simulation unit.

3.3. Production of scenario results

The model produces outputs at simulation unit level. To be meaningful for stakeholders, these outputs should undergo an adequate restitution process and should be compared to a reference situation. Defining adequate communication tools is a key ingredient of IAM and ICM [Parker et al, 2002], the scenario results restitution process is therefore crucial. It should consider the nature of the output(s) to be provided to stakeholders (which indicator?), how it should be estimated (average, median, sum etc.), and under which visual form (e.g. maps, curves, graphs, tables). To facilitate aggregation or averaging according to different levels (i.e. water management, administrative district, farm) or different criteria (i.e. crop, soil), it is important to store outputs at the lowest level, i.e. simulation units. A mean for facilitating stakeholders' understanding about scenario results is to compare the latest to a reference situation. The reference situation often corresponds to model outputs for a baseline scenario, which can be a situation observed in the past [Palang et al., 2000], the most expected situation in a given future [Rounsevell et al., 2006] or a situation where actual trends in land use changes were translated in a given future [Ewert et al., 2005].

4. STUDY CASE

4.1. Context

The study area is the Neste System (Figure 2). This 800 000 ha area gathers the catchment areas of 14 tributaries of the river Garonne artificially supplied in water by the Neste canal. The land is mainly dedicated to agriculture: 500 000 ha are cultivated from which 50 000 ha are irrigated [Hurand, 2000]. The main irrigated crop is maize. Thanks to the canal and European common agricultural policy, irrigated area has been multiplied by around three since the 70's. As a consequence, two years out of ten, the system faces a lack of water that threats the environmental equilibrium of the rivers and the satisfaction of agricultural needs in water.

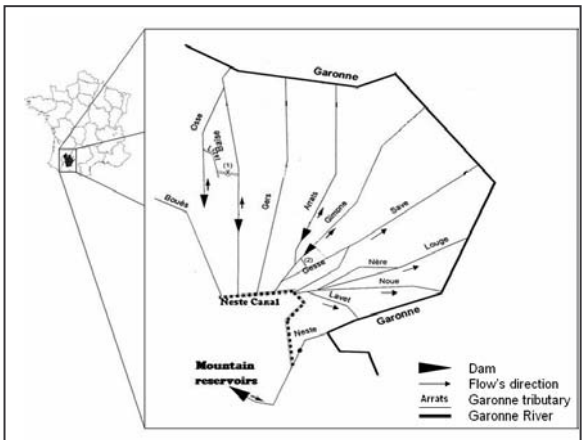


Figure 2 : The Neste System

The project of building a dam (the Charlas dam) to support ecological flows of Garonne and its tributaries led to a public debate that animated the whole south-western France about agriculture issues [Mandement, 2004]. Proponents of the project argued that the dam could not only sustain ecological flows, but also increase water offer for other water uses such as irrigation within the Neste system. Contestants argued for a reduction of 15% of the irrigated area in order to contain water deficits.

4.2. Indicators and model

Regarding this case context, we assumed that the assessment of an effective reduction (or increase) of the irrigated area relied on the estimation of the effective farmers water demand and resulting economic yield to balance environmental and economical arguments. We chose the biodecisional model MO_uSTICS for simulating these indicators. This model integrates (i) a crop model (STICS, [Brisson et al., 2002]) that simulates variables relevant for SCSDs' evaluation (production variables such as crop yield, and externalities such as water and nitrate loss) for different crop present within the Neste system and (ii) a decisional model that, instead of simulating the water demand of the considered crop, explicitly simulates irrigation applications through farmers' decision rules that depend on crop, soil, climate and water resource constraints [Bergez et al., 2001]. CSs, soil and weather information is required as input data for running this model.

For assessing an effective reduction (or increase) of the irrigated area, the strategy is to spatialise the MO_uSTICS model over the whole Neste system, permitting thus to apreciate the spatial heterogeneity of CSD. It is therefore necessary to identify where (e.g. soil type, farm type etc.) the stakeholders (e.g. proponents and opponents) intend to modify CSs, and which CS (i.e. crop and cultural practices) they intend to carry out.

5. PROPOSED METHOD FOR ORGANISING DATA

5.1. Simulation units (SU)

The SUs were defined as the intersection between water management units defined by CACG and Small Agricultural Regions (SAR) (Figure 3). CACG management units represent longitudinal sections of rivers' catchment areas. They are used by the operational water manager to optimise water releases between different rivers according to local demands (mainly irrigation and ecological demands). The SARs, also called "Agricultural Districts", have been defined in France in 1946 by the French National Institute of Statistics and Economic Studies in order to coordinate post-war agricultural development. They were delineated according to agriculture-related criteria (mainly soil and climate conditions). Although agricultural intensification turned some of them obsolete, after more than fifty years of existence the SAR have mostly retained their own characteristics and well represent uniform farming zones [Mignolet et al., 2007].

Thus, SUs have both agricultural and water management relevancy and are significative for water managers, land planners or farmers. As SARs are aggregation of local government areas (also called municipalities or "communes" in France), where crop and farming system data from the agricultural census is accessible, this choice also facilitates agricultural data collection and distribution.

The intersection was done by using Geographical Information System (GIS) software. Each micro-polygon (less then 1 % of the total area) has been inserted to its neighbour belonging to the same water management unit. It resulted in 76 SUs, with a mean area of 10 000 ha.

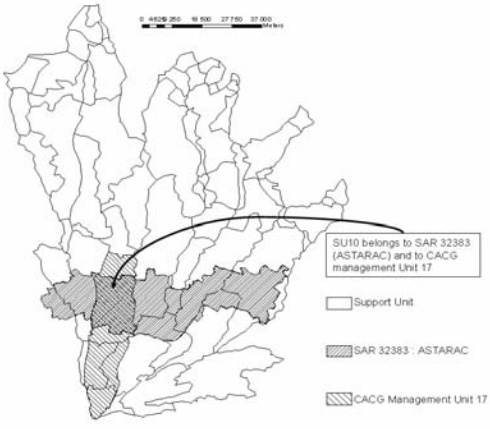


Figure 3 : The SUs

5.2. Data distribution

GIS was also used to distribute data within SUs by overlaying SU layer to other data layers, from which, soil and weather MO_uSTICS input data and other variables such as factors of CSD or CSs' technical modalities.

To handle soil variables, we defined four main soil types according to their agronomic potential and semi-distributed them within each SU. The reason for this was that SUs, by construction (longitudinal sections of river catchments), often included all soil types (from deep valley soils to shallow hill soils).

For distributing weather data, we affected to each SU a point where daily weather data is known for different years. The whole Neste system includes 121 grid-points [Prats and Pérarnaud, 2001]. Each SU was associated to the grid point closest to its centroid.

For distributing CSs, we used a farm typology because we assumed that the farm type could be a factor that would be used by some stakeholders to allocate CSs within the catchment. For each SAR, ten farm types were defined according to economic dimension, specialisation and irrigated area. They were characterised by average Utilised Agricultural Area (UAA) and percentage occupied by each CS (a CS being a crop related to a set of MOuSTICS decision rules). The number of farms of each type was then calculated for each SU in order to reconstitute the observed UAA (in 2000) of the considered type within the given SU.

Soil and farm types have been independently semi-distributed within SUs. No information was available on the number of farms of each type per soil type or on the proportion of each soil type per farm type. To reconstitute homogeneous soil, climate and CS conditions for running MO_uSTICS, we assumed, in a first stage, that for any farm all soil conditions were equally probable: if the UAA of an SU contains X% of farm type f and Y% of soil type s, the proportion of farm type f within soil s will be X * Y%.

5.3. Matrix architecture

CS, weather and soil information is necessary for running MO_uSTICS. Since the CS information is related to the farm type, the purpose was then, for any SU, to determine the distribution of crops within each (CS x soil type x farm type) combination.

720 combinations are possible for each SU (4 soil types * 10 farm types * 18 CSs). A 3-dimensions matrix was elaborated to calculate the area A_{cfs} of each combination (Figure 4). We will further refer to it as the "situation matrix".

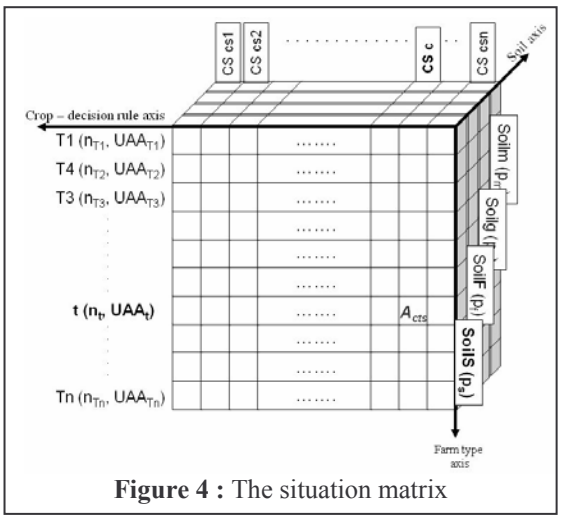


Figure 4 : The situation matrix

For each SCSD and each SU, the situation matrix represents: (i) the number n_f of each farm type f and its associated UAA, UAA_f, on the farm type axis; (ii) the relative proportion a_{ct} of each CS (c) within the farm type f, on its CS axis; and (iii) the proportion p_s of each soil S on its soil axis. "Reference" situation matrices were built for a CSD observed in the past (in 2000). The "scenario" situation matrices can be built from modifications of the "reference" situation matrices. For example, if some economical subsidies may lead to the development of one particular farm type F1 to the detriment of farm type F2, a simple change in the situation matrix can translate this idea: the proportions in F1 and F2 in the reference situation matrix are just modified and the areas of the combinations are re-calculated, leading to a new "situation" matrix for this particular scenario. Since a "situation" matrix is defined for each SU, it is possible to modify situation matrices in limited number of SUs. This concept of matrix is then a powerful tool to design scenarios. For example, if a scenario deals with regulations affecting particular water-deficit river catchments, only the matrices of the concerned SUs have to be changed.

5.4. Scenario results

At this stage we selected two outputs of particular interest, simulated by MO_uSTICS: daily water demand and crops yield. The daily water demand measures the impact of the SCSD on the water resource. Combined with the area concerned and with crop prices and input costs, the yield allows getting an economical evaluation of the studied scenarios. As these outputs are also calculated for the reference SCSD, a comparison of outputs is possible for any SCSD. This tool can then provide an indication about the evolution of these indicators for various scenarios in comparison to a reference situation known by all stakeholders.

The same matrix approach is used for scenario evaluation restitution. "Output" matrices with the 720 possible combinations were elaborated. These matrices store the model outputs estimated for each possible combination. These outputs are expressed per hectare. Multiplying them by corresponding areas contained in situation matrices allows calculating the expected results (i.e. daily water demand and crop production) for each situation. These results can then be stored in result matrices, following the same architecture: one matrix per SU, each matrix having the same 3 dimensions representing crop, soil and farm type. Therefore it is possible, for evaluating the scenario, to express outputs according to various indicators (using matrix axes) or to different perception levels (by aggregating SU to a superior level, e.g. CACG management unit or SAR).

6. DISCUSSION AND CONCLUSION

In this paper, we presented a method for organising spatially distributed data in order to facilitate the construction and evaluation of SCSD. Its originality lies on a stratification of the extent in SUs and on the use of a multi-criteria matrix approach. This allows accounting for different spatial levels and different criteria of perception, permitting thus a maximum of stakeholders to participate to the process of scenario construction and evaluation.

However, two points can be discussed. The way farms were distributed in relation to soil condition is exclusive. One farm is located on one unique soil type. This strong assumption should be revisited later: some farm types might be preferentially located on particular soil types. For example, in south-western France, it is more frequent to meet cattle uphill and large-scale farming in deep valleys. But conversely many farm areas cover the whole topographical sequence of soils [Maton et al, 2008], which certainly influences the CSD within the farm. To account for the actual distribution of soil types within farms, it would be necessary to study the correlation between soils and farm types and between soils and characteristics of different farm types.

One of the methodological issues was to create a sufficient number of factors and scales to allow any stakeholder constructing and evaluating SCSD. At this stage, we selected a finite set of factors and scales. It is now important to implement the proposed method with stakeholders because they may identify some additional factors or scales, and may lead to the modification of SUs and matrices. This step will also raise new issues related to stakeholders' participation that have not been discussed in this paper: the choice of stakeholders (e.g. institutions, managers, farmers), the way to involve them (e.g. collective or individual participation) and the mediating representations to be developed for facilitating scenario construction (e.g. game, computer interface etc.).

