

## Accepted Manuscript

Advances in clustering and visualization of time series using GTM through time

Iván Olier, Alfredo Vellido

PII: S0893-6080(08)00119-6

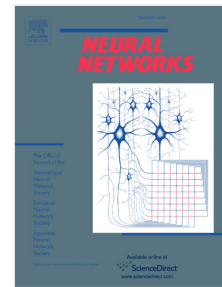
DOI: 10.1016/j.neunet.2008.05.013

Reference: NN 2439

To appear in: *Neural Networks*

Received date: 20 November 2006

Accepted date: 9 May 2008



Please cite this article as: Olier, I., & Vellido, A. Advances in clustering and visualization of time series using GTM through time. *Neural Networks* (2008), doi:10.1016/j.neunet.2008.05.013

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Advances in Clustering and Visualization of Time Series Using GTM Through Time

Iván Olier and Alfredo Vellido

{iaolier, avellido}@lsi.upc.edu

*Department of Computing Languages and Systems (LSI).*

*Polytechnic University of Catalonia (UPC).*

*C/. Jordi Girona, 1-3. 08034, Barcelona, Spain.*

## **Acknowledgements**

Alfredo Vellido is a research fellow within the Ramón y Cajal program of the Spanish Ministry of Education and Science.

## **Corresponding Author**

Alfredo Vellido

*Department of Computing Languages and Systems (LSI).*

*Polytechnic University of Catalonia (UPC).*

*C/. Jordi Girona, 1-3. 08034, Barcelona, Spain*

Phone number : +34 93 4137796

Fax number : +34 93 4137833

e-mail: avellido@lsi.upc.edu

# Advances in Clustering and Visualization of Time Series Using GTM Through Time

## Abstract

Most of the existing research on multivariate time series concerns supervised forecasting problems. In comparison, little research has been devoted to their exploration through unsupervised clustering and visualization. In this paper, the capabilities of the Generative Topographic Mapping Through Time, a model with foundations in probability theory that performs simultaneous time series clustering and visualization, are assessed in detail. The focus is placed on the visualization of the evolution of signal regimes and the exploration of sudden transitions, for which a novel identification index is defined. The interpretability of time series clustering results may become extremely difficult, even in exploratory visualization, for high dimensional datasets. Here, we define and test an unsupervised time series relevance determination method, fully integrated in the Generative Topographic Mapping Through Time model, that can be used as a basis for time series selection. This method should ease the interpretation of the time series clustering results.

**Keywords:** *Multivariate time series; Generative Topographic Mapping; unsupervised relevance determination; clustering; visualization; change point detection.*

## 1. Introduction

The data mining of multivariate time series has long ago become an established research area (Chatfield, 2000). Methods to deal with this problem have stemmed from traditional statistics and also from the machine learning field, where neural networks have provided some of the most fruitful approaches (Zhang, Patuwo & Hu, 1998). These methods usually consider the problem as supervised, being prediction the main goal of the analysis. In comparison, little research has been devoted to methods of unsupervised clustering for the exploration of the dynamics of multivariate time series. It is sensible to assume that, in many problems concerning time series, the states of a process may be reproduced or *revisited* over time; therefore, data grouping or clustering structure is likely to be found in the series. Furthermore, for exploratory purposes, it would be useful to visualize the way these series evolve from one cluster or region of clusters to another over time, as this could provide intuitive visual cues for forecasting as well as for the distinction between mostly stable states, smooth dynamic regime transitions, and abrupt changes of signal regime.

Some of the most interesting time series clustering results have been obtained with different variants of Kohonen's Self-Organizing Map (SOM: Kohonen, 2001) models in diverse contexts although, in general, without accounting for the violation of the independent identically distributed (i.i.d.) condition. Alternatively, SOM has been combined with data pre-processing techniques such as variants of dynamical embedding (Simon, Lee & Verleysen, 2006), which are somehow common in time series clustering in general (Keogh & Lin, 2005) and mostly suitable for univariate time series. Several extensions of SOM have been developed to explicitly accommodate time series, though, mostly through recurrent connectivity (Chappell & Taylor, 1993; Voetglin, 2002; Strickert & Hammer, 2005, Tiño, Farkas & van Mourik, 2006). The SOM was originally defined as a biologically inspired model but has long ago veered away towards general data analysis. Despite attempts to fit it into a probabilistic framework (e.g., Yin & Allinson, 2001; Kostiainen & Lampinen, 2002), it has mostly retained its heuristic definition, which is at the origin of some of its