The stratification in multi-criteria SU was justified by the diversity of stakeholders that may be involved in the construction and evaluation of SCSD. Indeed, stakeholders may mobilise different variables to build the scenarios. These variables are not all known a priori. This is why we specified the SUs as the intersection of layouts of very general signification, but not as the intersection of all model input and scenario criteria variables, which would had led to very small and numerous SUs and to a stratification of the catchment extent that may have changed from one user to the other.

The use of a matrix system is the consequence of this last choice. It is allowed by the fact that estimating irrigation demand does not require to account for neighbouring effects between fields. However, this matrix system could be used for constructing and evaluating SCSDs for surveying other environmental problems, such as water pollution or gene

transfer with the condition to define SUs that permit to run transfer models. In addition such matrix approach allows accounting for scale issues requirements by decision-makers.

REFERENCES

Agreste, Recensement général agricole 2000. Cartes thématiques Midi-Pyrénées. *CR-Rom édité par le Service Régional de la Statistique Agricole de Midi-Pyrénées*, 2002.

Alcamo, J., Scenarios as tools for international environmental assessments. 24, 1-31. European Environment Agency. Environmental issue report, 2001.

Benoît,M. and F. Papy, La place de l'agronomie dans la problématique environnementale. *Dossier de l'environnement de l'INRA*.53-62, 1998.

Bergez, J.E., P. Debaeke, J.M. Deumier, B. Lacroix, D. Leenhardt, P.Leroy, and D. Wallach, MODERATO: an object-oriented decision tool for designing maize irrigation schedules, *Ecological Modelling*. 137:43-60, 2001.

Brisson, N., F. Ruget, P. Gate, J. Jorgeou, and B.E.A. Nicoullaud, STICS : a generic model for simulating crops and their water and nitrogen balances. Model validation for wheat and maize, *Agronomie*. 22:69-92, 2002.

Ewert,F., M.D.A.Rounsevell, I.Reginster, M.J.Metzger, and R.Leemans, Future scenarios of European agricultural land use I. Estimating changes in crop productivity. *Agriculture Ecosystems & Environment*. 107:101-116, 2005.

Faivre, R., D. Leenhardt, M. Voltz, M. Benoît, F. Papy, G. Dedieu, and D. Wallach, Spatialising crop models. *Agronomie*. 24:205-217, 2004

Gorredale,E., The Dublin Statement on Water and Sustainable Development. *Environmental Conservation*. 19:181, 1992.

Hurand, P., La gestion opérationnelle d'un système complèxe: "Le système Neste". CACG. 1-23, 2000.

Jakeman, A.J. and R.A. Letcher, Integrated assessment and modelling: features, principles and examples for catchment management, *Environmental Modelling & Software*. 18:491-501, 2003.

Jouve, P, La dimension spatiale des systèmes de culture : comparaison entre agriculture tempérée et agriculture tropicale, *Cahiers Agricultures*. 15:255-260, 2007.

Klöcking, B. and U.Haberlandt, Impact of land use changes on water dynamics-a case study in temperate meso and macroscale river basins, *Physics and Chemistry of the Earth*. 27:619-629, 2002.

Mandement, A., Compte-rendu et bilan du débat public de Charlas, *Revue de l'agence de l'Eau Adour Garonne*.26-30, 2004.

Maton, L., D. Leenhardt and J.E. Bergez, Choix de précocité et pratiques de semis en maïsiculture irriguée du sud-ouest de la France : quelle diversité et comment l'expliquer? Submitted to *Cahiers Agriculture*, 2008.

Mendoza,G.A. and H.Martins, Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*. 230:1-22, 2006.

Mignolet, C., C. Schott, and M. Benoît, Spatial dynamics of farming practices in the Seine basin: Methods for agronomic approaches on a regional scale, *Science of The Total Environment*. 375:13-32, 2007.

Palang, H., H. Alumae, and U. Mander, Holistic aspects in landscape development: a scenario approach, *Landscape and Urban Planning*. 50:85-94, 2000.

Parker,P., R.Letcher, A.Jakeman, et al, "Progress in integrated assessment and modelling." *Environmental Modelling & Software*. 17:209-217, 2002.

Prats, S., and V. Pérarnaud, Projet ADEAUMIS. Contribution de Météo-France, September 2001, 19 p, 2001.

Rounsevell, M.D.A., I. Reginster, M.B. Araujo, T.R. Carter, N. Dendoncker, F. Ewert, J.I. House, S. Kankaanpaa, R. Leemans, and M.J. Metzger, A coherent set of future land use change scenarios for Europe, *Agriculture, Ecosystems & Environment*. 114:57-68, 2006.

Sebillotte, M., Système de culture, un concept opératoire pour les agronomes, In L.Combes and D.Picard, editors, *Les systèmes de culture*. INRA, 1990.

Sheppard,S.R.J. and M.Meitner, Using multi-criteria analysis and visualisation for sustainable forest management planning with stakeholder groups. *Forest Ecology and Management*. 207:171-187, 2005.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Deposition and Resuspension of Sediments in Near Bank Water Zones of the River Elbe

S. Prohaska[a] **and B. Westrich**[b]

[a] *Institute of Hydraulic Engineering, University of Stuttgart, Pfaffenwaldring 61, 70550 Stuttgart, Germany (sandra.prohaska@iws.uni-stuttgart.de)*

[b] *Institute of Hydraulic Engineering, University of Stuttgart, Pfaffenwaldring 61, 70550 Stuttgart, Germany (bernhard.westrich@iws.uni-stuttgart.de)*

Abstract: Catchment processes, such as spatially and temporally distributed rainfall, influence quantity and quality of river runoffs and thus, the transport of particulate sediments and associated contaminants in a river network. During flood events resuspension of contaminated sediments impacts the water quality and thus, it is important to determine the erosion probability of such deposits. The research is focused on the River Elbe deposits in groyne fields and their sediment dynamics. High percentage of contaminated deposits in the groyne fields originate from the tributary Mulde, which discharges high concentrations of arsenic and lead into the Elbe [Kraft, 2007]. Various hydrological scenarios were investigated for a 112 km long river reach of the Middle Elbe to predict the spatial distribution of the particulate suspended sediments and associated contaminants. The scenario results gave the estimation of erosion probability and pointed out the importance of the Mulde as the main contributor to the groyne field deposition and their potential source of contamination.

Keywords: river Elbe, transport modeling, suspended sediments, groyne fields.

1. INTRODUCTION

Describing the movement of suspended particulate matter in a river is important for ecological and water quality issues due to toxicity of adsorbed sediment bound contaminants. Numerical modeling of the suspended sediment transport contributes to better understanding of spatial and temporal distribution of ecologically interesting parameters, such as concentration in the water column, particulate deposition/erosion, deposited/eroded mass, etc. Fluvial hydraulics is the control mechanism of sediment transport. Therefore, modeling of different hydrological scenarios is a prerequisite for assessing ecological effects of toxicants to biodiversity. Various hydrological scenarios were applied to the River Elbe to predict the behavior of particulate suspended sediments and associated contaminants. The goal is to estimate the effect of the Elbe discharge on sedimentation and erosion in near bank groyne fields, and influence of a flood on dispersion and transport distance; and to determine the effect of contaminant inflow from tributaries and their spatial deposition in the Elbe. The scenarios results were used to estimate erosion probability as an important tool for decision making. The idea is to present a simple method to characterize the probability of erosion for contaminated sediment deposits in the groyne fields in terms of discharge probability.

2. HYDROLOGICAL SCENARIOS

The current research is focused on the Middle Elbe as a representative reach, for which the flow field is influenced by groyne structures on both river banks, see left panel of Fig. 1. The simulated domain extends from Wittenberg (km 214.1) to Magdeburg (km 326.6), comprising the influence of the tributaries Mulde (km 259.6) and Saale (km 290.7). The model used for numerical simulations is a 1D multi-strip model developed to predict the transport of suspended sediments in rivers trained by groynes [Prohaska and Westrich, 2006]. The model allows the subdivision of cross sections into three compartments in order to comprise the influence of groyne fields on flow and transport, refer to the right panel of Fig. 1.

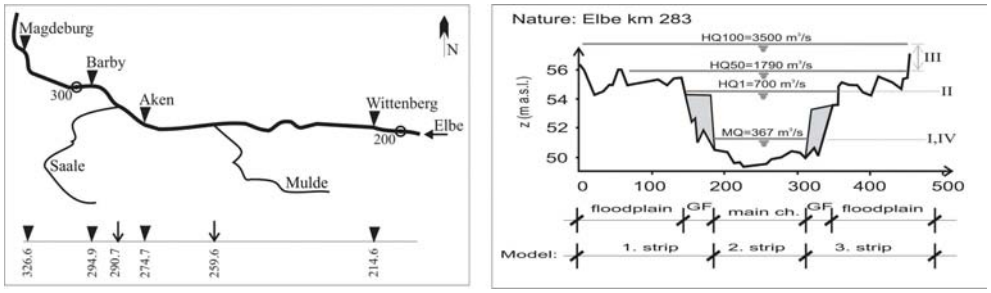


Figure 1. Left panel: Middle Elbe and model area from Wittenberg to Magdeburg; right panel: Typical cross section of the River Elbe and simulated scenarios (I-IV) with model simplification and subdivision into strips.

The model parameters are listed in Tab. 1. The mass exchange parameter (ε), which accounts for the exchange of suspended sediments between adjacent strips, was measured in the laboratory, as well as the averaged sedimentation parameter (ξ). The geometry conditions and model parameters were kept constant during all simulations. In all simulations the discharge of the Saale had a constant value of 115 m^3/s, and no suspended load. Therefore, the influence of the Saale was only by dilution, whereas the effect of a tributary on the transport processes was analyzed by the Mulde. Inflowing suspended sediment concentration and discharge in both the Elbe and the Mulde varied depending on the applied hydrological scenario. All input hydrographs are based on measured data. The model parameters are listed in Tab. 2.

Table 1. Model parameters for flow and transport used in the numerical simulations.

parameter	value	unit	comment
n*	0.028	$m^{-1/3}$ s	calibrated
n**	0.200	$m^{-1/3}$ s	calibrated
d_{50}	50	µm	measured
ξ	0.3	-	measured
ε	0.018	-	measured
$\tau_{c,E}$*	14.88	Pa	measured
$\tau_{c,E}$**	3.05	Pa	measured

*main channel; **side strip

Table 2. Hydrological scenarios with input data: discharge Q [m^3/s], suspended sediment concentration [mg/l].

River:	Elbe		Mulde		Saale	
Scenario	Q	c	Q	c	Q	c
I	367	58	64	30	115	0
II	700	58	64	30	115	0
III	1790-3500	58-150	64	23	115	0
IV	367	0	61-150	30-60	115	0

Estimation of the main river effect on the contaminant sediment dynamics was performed focusing on different scenarios: (I) discharge that is smaller than erosive, (II) critical discharge at which erosion starts, (III) an extreme flood discharge, and (IV) the influence of a tributary on sediment deposition, see Tab. 2.

3. MODELING RESULTS

3.1 Constant discharge periods

In order to explore the long lasting deposition of contaminated sediments in groyne fields and whether erosion occurs in groyne fields for a flood of less than one year return period (Q_1=876 m^3/s), numerical simulations of a mean discharge (scenario I) and small flood discharge with water level above groyne crests, but still lower than the bank elevation (scenario II) were performed. In both scenarios a constant inflow of suspended sediments in the Elbe was applied.

During 30 days (scenario I) around 367 m^3/day of sediments were deposited in the groyne fields. In both the side strips there is a constant total sedimentation rate of about 172 t/day. The domain which is influenced by the Elbe sediment inflow is longer than the computational domain, i.e., small deposition in groyne fields still occurs 100 km downstream. In the upstream section of about 10 km the trapping effect of the near bank groyne field is very strong due to high lateral exchange with groyne fields. About 28 % of total sediment inflow is deposited in groyne fields. Further downstream, at the Mulde mouth the deposition gradient again increases due to sediment inflow from the tributary.

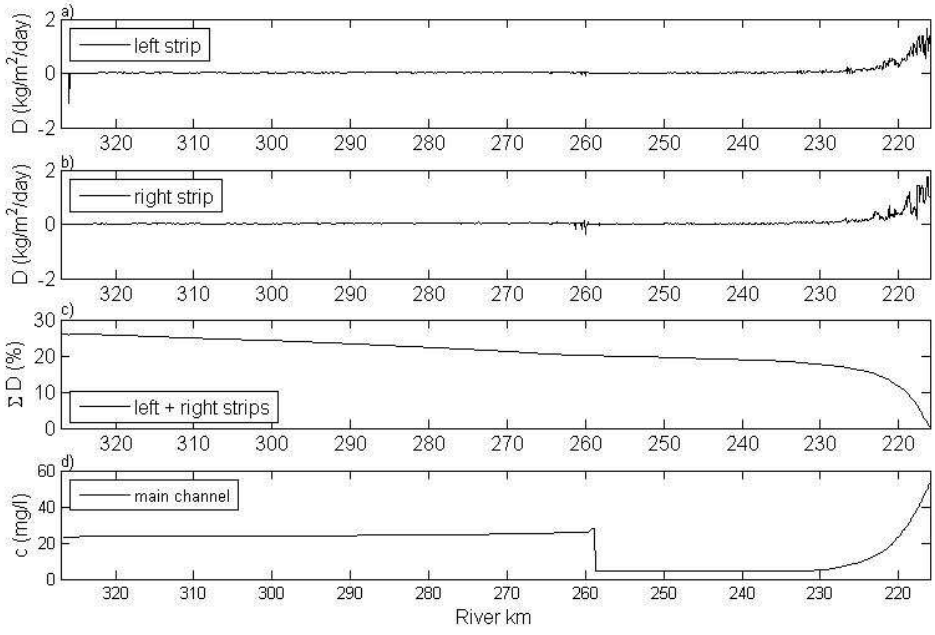


Figure 2. Longitudinal profiles after 4 days, scenario II: a) deposited and eroded mass of sediments per unit area of a side strip per day, for the left and b) right strip; c) cumulative mass of deposited sediments in both side strips in percentage of total sediment inflow; and d) suspended sediment concentration in the main channel.

After four days of constant discharge of 700 m^3/s, around 83 m^3/day of particulate sediments were deposited in the groyne fields. A deposition rate of 159 t/day is smaller compared to the previous scenario I, due to increased discharge and erosion occurrence. Downstream of the Mulde mouth erosion occurs in some groyne fields, see Figs. 2a) and

b). However, the amount of eroded sediments is not big enough to influence the total mass balance. This is due to the fact that erosion occurs in only 1 % of all the groyne fields during the Elbe discharge of 700 m^3/s, which is considered as erosive discharge. As in the previous scenario I, the highest deposition occurs in the upstream section of the river (see Fig. 2d)) at the first 10 km. The deposition rate decreases nearly exponentially with increasing distance from the upstream model boundary, because of the relatively high lateral gradient dispersion of suspended sediments into the dead zones. In Fig. 2c) is shown that around 25 % of the total sediment inflow was deposited in groyne fields over 4 days, which is somewhat smaller compared to the scenario I. The dilution by the Saale and decrease of the slope at km 290.7 is not clearly indicated due to small discharge increase after the Saale by 16 % only.

3.2 An Extreme Flood Event with Overbank Flow

Recently, severe precipitation for a number of days and snowmelt induced several extreme floods in the Elbe basin. These rare events, which occurred in summer 2002 and spring 2006, caused record water levels and worst flooding in the Elbe since the last 100 years. During the flood in 2002, transport characteristics have changed due to dyke breaching and top layers of fine sediments from groyne fields have been eroded [Baborowski et al., 2004; Schwartz, 2006]. Therefore, after the flood in 2002 many special measuring programmes were launched to assess sediments and pollutants. Assuming that top layers have been eroded and more resistant layers remained, less erosion could be expected if the same flood discharge occurs again. This was confirmed by measurements during the flood in spring 2006, i.e., in mass balance deposition was higher than erosion. The goal of the performed scenario is to determine transport characteristics of the Elbe during an extreme flood event (scenario III). A four days flood wave was applied starting with a discharge of 1790 m^3/s and reaching a peak of 3500 m^3/s, which corresponds to a 100 years return period flood. Inflow suspended sediment concentration follows the hydrograph.

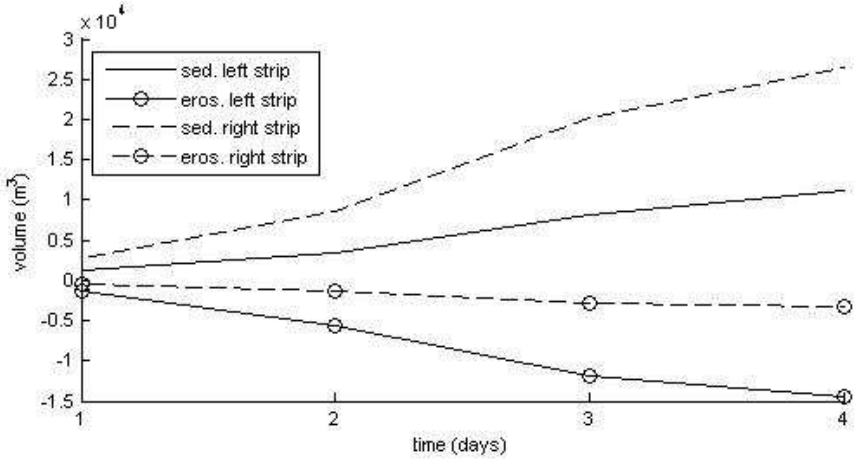


Figure 3. Deposited and eroded sediment volume in the side strips along the 112 km river stretch, scenario III.

In both side strips erosion and deposition occur depending on geometrical and flow conditions, see Fig. 3. With increasing discharge deposition increases due to higher lateral dispersion. However, in some groyne fields erosion increases due to the increase of flow velocity in side strips itself. Maximum increase of both erosion and deposition rates (kg/day) occur during peak discharge. Afterwards, discharge decreases and consequently, erosion and deposition rates decrease as well.

The results show that both erosion and deposition are significantly higher compared to scenario II. The number of groyne fields with erosion is increased to 30 %, whereby 63 % of all groyne fields have deposition and 7 % show both deposition and erosion depending on time. The areas of erosion and deposition can be roughly estimated, refer to Fig. 4. In

the first 35 km mainly deposition takes place, whereby in the next 10 km erosion occurs until the Mulde mouth. Inflow suspended sediment concentration from the Mulde produces predominated deposition, which extend about 100 km downstream. At the end of the simulated river reach of 112 km, 33 % of the total inflow sediment mass was deposited. However, it is assumed that most of the sediments are deposited in flood plains. Keeping in mind the simplification made by the multi-strip model that side strips of the model comprise both the groyne fields and the flood plains, sediments in the groyne field might be eroded and deposited on flood plains. However, in total mass balance for the side strip deposition occurs.

The numerical estimation of the net deposition can be compared with the measurements during the spring flood in 2006, which had a maximum discharge of 3000 m^3/s at Rosslau (km 258) and lasted for 14 days. The inflow of suspended sediments from the Mulde and Saale was measured together with the suspended sediment concentration at Rosslau and Magdeburg (km 326.6). A robust estimate of total mass balance for the river stretch of 69 km indicated that a net deposition occurred during the flood, i.e., that 27 % of total sediment inflow was deposited in the domain, mainly on floodplains. In summary, the simulated extreme flood had the same peak discharge as the measured one and good agreement could be found.

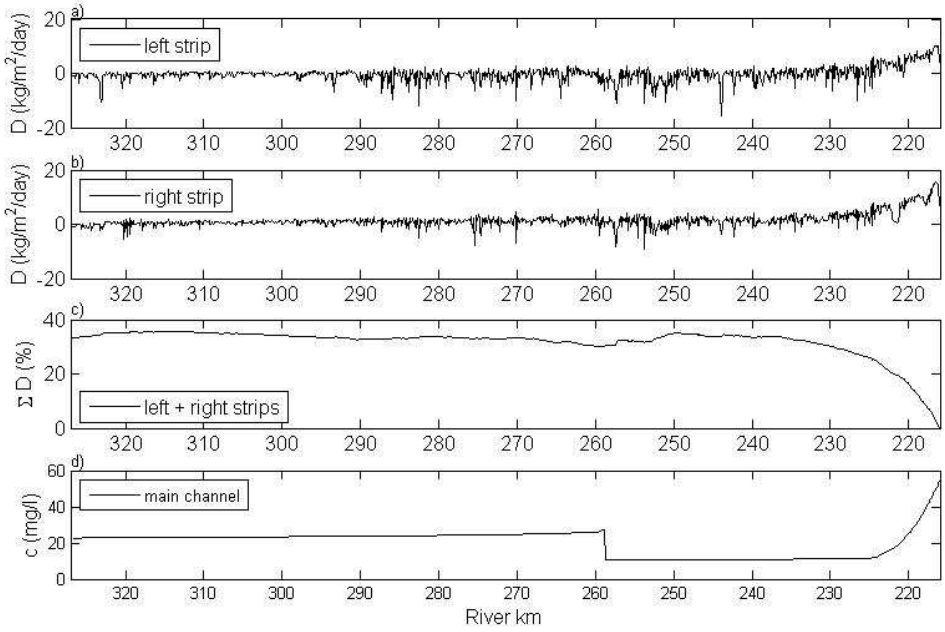


Figure 4. Longitudinal profiles after 4 days, scenario III: a) deposited and eroded mass of sediments per unit area of a side strip per day, for the left and b) right strip; c) cumulative mass of deposited sediments in both side strips in percentage of total sediment inflow; and d) suspended sediment concentration in the main channel.

3.3 Impact of a Particulate Contaminant Load Associated with a Flood in the Tributary

The goal is to estimate the effect of particulate contaminant inflow from a tributary and its influence domain in the Elbe (scenario IV), excluding adsorption and degradation processes. Highly contaminated deposits in the groyne fields along the Elbe presumably originate from the tributary Mulde, which discharges high concentrations of arsenic and lead [Kraft, 2007]. Therefore, a hydrological flood scenario was applied on the Mulde in order to determine the effect and spatial extension of its influence with regard to potential deposition of particulate contaminant. It is important to emphasize that the 1D multi-strip

model cannot account for the lateral mixing in the main channel, which can be up to about 40 km in the case of a river trained by groynes [van Mazijk, 1996]. In the multi-strip model a tributary inflows into the main channel and a fully mixed situation occurs resulting in deposition in both side strips. In a reality, deposition would occur mainly in the left groyne fields until a fully mixed situation is reached.

A four days flood scenario was applied on the Mulde, starting with a mean discharge of 64 m^3/s and reaching a peak discharge of 150 m^3/s after one day. Water levels were below the groyne crests along 112 km of the Elbe reach, meaning that the contribution of the Mulde flood on the Elbe discharge was insufficient to cause overtopping of the groynes. Suspended sediment concentration from the Mulde follows the discharge hydrograph.

The results show that the flood produces an immediate deposition in groyne fields along the Elbe. Increased discharge in the Mulde increases the flow velocities in the Elbe main channel and thus, the lateral transport into the adjacent strip is higher and consequently the deposition in groyne fields. Deposition patterns follow the concentration decay in the main channel (Fig. 5) whereby higher suspended sediment concentration leads to a higher concentration difference between main channel and groyne fields and consequently to a higher deposition in side strips. At the end of the simulated river reach deposition in groyne fields is still taking place. The longitudinal dimension of the deposition process can be estimated on 90 km. Furthermore, the cumulative mass of deposited sediments in side strips (Fig. 5c)) indicates that deposition ceases about 20 km downstream of the model boundary. It is important to emphasize that 44 % of the total sediment inflow is deposited in groyne fields at the end. In summary, the Mulde flood has an impact on deposition in groyne fields, which is relevant for contaminant management strategies.