limitations. On the contrary, the Generative Topographic Mapping (GTM: Bishop, Svensén & Williams, 1998a) is a stochastic latent model (Grabmeier & Rudolph, 2002) that was originally devised as a probabilistic alternative to SOM, aiming to overcome its aforementioned limitations. The GTM, which can also be understood as a constrained finite mixture model, is suited for data clustering but also, as a latent variable model, is embodied with visualization capabilities that are akin to those of the SOM, which have been extensively studied (Vesanto, 1999). Its probabilistic setting has enabled the definition of principled extensions for hierarchical structures (Tiño & Nabney, 2002), missing data imputation (Carreira-Perpiñan, 2000; Vellido, 2006), adaptive regularization (Bishop, Svensén & Williams, 1998b; Vellido, El-Dereby & Lisboa, 2003), discrete data modelling (Bishop, Svensén & Williams, 1998b; Girolami, 2002), robust outlier detection and handling (Bullen, Cornford & Nabney, 2003; Vellido, 2006) and others (Bishop, Svensén & Williams, 1998b).

The GTM Through Time (henceforth referred to as GTM-TT: Bishop, Hinton & Strachan, 1997) is one such extension of GTM. It explicitly accounts for the violation of the i.i.d. condition, given that it is defined as a constrained Hidden Markov Model. The GTM-TT was defined for the exploratory analysis of multivariate time series, but its capabilities for clustering and visualization have never been assessed in detail. In this paper we first intend to carry out such assessment by implementing the GTM-TT model and performing several experiments with a diverse array of publicly available multivariate time series datasets, as well as with synthetically generated ones. These experiments focus on the visualization of the evolution of signal regimes and the exploration of sudden transitions or change points, for which a novel identification index is defined.

Feature selection, as a data reduction technique, plays an important role in pattern recognition and data analysis. In the context of multivariate time series analysis, one approach to data reduction would be the selection of a subset of time series on the basis of a relevance ranking. Some strides have already been made in feature selection for time series prediction (See, for

instance, Bengio & Chapados, 2003) but little has been accomplished in unsupervised time series clustering settings (one exception is the recent work by Yoon, Yang & Shahabi, 2005). Data visualization can be especially important in the exploratory stages of an analytical data mining process (Wong, 1999), but the interpretation of the GTM-TT clustering results through exploratory visualization might become difficult for data sets consisting of a large number of time series. The data analyst would benefit from a method that allowed ranking the time series according to their unsupervised relative relevance and, ultimately, from a time series selection method based on it. Recently, an unsupervised feature relevance determination method for GTM was defined in (Vellido, 2005; Vellido, Lisboa & Vicente, 2006). Individual data features are deemed relevant by this method only if they explain the specific clustering structure provided by GTM. In this paper, we extend this approach and define a method for time series relevance determination (TSRD) that is an integral part of the GTM-TT model data fitting process. The proposed relevance determination method should therefore ease the interpretation and improve the actionability of the time series clustering results.

The rest of the paper is structured as follows: In section 2, an introduction to the GTM as a constrained mixture of Gaussians is provided; this is followed by a description of the GTM-TT variation on the standard model and by the definition of the time series relevance determination method for GTM-TT. In section 3, several experiments for the assessment of the GTM-TT performance are first described, and their results presented and discussed; then, the performance of the proposed TSRD method for GTM-TT is tested and the corresponding experimental results discussed. The paper wraps up with a conclusion section.

## **2. Generative Topographic Mapping and its extensions**

The GTM was originally conceived as a model that could provide most of the functionality of SOM, while overcoming some of its limitations through a probability theory-based definition. In

this section, we provide a brief introduction to the standard GTM model and its extension for multivariate time series analysis, followed by the definition of its extension for TSRD that would be the basis for series selection procedures.