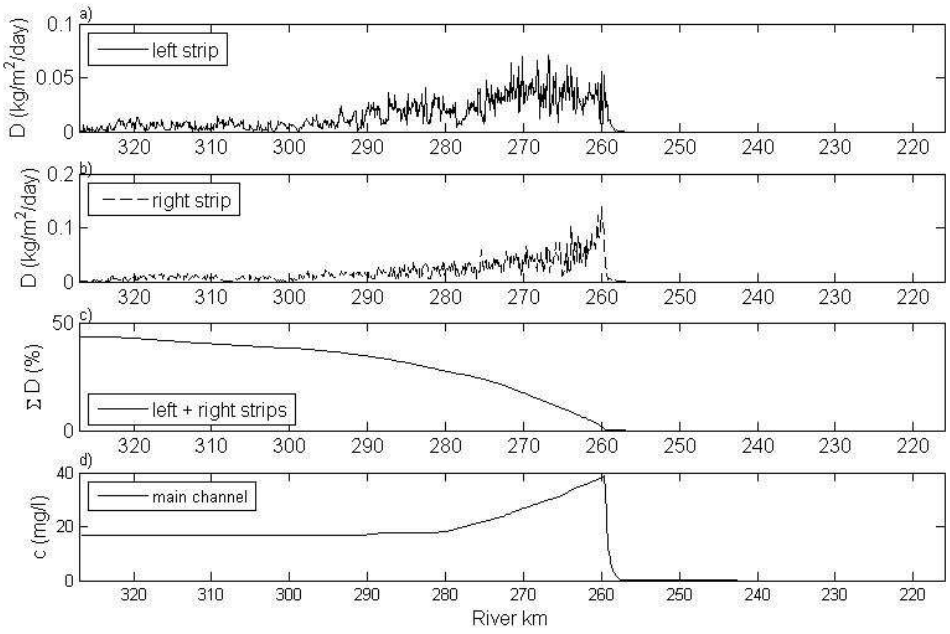


Figure 5. Longitudinal profiles after 4 days, scenario IV: a) deposited and eroded mass of sediments per unit area of a side strip per day, for the left and b) right strip; c) cumulative mass of deposited sediments in both side strips in percentage of total sediment inflow; and d) suspended sediment concentration in the main channel.

3.4 Estimates of Erosion Probability

The results of the performed hydrologic scenarios allow an estimation of the erosion probability. The frequency of erosion can be easily derived from discharge hydrology by hydraulic flow model. The proposed method uses a cumulative frequency distribution of

the discharge, which correlates to the cumulative frequency of erosion in order to estimate the probability of erosion in groyne fields. The cumulative frequency curve at Wittenberg is obtained based on measurements from 1936 to 1995. However, data from the recent extreme flood events are not included.

Flow velocity and water depth in the main channel are significant hydraulic parameters to characterize the river flow, refer to Fig. 6. Therefore, the results are shown in terms of the main channel flow velocity and the water depth to enable a transfer of results to other rivers with similar characteristics. Erosion starts in some groyne fields at low main channel flow velocity and it increases with increasing flow velocity, see Fig. 6b). The results indicate that during the period of low discharge, i.e., with less then one year return period, erosion of groyne field sediments is low. Increased discharges produce higher and intensive erosion however, their probability of occurrence is lower. Although the analysis presented here is simplified, it clearly illustrates the large spatial and temporal variation of sedimentation and resuspension in groyne fields.

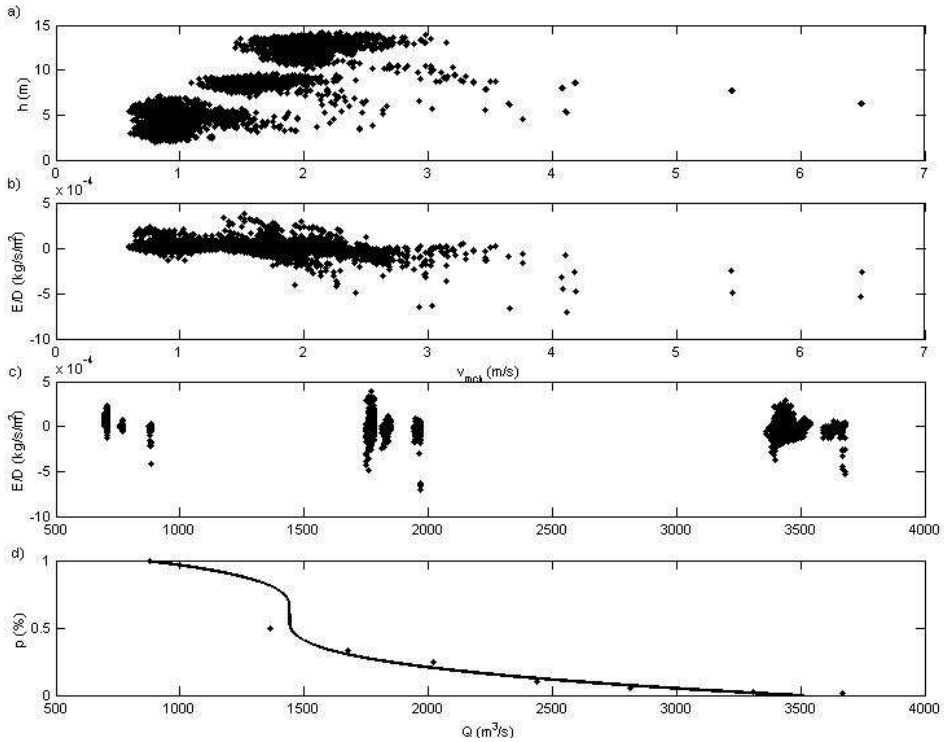


Figure 6. Overview of significant hydraulic parameters for the whole investigated river stretch of 112 km at mean, high, and extremely high discharge: a) water flow depth and flow velocity in the main channel; b) erosion and deposition rates depending on main channel flow velocity; c) erosion and deposition rates for different discharges; and d) discharge probability curve at Wittenberg.

4. CONCLUSIONS AND REMIDIATION MEASURES

Typical hydrological scenarios were applied on the 112 km long river reach of the Middle Elbe to estimate the longitudinal distribution of particulate suspended sediments, erosion and deposition. Erosion is governed by the Elbe discharge and starts at about 700 m^3/s. Considerable erosion occurs during overbank flow, and increases with increasing discharge. The domain affected by deposition is greater for higher discharges. The number of groyne fields subject to erosion increases with increasing flood, whereas about 30 % of inflow sediments are deposited in the near bank stagnant water bodies.

The Mulde flood strongly effects the deposition in the groyne fields along the Elbe. Keeping in mind the contaminant inflow discharged from the tributary, 44 % of deposited sediments in groyne fields indicate severe potential pollution. The Mulde influence domain is estimated to approximately 90 km, which confirms that the tributary is the main contributor to the Elbe pollution. Even though, the calculations were performed in a conservative way neglecting adsorption and degradation processes, the presented approach gives an estimation of deposition of particulate contaminants.

The assessment of erosion probability is a complex issue that needs to involve the interaction of the physical, chemical and biological sediment properties together with hydraulic bed shear stress. The results presented here as an example, are a preliminary attempt to address the erosion probability.

ACKNOWLEDGEMENTS

This work was supported by European Commission; project MODELKEY (Contract No 511237-GOCE).

REFERENCES

Baborowski M., W. von Tümpling and K. Friese, Behaviour of suspended particulate matter (SPM) and selected trace metals during the 2002 summer flood in the River Elbe (Germany) at Magdeburg monitoring station. *Hydrology and Earth System Sciences*, 8, 135-150, 2004.

Kraft C., Schwermetallbelastung und Elementspeziation in Sedimenten der Mulde und Elbe unter besonderer Berücksichtigung von Arsen und Blei. Ph.D. thesis, University of Braunschweig, Germany, 2007.

Prohaska S., and B. Westrich, One-dimensional multi-strip transport model for long distance sediment transport in navigation channels, case study the River Elbe, paper presented at 23. Conference of the Danubian Countries on the Hydrological Forecasting and Hydrological Bases of Water Management, Belgrade, Serbia, August, 2006.

Schwartz R., Geochemical characterisation and erosion stability of fine-grained groyne field sediments of the Middle Elbe River, *Acta hydrochim. hydrobiol.*, 34, 223 – 233, 2006.

van Mazijk A., One-dimensional approach of transport phenomena of dissolved matter in rivers, Ph.D. thesis, Technical University Delft, 1996.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Challenges of alpine catchment management under changing climatic and anthropogenic pressures

C. de Jong[a], P. Masure[a], and T. Barth[a,b]

[a] The Mountain Institute, University of Savoy, Le Bourget du Lac, France (carmen.dejong@institut-montagne.org)

[a, b] CISM, Master Gaia, University of Savoy, Le Bourget du Lac, France, (barth.thierry@wanadoo.fr)

Abstract: Water resources in mountain areas are particularly important with respect to new and increasing pressures in the headwater areas of catchments. Climate change has resulted in a decrease in precipitation and decrease in groundwater recharge in the western and southern Alps over the last few decades. This is coupled with changing anthropogenic pressures such as an increase in water abstraction for tourism and artificial snow in addition to the water demands of agriculture and hydroelectricity. Water consumption by tourism in winter is several times higher than that of the permanent population. To avoid the development of water shortages with relation to other users, a clear quantification of parameters within the water cycle is necessary. However, measurements of classical hydrological and biological components for model validation are often missing and terrestrial photogrammetry is not yet purposefully applied. An integrated hydrological model is applied to the ski resort of Les Arcs, Bourg-St-Maurice in Savoy, France, taking into account the impacts of artificial snow on torrent discharge under different climatological and anthropogenic scenarios with relation to interbasin water transfer. The classical arguments for the expansion of artificial snow are the necessity of winter tourism to maintain the local and regional economy but no empirical relation exists between snow production and ski frequentation. If the clear trend of increasing temperatures in the mountains continues, less and less reliable snowmaking days will be available and some regions may gradually turn into unprofitable zones depending on wind (e.g. foehn conditions), evaporation, aspect and altitude. New adaptation strategies for diversification of tourism in mountain areas, e.g. through a four seasons approach, have to be developed. Since mountain areas are highly fragile, powerful decision support systems are required to control hydrological modifications and prevent water conflicts. Scientists cooperating with stakeholders will be increasingly confronted with issues such as economical benefits versus environmental impacts. Even before the stage of developing decision support tools with the stakeholders concerned, important work is necessary to develop an interdisciplinary and intersectorial problem consciousness and acceptance. There is an urgent need for sub-scale catchment management plans that take into account local hydrological and economical requirements and conflicts that are linked to, but not overshadowed by large-scale catchment plans.

Keywords: Water management, artificial snow, tourism, hydrological modelling, economy.

1. INTRODUCTION

According to the most recent OECD report, approximately half the world's population will be living under water stress by 2030 if no appropriate measures will be undertaken [OECD, 2008]. Nevertheless, integrated watershed management plans based on modelling approaches do not exist for most regions of the Alps and specific mountain water laws still have to be developed. This is due to the assumption that unlimited water supply exists for all purposes, including household consumption, agriculture, hydroelectricity, industry, tourism and artificial snow production with enough remaining as minimal ecological discharge. However, often this is not the case, in particular at the sub-catchment scale, where water is naturally limited both due to seasonal runoff constraints and lack of well developed groundwater reservoirs as a function of altitude [de Jong 2007]. The rapidly changing cryosphere, with both glacier retreat and changes in seasonality of snow, has considerable effects on hydrology [Böhm et al 2007, de Jong et al in press, Koboltschnig et al 2007] and tourist attractiveness. According to the Alpine Space Operational Programme [2007] water availability has not been a problem in the past but the expected impacts of climate change may considerably change this favourable situation, especially in the southern Alps which may be effected by severe water scarcity in the future. Time series analyses over the past 100 years indicate increasing temperatures and decreasing precipitation in the southern Alps as well as up to 25% decrease of groundwater recharge in certain regions of Carinthia, Austria untouched by human development [Harum 2001]. However, in most catchments, representative climatological and hydrological measuring stations are missing and remote sensing techniques are not purposefully developed or applied [de Jong & Barth 2008]. In addition, there is a lack of modelling of the periodicity of flood and drought events in mountain catchments, primarily caused by a lack of data. Based on historical records, sedimentological and other evidence, their recurrence intervals can be estimated and should form the basis for modelling. This is essential to better predict the effects of future scenarios on the magnitude of such events.

Since water resources are not only influenced by climate change but more importantly by changing anthropogenic use, integrated water management is very important to avoid the development of conflicts that are already emerging locally in several regions of the Alps. Apart from assuming that water is ubiquitous, decision makers often compare water availability, demand and use in small sub-catchments to that of the whole basin, indicating that proportions are nearly insignificant and therefore negligible. By strongly understating the local significance of water resources, this classical argument is used as a pretext to continue unlimited water resources exploitation. However, territorial conflicts and water stress are increasingly developed at the local scale, where space and water is limited and are related to altitude. At high altitudes, water storage is poor (restricted to fractures) and the quantity of rain- or snowfall is limited by the limited surface areas of the basin. Therefore, in the context of integrated catchment management, it is important to change the common assumptions and mentalities of water consumers and managers in mountain catchments, in particular those that are highly frequented by tourists. A fundamental shift in focus is essential away from "whole" catchment concept with its diluting effects to the sub-catchment scale with its intensive local use. Misleadingly, the role of mountains as "water towers" both for mountains catchments and those further downstream is constantly emphasised. As a result, decision makers and managers, especially those involved in the snow making industry, assume that water can be exploited continually even when there is a lack of precipitation and surface water retention. Consequently, groundwater reservoirs are increasingly sapped with possible long term effects such as drying up of springs in areas with vulnerable geology. It is therefore crucial to build up consciousness on the long term spatial and temporal consequences of intensive water use in these highly dynamical and fragile zones. In addition, rapid flow of climate change information from scientists to water managers is critical for planning as is training in non-stationarity and uncertainty of hydroclimatological data [Milly et al 2008].

In this paper the hydrological and ecological effects of dam reservoirs will not be discussed.

2. ANTHROPOGENIC AND CLIMATIC PRESSURES

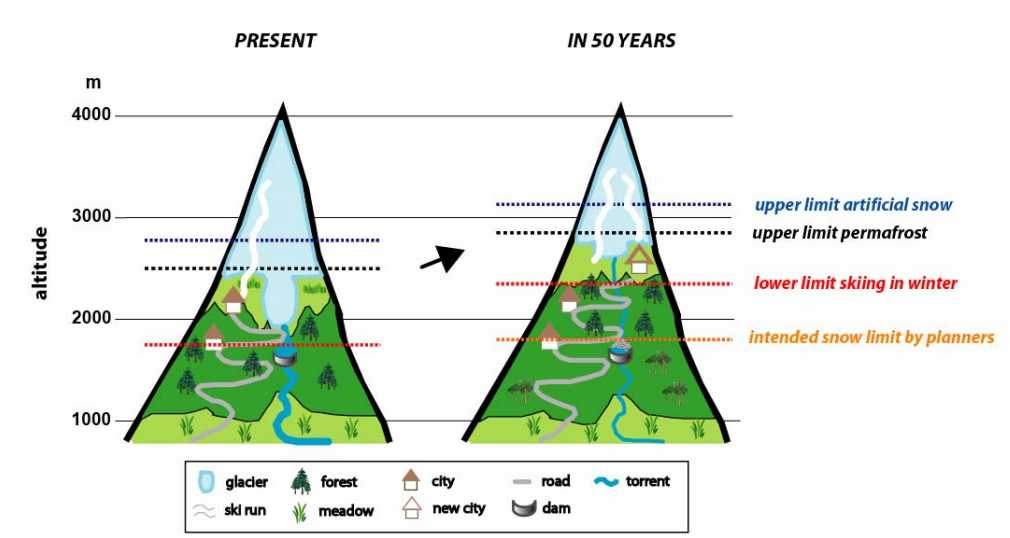


Figure 1 Model of increasing pressures on different mountain zones at present and in 50 years according to global change scenarios adapted to mountain regions. The image represents face of a typical alpine mountain during the winter season. With the recession of glaciers and snowfields, the average limit of the highest city, infrastructure and vegetation will increase; however, these will be constrained by altitude.

Figure 1 illustrates the increasing pressures exerted on different mountain zones at present and those projected in 50 years. It reflects the core assumption by planers associated with the economic development of tourism regions: that the present type of tourism offer, in particular skiing, is maintained. Since climate change and the associated reduction and increased uncertainty of snowfalls have become a reality, skiing is becoming more and more dependant on artificial snow production (Figure 2 & 3). As well as permanently increasing the upper limit of artificial snow production, planners also want to ensure the preservation of the present lower limit to avoid any disruptions. However, the maintenance of this lower limit will not be possible for two reasons. Firstly, under the current and future predicted effects of global warming, conditions will be too warm to manufacture or maintain artificial snow down to the lower limit.



Figure 2 Image of Les Menuires on 25th October 2007 several weeks before the onset of winter. Note the newly produced fields of artificial snow in preparation for the oncoming winter above the new tourist housing area. Intensive slope failure (grey areas) on bare ski runs is visible at the left of the photo.

Secondly, water resources will not suffice to satisfy both the needs of the increasing spatial and temporal extend of snow covered surfaces as well as the growing water demand by tourists. Snow production is being pushed to higher and higher limits, including over glaciers. Available water resources become more and more limited at these altitudes, restrained by the reduced catchment area and shorter discharge season. To compensate the lack of water, water is increasingly pumped over longer distances, either from the valley floors up the slopes to where it is re-distributed, or via inter-basin transfer. This has serious consequences on the hydrological cycle, in particular on flood discharge and the seasonal distribution of discharge (Section 4).

3. WATER CONSUMPTION AND TOURISM



Figure 3 a) Vine irrigation in the Upper Rhône, Valais, Switzerland (Photo: C. Dupont) and **b)** artificial snow production on ski runs in Savoy, France (Photo A. Marnezy), whose water consumption is approx. 3 x higher than irrigated maize or 4 x higher than vine.

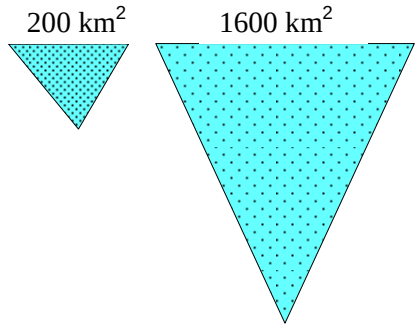


Figure 4 Catchment of 1600 km^2, Upper Rhône, with a relative winter discharge of 10 L/s/km^2 and tourist density of 190 beds/km^2. Its subcatchment of 200 km^2 has the same relative discharge but 300 tourist beds/km^2 Dot density indicates tourist density.

It is now estimated that there are more than 340 million tourist overnight stays in the Alps, of which approximately half are in summer [Keller & Förster 2007]. Some higher altitude ski tourist resorts count more than 50 000 overnight stays in winter. Therefore it is important to compare the hydrological pressure exerted by tourists at different scales. At the subcatchment scale, the actual amount of water consumed may not be high compared to that of the total catchment (Figure 4) however, the relative amount of water consumed by the number of tourists compared to the available resources is nearly double as high. This explains why water conflicts have started to arise at the local, subcatchment level.

Differences in water consumption are presented (Fig. 5) for Albertville, a holiday gateway town with a population of 18000, a typical non ski village, Les Chapelles with a population of 400 and two typical tourist resorts, Bourg St. Maurice and St. Martin de Belleville, with a permanent population of 6750 and 2532 respectively. Table 1 gives an indication of the maximum drinking water consumed per person during the month of December, with the highest tourist frequentation of the year. With a tourist population 5 times higher than its permanent population over a period of 2 weeks, the ski resort of Bourg St. Maurice consumes triple the amount of water for December. Note that the relative consumption of water of a ski tourist resort compared to a non-ski resort is

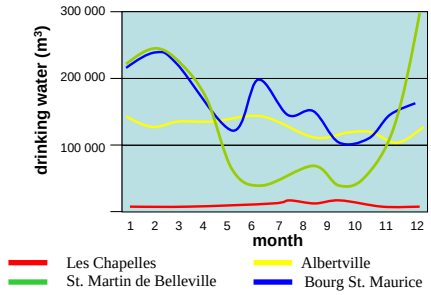


Figure 5 Seasonal water consumption for the city of Albertville compared to a non-ski resort (Les Chapelles) and two larger ski resorts (St. Martin de Belleville and Bourg St. Maurice) (modified after APTV 2007).

Table 1 Maximum drinking water consumption for the month of December

Town	permanent/ average tourist population	Max. drinking water in winter (m^3)	Ratio water (m^3)/ person
Albertville	18 000	140 000	7.8
Les Chapelles	400	2 000	5
Bourg St. Maurice	6750 20 000	160 000	23.7/ 8

1.5 times higher. In the region of Flaine, Haute Savoy, France, the increase in the number of overnight stays and the associated increase in water consumption does not allow sufficient residual flow in the Vernant torrent for artificial snow production [CIPRA, 2007]. Since stream discharge is to be reserved solely for potable water, groundwater reservoirs are being sampled for potential exploitation for snow production. Their potential may be limited due to their kartic nature. Apart from tourism, agriculture can also be water intensive, for example irrigation of vines on slopes (Figure 3) and maize in the valleys. In Savoy, French Alps, the reduction in precipitation over the last 6 years has seriously menaced the livelihood of cattle farmers both due to the direct shortage of drinking water for cattle as a result of drying up of springs and due to the reduction in good quality pasture as a basis for cheese production [Blandon per comm.].

4. HYDROLOGICAL AND ECONOMICAL MODELLING OF SNOWMAKING

4.1 Modelling the hydrological effects of snowmaking in Les Arcs

To manage the exponential expansion of artificial snow production on ski runs in the Alps and its effects on discharge of mountain torrents, a combination of hydro-ecological monitoring and modelling under different scenarios is a prerequisite. Water is increasingly stored in new, high alpine artificial snow-water reservoirs with dimensions resembling medium-sized dam reservoirs. This affects both local hydrology, biodiversity and pasture areas since wetlands and cattle grazing areas are very limited at these altitudes. Water and important nutrients are stored as artificial snow over many months, inevitably subject to high evaporation losses and delayed snowmelt. Such modifications not only affect the regional hydrology but also have direct impacts on drinking water availability. Thus it is essential that integrated catchment management is considered within a wider territorial context in the future.

The “Societé le Montagnes de l’Arc” (SMA or Association of the Mountains of the Arcs) was the first in the region of Savoy to initiate a scientific hydrological study at the University of Savoy on the possible impacts of the interbasin water transfer for artificial snow production in Les Arcs. The impacts of water transfer from the Arc 2000 basin to the Arcs 1800 basin was modelled for six torrents using a global, conceptual, rainfall runoff model, the GR (Genie Rural) developed by the CAMAGREF and adapted by Barth in 2007. This low parameter model is run at daily time steps and considers the catchment as a black-box consisting of a number of reservoirs, with one known input, precipitation and one known output, discharge. Module optimisation is difficult, since no discharge measurements are available for the different torrents. Therefore, parameters are optimised mathematically and return periods with similar discharge are classified according to return periods with similar precipitation. Daily and maximum discharge are then calculated for the different torrents. A special semi-distributive module for the simulation of melt from manmade snow was developed [Barth 2007] based on the degree-day method and classified according to 100 m iso-altitudinal bands. Its parameters include temperature, precipitation, daily snow melt and the contribution of artificial snow. Surfaces covered by artificial snow are subdivided into different altitudinal bands with relation to the amount of water necessary for snow production. Altitudinal bands with artificial snow are adjusted to those with natural snow, so that excess water produced by snowmelt from artificial snow is added to the natural torrent regime. The hydrological model is then calibrated with measured discharge data from the adjacent torrent Pissevieille (Arcs 2000), on the basis of the specially developed snow module and additional meteorological data.

The results show that the difference in natural and artificial snowmelt induced discharge are negligible for the autumn and winter months (Figure 6) but significant for the summer months, especially July and August [de Jong & Barth 2008]. The possible evolution of monthly discharge under different climatic and snow-making scenarios show that during the peak snowmelt months between April to June, discharge increases by 10-20% due to melt contribution from artificial snow. During mild winters the impact of artificial snow is strongest. In all scenarios, a 20-30% increase in discharge due to the anthropogenic influence of snowmaking is apparent between the months of May and August. In general, water abstracted for snowmaking in autumn may be delayed by 8-10 months before returning to the water cycle. Model accuracy should be increased through an improved knowledge of evaporation. In terms of validation and development of existing hydrological models, it is suggested that detailed, low altitude flight experiments are necessary for routine measurements of evaporation loss during artificial snow production. The first experiment of this kind has been carried out in the Austrian Alps [Arabas et al 2008]. In future, the ablation of prolonged snow cover from artificial snow runs should be monitored continually and regionally through radar technology and oblique photography, since remote sensing techniques lack a high enough spatial resolution [de Jong & Barth 2008]. Thus, modelling should be used to synthesise observations but can never replace them [Milly et al 2008].

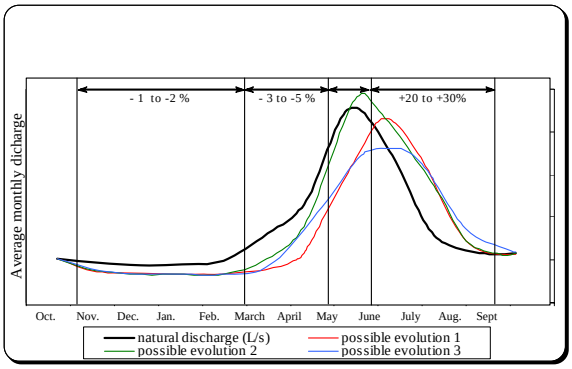


Figure 6 Possible monthly discharge evolution under the impact of artificial snow melt for torrents in Arcs 1600-1800.

4.2 Modelling the economical effects of snowmaking in Grisons

Studies in Grisons, Swiss Alps show that there is no empirical relation between the production of artificial snow and the daily number of skiers over the past 6 years [Teich et al 2007]. This result is important when related to the investment and running costs of snow infrastructure and the amount of water used to produce snow. The attractiveness of winter tourism e.g. in Davos seems more dependant on the general snow situation than on the presence or absence of artificial snow. Therefore in future, instead of seeking for relationships between tourist numbers and artificial snow, it would be important to investigate the relation between tourist numbers and the general snow situation, including natural and artificial snow. According to Teich [2007], snow production can prevent a reduction in winter tourists in poor winters but a classical cost-benefit analysis cannot be carried out since influencing parameters are difficult to parameterise, economical benefits are unequally distributed and it has not been considered whether this money could be better invested elsewhere. Prognoses from scenarios with an average global warming of 1°C for 2030 and 1.8 °C for 2050 for the valley stations in the regions of Davos and Scuol, Switzerland show a reduction in the number of potential snow making days by between 15 to 50 % for the months of November and December [Teich et al 2007]. Thus the reliability of snow making under the predicted future climatic conditions will no longer be feasible for mid-altitude stations (1500m) by the year 2050. Another economical risk is that it takes 15-20 years on average to amortize the costly artificial snow infrastructure, taking into account 100 snowmaking days, i.e. days with subzero temperatures, per season. In future, it may be important to regard climate change as a chance [Masure 2008] rather than a pessimistic scenario that makes strong demands on the maintenance of the cryosphere.

4.3 Relations between drinking water, snow making and minimal flow

The proportion of water used for artificial snow production is increasing at the annual, seasonal and daily scale. Thus, in the Austrian and Swiss Alps, between 20-40% of the total water annual water consumption goes into snow production, which is the equivalent of more than 50% of the total drinking water consumption [Teich et al 2007, Vanham et al 2008]. Furthermore, in some regions of the French Alps, more than 50% of the available drinking water is directly used for snow production at a daily scale. Since the remaining water is used for drinking water it does not essentially cover minimal ecological discharge of local torrents. This underlines once more the small-scale hydrological and ecological facet of the problem. Up to 75% of the minimum low flow discharge can be reduced during the winter due to water abstraction for artificial snow [Strasser 2008, Campion 2002] which has extreme consequences for ecology and downstream water use. If these figures are relativised, the total amount of water consumed for snow production would only amount to 0.5% of the total national drinking water. The situation is particularly problematic when water retention reservoirs for snow production do not contain enough water towards the end of the winter season. Often they are directly refilled from streams or springs which results in their naturally low winter discharge being reduced below the ecological threshold. [Teich et al 2007]. From a hydrological point of view, water is abstracted from the streams when flows are lowest during the winter and released again when flows are naturally high, e.g. during the spring snowmelt. The perturbations to the water cycle can persist practically over the whole year. Water diverted during the autumn and immobilised as snow on the ski runs may not be released back into the water cycle and drink water intakes again until the following summer season [APTV 2007, de Jong & Barth 2008]. Although laws on minimal flow exist for nearly all alpine countries, it is not clear whether they are respected and how they are monitored and controlled. In addition, water agencies and water tax offices assume that between 30-50% of water can be lost by evaporation related to snow production including from retention reservoirs, during actual snow production and by sublimation and evaporation from the snow surface [de Jong 2007]. The seasonal reduction in discharge and evaporation losses can cause real water scarcity problems in certain alpine regions, e.g. downstream of certain ski resorts where nearly all natural discharge and humid zones have disappeared. A common difficulty is that inventories of humid zones often did not exist before water retention reservoirs were built for snow production and that the resolution of remote sensing images is not high enough, so that it is difficult nowadays to reconstruct or access the situation.

5. INTEGRATING QUALITATIVE DATA IN BASIN MANAGEMENT

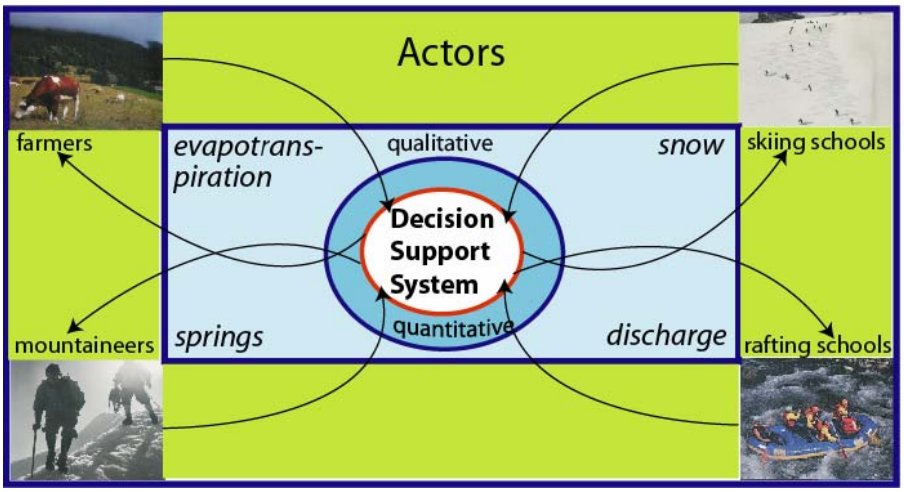


Figure 7 The potential of integration of qualitative data from stakeholders as well as traditional quantitative data for decision support systems.

If research is to be applied in decision support systems, it has to integrate stakeholder knowledge as well as both qualitative and quantitative knowledge (Figure 7). Stakeholder knowledge can be secured through participatory meetings with the local, long-term inhabitants of local mountain villages or towns as well as those organisations that are directly involved in seasonal water related issues, e.g. hydroelectric companies. Examples could include skiing schools (for snow conditions and the duration of snow), rafting schools (for river stage, discharge and flooding characteristics), farmers (related to evapotranspiration e,g for the drying of their hay) as well as information from long-term tourists, hikers and farmers on e.g. the drying up or emergence of new springs. Both "hard" (quantitative) and "soft" (qualitative) data gained from such observations should then be integrated into flexible integrative hydrological modelling frameworks. Such an approach would facilitate the development of decision support systems according to the "soft" needs of the stakeholders involved and not remain rigid with the software-focus of scientists alone.

6. CONCLUSIONS AND PERSPECTIVES

The way forward for integrated watershed management in alpine catchments is to intensify long-term observational networks in combination with short-term in situ measurements and integrate hydrological "soft" (or qualitative) stakeholder knowledge. Care should be taken to expand traditional hydrometeorological stations by adding automated evaporation pans and lysimeters to measure evapotranspiration and by carrying out more direct measurements on plant and forest physiology and their mechanisms of water exchange. With the expected climate change and resulting modification of the quantity and seasonality of snow, the intensification of snow monitoring through snow pillows or snow lysimeters is essential. In addition, the significant spatial and temporal re-distribution of water through anthropogenically induced activities such as snow production or temporary water consumption for tourism should be integrated in future hydrological management and research. Adaptation and mitigation strategies to these new environmental concerns need to be developed. Economical factors related to snow production and tourism have to be analysed and innovative alternatives considered such as diversification of tourism, e.g. through a four seasons approach. In mountain regions, there is an urgent need for sub-scale catchment management plans that take into account local hydrological and economical requirements and conflicts that are linked to, but not overshadowed by large-scale catchment plans. The challenge is to develop an interdisciplinary and intersectorial problem consciousness and acceptance.

ACKNOWLEDGEMENTS

This study was carried out in the frame of the UMS Montagne (Mountains) CNRS 3046 at the Mountain Institute. The modelling was part of the Master thesis of Thierry Barth.

REFERENCES

Alpine Space Operational Programme 2007. European territorial cooperation 2007-2013. http://www.alpine-space.eu/uploads/media/Operational_Programme_ASP_01.pdf.

APTV. Bilan quantitatif de la ressource en eau sur le bassin versant de l'Isère en amont d'Albertville (Quantitative water balance of the Isere cathment upstream of Albertville), (Sepia Conseils, ANTEA, EDACERE) *Report* pp.78, 2007

Arabas, S., P. Paccard, L. Haga, W. Junkermann, B. Kulawik, C. de Jong . Signatures of Evaporation of Artificial Snow in the Alpine Lower Troposphere (SEASALT), *EGU Abstract* 11002, April 2008.

Barth, T. "Etude de l'impact du transfert d'eau du basin versant d'Arc 2000 au basin versant des Arcs 1600-1800, dans le cadre de l'enneigement culture." (Study of the impact of water transfer of the basin of Arc 2000 to Arcs 1600-1800 in the frame of

artificial snow production) *Master Thesis* Science and Technology. Applied Mountains Sciences, University of Savoy, pp. 286, September 2007.

Böhm, R., W. Schöner, I. Auer, B. Hynek, Ch. Kroisleitner, G. Weyss, *Gletscher im Klimawandel*. Vom Eis der Polargebiete zum Goldbergkees in den Hohen Taurern. ZAMG. pp. 111, 2007

Campion, T. 2002 Impact de la neige de culture. Agence de l'eau Rhône Méditerranée Corse. *Report*, pp. 67.

CIPRA Water scarcity has heavy consequences. *Alpmedia newsletter*, 5, March 2007.

de Jong, C. & T. Barth, Challenges in Hydrology of Mountain Ski Resorts under Changing Climatic and Human Pressures. *ESA Proceedings*, Geneve, pp. 6, 2008.

de Jong, C. "Artificial snow drains mountain resources" *EnvironmentalResearchWeb*, Talking Point Article. http://environmentalresearchweb.org/cws/article/opinion/30703, August 2007.

de Jong, C, R. Essery, and D. Lawler (eds) Mountain Hydrometeorology and Snow Seasonality. Special Issue. *Hydrological Processes*, in press.

Harum, T. Abschätzung der nachhaltig nutzbaren Quellwasserdargebotes im alpinen Raum Österreichs (Estimation of the available sustainable spring water offer in the Austrian Alps). *Wasserwirtschaftskataster*, Bundesministerium für Land, Forstwirtschaft, Umwelt und Wasserwirtschaft, Vienna, 2001The Innsbruck University Press Monographs. 2007 .

Keller, L. and K. Förster, 1 x 1 der Alpen: 101 Regionen von Monaco bis Wien, *The Innsbruck University Press Monographs*, pp. 496.

Koboltschnig, G.R, W. Schöner, M. Zappa, and H. Holzmann. Contribution of glacier melt to stream runoff: of the climatically extreme summer of 2003 had happened in 1979. *Annals of Glaciology*, 46 2007.

Masure, P. Le changement climatique, une opportunité. (Climate Change, an opportunity). *Les dossiers de demain*. L'agence d'urbanism de la region Grenobloise. No. 6, January 2008.

Milly, P.C.D., Batabcourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmeier, D.P., Stouffer, R.J.. Stationarity is dead: Whither water management? *Science*, Climate Change Policy Forum, Vol. 319, February 2008, p. 573-574.

OECD, *Environmental Outlook to 2030*, March 2008.

Teich, M., C. Lardelli, P. Bebi, D. Gallati, S. Kytzia, M. Pohl, M. Pütz, and Ch. Rixen, "Klimawandel und Wintertourismus: Ökonomische und ökologische Auswirkungen von technischer Beschneiung. Report, *Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL*, Birmensdorf, pp. 169, 2007.

Vanham, D. E. Fleischhacker and W. Rauch **"**Technical Note: Seasonality in alpine water resources management – a regional assessment"**.** *Hydrol. Earth Syst. Sci*., 12, 91-100, 2008.

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Ecosystem Stability in Integrated Catchment Management

P. A. Khaiter[a] **and M. G. Erechtchoukova**[a]

[a] *The Atkinson Faculty of Liberal and Professional Studies, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3,Canada (pkhaiter/marina@yorku.ca)*

Abstract: Freshwater is vital for human civilization. Growing population and societal development demand for more water resources. Freshwater is not considered globally scarce, but its distribution across the planet is not even and many countries suffer from water shortage. Human activities constantly produce an impact upon natural waters causing their destruction, degradation and pollution. Freshwater is a multi-purpose resource and this circumstance dictates a need to coordinate the interests of various water users. Integrated catchment management is a framework with the explicit objective to improve natural resources management on a catchment basis. Water resources are an integral part of ecosystems and their sustainable management is only possible within an ecosystem-based approach. It is important to predict ecosystem stability properties in response to exogenous stress. The paper discusses ecosystem stability from mathematical and ecological perspectives and demonstrates invariant patterns in ecosystem stress reactions. Using the typology of the stress reactions, the tasks of integrated catchment management can be viewed as a prediction of a particular type of ecosystem behaviour in response to the expected effects of planned activities.

Keywords: Global warming; Anthropogenic pollution; Freshwater resources; Natural ecosystem; Stability.

1. INTRODUCTION

Water resources play a vital role for human survival, health and development. While being not scarce globally, in many countries, resources of freshwater are limited and insufficient. Freshwaters account for only three per cent of the total planetary water reserve, of which 87 per cent are not accessible. More than two billion people are affected by water shortages in over 40 countries. It is predicted [WCU, 2000] that by the year 2025, water abstractions will increase by 50 per cent from 2,600 to 3,800 km^3/year in developing countries and by 18 per cent from 1,200 to 1,400 km^3/year in developed countries. Human activities negatively affect both water quantity and quality causing destruction, degradation and pollution of natural waters. A compounded stress of population growth, global warming and deforestation results in such extreme events as flooding and drought. For example, human and livestock wastes and airborne pollutants have caused eutrophication and acidification of many water bodies worldwide and contaminated them with pesticides, mercury, other toxic chemicals and pathogens that greatly increase the cost of water treatment and health costs from waterborne illnesses. Two million tones per day of human waste are deposited in water courses. 90 per cent of natural disasters in the 1990s were water related.

According to UNEP data [Water, 1995], 90 per cent of all the world's lakes are suffering from eutrophication as a result of anthropogenic pollution. Southern Scandinavia, north-eastern United States/eastern Canada, and eastern China are declared as areas where acidification of natural waters is a major issue. Projected rapid economic growth and related increase in atmospheric pollutants emission are likely to create future problem areas

in Nigeria, India, Venezuela, southern Brazil and south-east Asia. As predicted by Schindler [2000], acute water problems in the United States, Middle East and other parts of the world will even threaten the security of water-rich countries like Canada.

Recently, these problems are being coupled with and amplified by the cumulative effects of global warming and, in a broader sense, of climatic change. Simulations by 8 different climate models (i.e., CCSR/NIES, CCCma, CSIRO, HADCM3, GFDL, MPI-M, NCAR PCM, NCAR CSM) demonstrated that, under different scenarios of greenhouse gas emission, the average global temperature from 1990 to 2100 is expected to rise between 2.2-4.7°C. The lowest predicted warming is 0.55°C south of South America and the highest is 9.2°C in the Arctic Ocean (Figure 1).

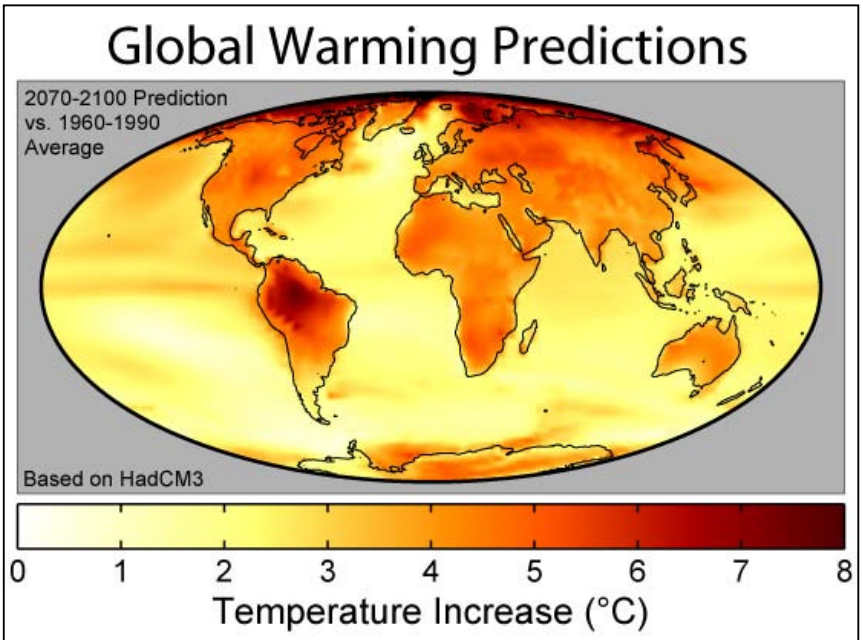


Figure 1. The predicted global warming over the 21st century according to the HADCM3 climate model [HCCPR, 1999].

A critical situation with water resources dictated a necessity to overcome a conventional top-down and sectoral approach that failed to produce desired results and often led to further environmental degradation [WCU, 2000]. Integrated catchment management is viewed as an alternative aimed at getting to sustainable water use.

2. INTEGRATED CATCHMENT MANAGEMENT

Objectively, there are different users of water resources: industry, agriculture, municipal water supply and sanitation, energy generation, recreation, forestry, fishery, shipping, *etc*. Natural waters also play an important environmental role in maintaining the functioning of terrestrial and aquatic ecosystems. A multi-purpose role of water resources and interdependency of their various users necessitate a coordinated approach to all water-related issues including decisions regarding water conservation, restoration, allocation, withdrawal, redistribution, monitoring and management. Such an approach should be able to take into account social, economic, and environmental goals and lead to the achieving of sustainable development.

Over the past two decades, the term *integrated catchment management* has evolved into an umbrella concept for a multi-objective approach to environmental problems with the explicit objective to improve natural resources management on a catchment basis. The

interdependency of water cycle elements and processes of water use requires to consider a basic hydrological unit (i.e., a catchment) as a subject of management efforts.

Geographically, a catchment can be defined as an area of land and water drained by a watercourse and its tributaries to a single defined point. It can range in size from the slope of a hill side through to the runoff area of a major river system. For example, Canada has four major watersheds: Hudson Basin, Pacific, St. Lawrence, and the Arctic. Physically, a catchment is made up of soil, water, air and vegetation; it provides habitat and supports biodiversity. Together these components support life and make up an ecosystem in which a whole range of biophysical processes operates. Carbon based materials (the 'energy' materials), nutrients and water cycle within the catchment area. Each component is linked so that changes made to one will ultimately affect one or more others. Water resources are an integral part of ecosystems and their sustainable management is only possible within an ecosystem-based approach.

Integrated catchment management is a framework to coordinate development and management of water, land and related resources to enhance economic and social welfare without jeopardizing the sustainability of the ecosystems (Figure 2).

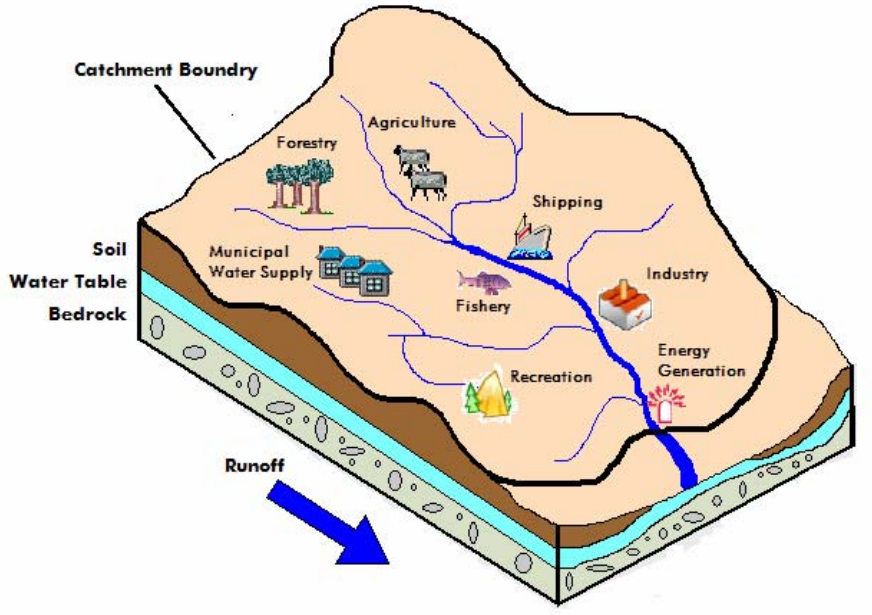


Figure 2. Integrated catchment management.

For practical implementation of the framework, it is important to understand the possible effects of human activities onto the ecosystems and predict ecosystem behaviour under the conditions of anthropogenic stress. The topic of ecosystem dynamics under exogenous disturbances has received considerable attention in the literature mostly within the studies on ecosystem stability.

3. ECOSYSTEM STABILITY CONCEPT

As Odenbaugh [2001] noted, stability has many different meanings and hence ecological theories are imprecise and inapplicable to environmental policy. Moreover, if stability has many different meanings, then theories that employ stability concepts will lead to different conservation strategies. Intuitively, ecosystem stability is understood as an ability to persist in the course of a sufficiently long time in spite of perturbations coming from the environment. Many researchers noted that, though intuitively clear, the notion of ecosystem stability can scarcely be defined in a formal and unambiguous way. A formalization of

ecosystem stability is sought in a well-developed mathematical theory of stability, first of all, in a sense of Lyapunov stability.

3.1 Lyapunov Stability

In general case, an ecological system can be represented at any given moment in time t by a non-negative vector $\mathbf{x}(t) = (x_1(t),..., x_n(t))$ in an abstract *n*-dimensional state space E. The coordinates of vector $\mathbf{x}(t)$ designate quantitatively the components (or sub-systems) of the ecosystem and their properties, such as species numbers and composition, concentrations of organic and inorganic matters and polluting substances, biomass, primary production, dissolved oxygen, nutrients, *etc*.

Due to the complexity of real-world ecosystems, a direct prediction of the system trajectory $\mathbf{x}(t)$ is seldom possible. For this purpose, a mathematical model of an ecosystem is constructed and the question of ecosystem stability or instability, thus, becomes an exercise in mathematical (extreme) properties of a model solution $\bar{\mathbf{x}}(t)$, first of all, in a sense of Lyapunov stability.

Following a unified notation proposed by Ide *et al.* [1997], a model for the natural evolution of the system is governed by an equation:

$$M[t, \bar{\mathbf{x}}(t), \mathbf{p}] = 0 \tag{1}$$

with the initial conditions $\bar{\mathbf{x}}(0) = \bar{\mathbf{x}}_\mathbf{0}$. Here M is the model dynamics operator and $\mathbf{p}$ is the vector of model parameters. Depending on the aim of research, a particular ecosystem being modelled and observation data available, the operator M may be in a form of an algebraic expression, differential or integral operator. Often in ecological applications, M characterizes ecosystem dynamics in terms of ordinary differential equations and then formula (1) can be rewritten as:

$$\frac{d\bar{\mathbf{x}}}{dt} = \mathbf{F}(\bar{\mathbf{x}}(t), \mathbf{p}), \tag{2}$$

where vector-function $\mathbf{F}$ represents the magnitude and change of direction it induces on model trajectory $\bar{\mathbf{x}}(t)$. Points in E for which $\mathbf{F} = 0$ are called equilibrium points and the corresponding solution $\bar{\mathbf{x}}^*$ nullifying the right-hand side of (2) are equilibrium solutions. Lyapunov stability is a property of system behaviour in neighbourhoods of equilibria.

Definition. An equilibrium solution $\bar{\mathbf{x}}^*$ ($\bar{\mathbf{x}}^* \in E_x \subseteq E$) is said to be *Lyapunov stable* if

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\left|\bar{\mathbf{x}}_0 - \bar{\mathbf{x}}^*\right| < \delta \Rightarrow (\forall t \geq t_0)(\left|\bar{\mathbf{x}}(t, \bar{\mathbf{x}}_\mathbf{0}) - \bar{\mathbf{x}}^*\right| < \varepsilon)), \tag{3}$$

where ε and δ are real values, $\bar{\mathbf{x}}(t, \bar{\mathbf{x}}_\mathbf{0})$ designates the solution corresponding to the initial state $\bar{\mathbf{x}}_\mathbf{0}$, and '$\left|\cdot\right|$' designates a Euclidian distance metric on E. If, in addition to (3),

$$\lim_{t\to\infty} \bar{\mathbf{x}}(t) = \bar{\mathbf{x}}^*, \tag{4}$$

it is said that $\bar{\mathbf{x}}^*$ is *asymptotically Lyapunov stable* in E_x. The subspace E_x of E within which the system is (asymptotically) Lyapunov stable is called the (*attraction*) *stability domain* of $\bar{\mathbf{x}}^*$. If the (attraction) stability domain is all of E, $\bar{\mathbf{x}}^*$ is said to be (asymptotically and) globally Lyapunov stable.

3.2 Ecological Stability

Lyapunov defined the concept of stability to describe equilibrium behaviour of the solar system. But his definition is used outside its original context, particularly, to analyse mathematical models of biological communities [Svirezhev and Logofet, 1978] or relationships between stability and complexity of such models [May, 1974].

Justus [2006] noted clear advantages of the definition of Lyapunov stability as it integrates ecological stability into a thoroughly studied mathematical theory and supplies analytical techniques to assess stability properties of community models. He, however, argued that ecological stability should not be defined as Lyapunov stability. Justus [2006] considered four types of Lyapunov stability: (1) local Lyapunov stability; (2) non-asymptotic stability within a non-local domain; (3) asymptotic stability within a non-local domain; and (4) global asymptotic stability and concluded that none of them captures perturbed behaviour of ecological systems because the latter are not completely analogous to physical systems for which Lyapunov's methods are so successfully applied.

It should also be taken into account that a mathematical model of biological communities does not represent an ecosystem as a single whole. From this perspective, a conclusion of an ecosystem stability or instability derived from an analysis of the extreme properties of a mathematical model is a mathematical artefact rather than a real characteristic of ecosystem behaviour and, thus, is a matter of mathematical convenience rather than ecological usefulness [Straškraba, 1995].

Therefore, it is much more important for integrated catchment management to predict the behaviour of the ecosystem state vector $\mathbf{x}(t)$ rather than trajectory of model's solution $\bar{\mathbf{x}}(t)$.

4. ECOSYSTEM STRESS DYNAMICS

Let $\mathbf{x}^A(t)$ be ecosystem state vector under stress conditions. If u denotes management activity that perturbate the ecosystem, its anthropogenic trajectories can be found as:

$$\mathbf{x}^A(t) = \mathbf{F}^A(\mathbf{x}(t), u) \tag{4}$$

It has been demonstrated [Khaiter and Erechtchoukova, 2007] that there are certain common patterns in the behaviour of ecosystems as they respond to anthropogenically caused perturbations and, following classical papers by Holling [1973] and Odum [1983], five scenarios in ecosystem behaviour can be determined.

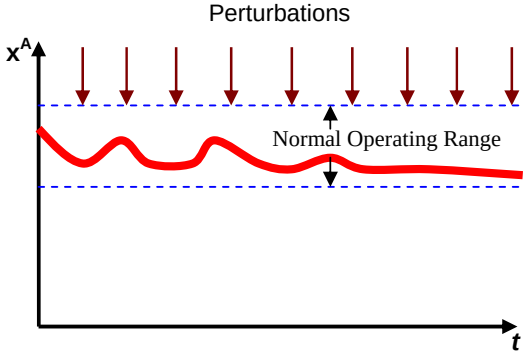


Figure 3. Resistance stability (RT-type of dynamics).

1. *Resistance* (RT-type). RT-type of dynamics occurs when the capacity of a system to sustain perturbations allows it – due to the built-in buffering mechanisms – to maintain its structure and functions within a certain "normal" operating range, and fluctuations of $\mathbf{x}^A(t)$ remain within a bounded domain at any time (Figure 3).

2. *Deformation* (DF-type). DF-type of dynamics represents the situation where external disturbances are applied to the system, which initially resists the perturbation demonstrating, thus, RT-type of dynamics. However, starting from some critical point in time, t_1, the system with DF-type of dynamics will leave its "normal" operating range (Figures 4*a*, *b*).

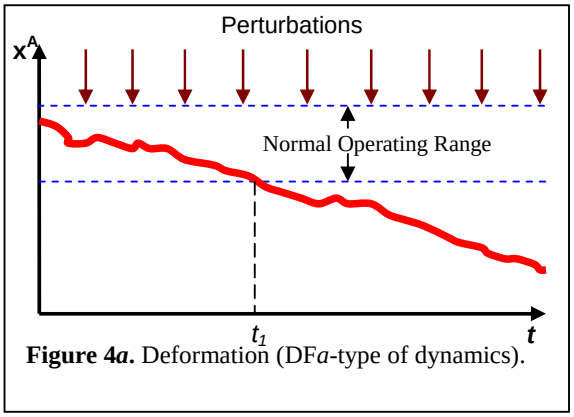


Figure 4*a*. Deformation (DF*a*-type of dynamics).

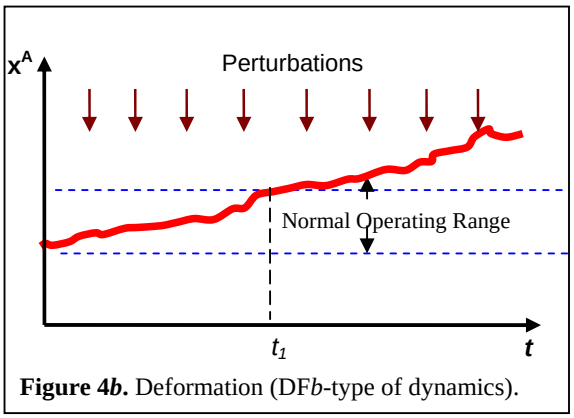


Figure 4*b*. Deformation (DF*b*-type of dynamics).

3. *Resilience* (restoration) RL-type. RL-type of dynamics represents the situation where external disturbances are applied to the system, which initially resists the perturbation exhibiting thus RT-type of dynamics. Starting from some point in time t_1 the

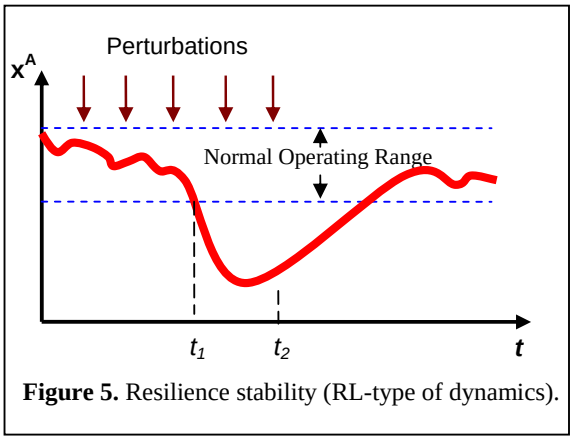


Figure 5. Resilience stability (RL-type of dynamics).

system leaves its normal operating range as in the case of DF-type of dynamics. However, in the subsequent time $t > t_2$, after the stress is terminated, the system is able to recover and return to its "normal" operating range (Figure 5).

4. *Degradation* (DR-type). DR-type of dynamics occurs under the conditions of the previous case (i.e., RL-type of dynamics), except the system stress-compensating capabilities are unable to cope with the stress, recover the damages and return the system to its "normal" operating range even after an external perturbation is discontinued, that is for $t > t_2$ (Figure 6).

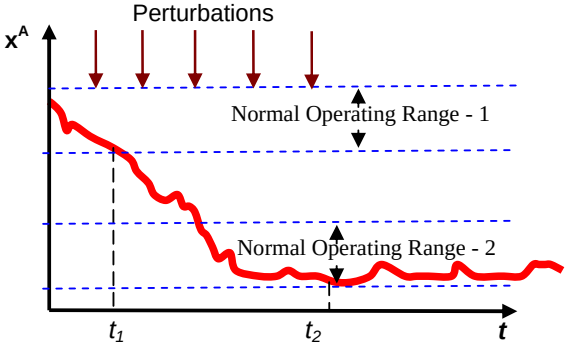


Figure 6. Degradation (DR-type of dynamics).

5. *Shift* (SS-type). SS-type of dynamics represents the situation where the magnitude and/or duration of the stress exceeds the ecosystem ability to sustain perturbations in terms of both RT- or RL-types of dynamics, eventually leading to the destruction of the initial ecosystem S1 and to its replacement by a new ecosystem S2. In most cases, a series of DR-type of changes will take place in the ecosystem S1 preceding the actual transformation S1 → S2 (Figure 7).

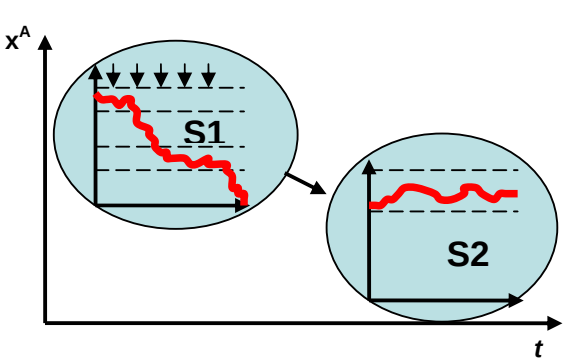


Figure 7. Shift (SS-type of dynamics).

5. CONCLUSIONS

Understanding the ecosystem dynamics and stability properties in response to exogenous perturbations is a crucial part of integrated catchment management. The patterns in the ecosystem behaviour allow for prediction of the under-the- and post-the-stress dynamics of

ecological systems in study. They make also possible to assess the maximum allowable levels of anthropogenic impact the system would sustain without drastic, sometimes irreversible, alterations of its structure and functions as well as to judge on what will happen to the system if a critical level of stress is exceeded.

One of the main tasks of integrated catchment management is a restriction of anthropogenic impact onto the ecosystems so that their rehabilitation abilities were not exceeded, thus, preventing them from deterioration and destruction. A sustainable way of human development and maximum allowable level of anthropogenic load it produces should not, therefore, lead natural objects in question to either DR-type or SS-type of dynamics. In the practice of integrated catchment management, it is necessary to determine the threshold values of disturbances that trigger an ecosystem to DR- or SS-types of dynamics and select strategies generating lesser harm to the ecosystem.

ACKNOWLEDGEMENTS

The authors wish to thank anonymous reviewers for their helpful suggestions and comments on the manuscript. Part of this work was supported by the Atkinson Faculty Council through the 2006-2007 and 2007-2008 individual grants to each co-author.

REFERENCES

Hadley Centre for Climatr Predictions and Research (HCCPR), Climate change and its impacts. Stabilisation of CO in the atmosphere, Met. Office, Bracknell, 1999.

Holling, C.S., Resilience and stability of ecological systems, *Annual Review of Ecology and Systematics*, 4, 1-23, 1973.

Ide, K., P. Courtier, M. Ghil and A.C. Lorenc, Unified notation for data assimilation: operational, sequential and variational, *Journal of the Meteorological Society of Japan*, 75(1B), 181-189, 1997.

Justus, J., Ecological and Lyapunov stability, 2006. URL: http://philsci-archive.pitt.edu/archive/00002987/01/PSA_2006_Justus_10-15-06.pdf

Khaiter, P.A., and M.G. Erechtchoukova, Environmental assessment of anthropogenic impact through the patterns of ecosystem reactions, *International Journal of Environmental, Cultural, Economic and Social Sustainability*, 3(4), 181-190, 2007.

May, R., Stability and complexity in model ecosystems, 2nd ed., Princeton University Press, Princeton, 1974.

Odenbaugh, J., Ecological stability, model building, and environmental policy: a reply to some of the pessimism, *Philosophy of Science*, 68, 493-505, 2001.

Odum, E.P., Basic ecology, Saunders College Pub., Philadelphia, 1983.

Schindler, D.W., The cumulative effects of climate warming and other human stresses on Canadian freshwaters in the new millennium, *Can. J. Fish. Aquat. Sci.*, 58, 18-29, 2000.

Straškraba, M., Cybernetic theory of complex ecosystems, In Complex ecology: the whole-part relation in ecosystems, eds. B.C. Patten and S.E. Jørgensen, Prentice Hall, Englewood Cliffs, N.J., 104-129, 1995.

Svirezhev, Y.M., and D.O. Logofet, Stability of biological communities, Nauka, Moscow, 1978.

Water quality of world river basins, UNEP, Nairobi, (UNEP Environmental Library No 14), 1995.

World Conservation Union (WCU), Vision for water and nature. A world strategy for conservation and sustainable management of water resources in the 21st century, 2000. URL: www.iucn.org