### 2.1 The standard Generative Topographic Mapping for static data

The neural network-inspired GTM is a nonlinear latent variable model of the manifold learning family with sound foundations in probability theory. It performs simultaneous clustering and visualization of the observed data through a nonlinear and topology-preserving mapping from a visualization latent space in  $\mathfrak{R}^L$  (with  $L$  being usually 1 or 2 for visualization purposes) onto the  $\mathfrak{R}^D$  space in which the observed data reside. The mapping that generates the embedded manifold takes the form:

$$\mathbf{y} = \mathbf{W}\Phi(\mathbf{u}) \quad [1]$$

where  $\mathbf{u}$  is an  $L$ -dimensional point in latent space,  $\mathbf{W}$  is the matrix that generates the mapping, and  $\Phi$  consists of  $S$  basis functions  $\phi_S$  (radially symmetric Gaussians in the standard model for continuous data). To achieve computational tractability, the prior distribution of  $\mathbf{u}$  in latent space is constrained to form a uniform discrete grid of  $M$  centres, analogous to the layout of the SOM units, in the form:

$$p(\mathbf{u}) = \frac{1}{M} \sum_{i=1}^M \delta(\mathbf{u} - \mathbf{u}_i) \quad [2]$$

This way defined, the GTM can also be understood as a constrained mixture of Gaussians model. A density model in data space is therefore generated for each component  $i$  of the mixture, which, assuming that the observed data points  $\mathbf{x}_n$  are i.i.d., leads to the definition of a complete log-likelihood in the form:

$$L_c(\mathbf{W}, \beta | \mathbf{X}) = \sum_{n=1}^N \ln p(\mathbf{x}_n | \mathbf{W}, \beta) = \sum_{n=1}^N \ln \left\{ \frac{1}{M} \sum_{i=1}^M p(\mathbf{x}_n | \mathbf{u}_i, \mathbf{W}, \beta) \right\} \quad [3]$$

where,

$$p(\mathbf{x}_n | \mathbf{u}_i, \mathbf{W}, \beta) = \left( \frac{\beta}{2\pi} \right)^{D/2} \exp \left\{ -\frac{\beta}{2} \|\mathbf{y}_i - \mathbf{x}_n\|^2 \right\} \quad [4]$$

In Eq. 4,  $\mathbf{y}_i = \mathbf{W}\Phi(\mathbf{u}_i)$  is a  $D$ -dimensional point in the manifold embedded in data space: the centre of the  $i^{\text{th}}$  constrained mixture component, while  $\beta$  is the estimated common inverse variance of the isotropic Gaussian distributions in data space whose centres are  $\mathbf{y}_i$ . The adaptive parameters of the model can be optimized using the EM algorithm within the Maximum Likelihood framework. Matrix  $\mathbf{W}$  is updated as the solution of the following system of equations:

$$\Phi^T \mathbf{G}_{old} \Phi \mathbf{W}_{new}^T - \Phi^T \mathbf{R}_{old} \mathbf{X} = 0 \quad [5]$$

where  $\Phi$  is a  $M \times S$  matrix with elements  $\phi_s(\mathbf{u}_i)$ ;  $\mathbf{R}$  is the *matrix of responsibilities*, with elements  $\mathbf{R}_{in} = p(\mathbf{u}_i | \mathbf{x}_n)$  that define the probability of the data point  $\mathbf{x}_n$  being generated by the latent point  $\mathbf{u}_i$ ;  $\mathbf{G}$  is a matrix with elements  $\sum_{n=1}^N R_{in}$  in the diagonal and zeros elsewhere; and  $\mathbf{X}$  is the observed data matrix. Finally, parameter  $\beta$  is updated according to:

$$(\beta^{new})^{-1} = \frac{1}{ND} \sum_{n=1}^N \sum_{i=1}^M R_{in} \|\mathbf{y}_i - \mathbf{x}_n\|^2 \quad [6]$$

See (Bishop, Svensen & Williams 1998a,b) for further details on all these calculations.

## 2.2 Generative Topographic Mapping Through Time: the GTM-TT

Multivariate time series are not i.i.d. data and, therefore, the standard definition of the GTM summarized in the previous sub-section can only provide a rough approximation to their proper

modelling. A variation on the standard model was defined as a topology-constrained Hidden Markov Model (HMM) in (Bishop, Hinton & Strachan, 1997) to deal with this limitation, namely the GTM Through Time.

In GTM-TT, points  $\mathbf{u}_i$  in latent space are considered as hidden states and temporal dependencies are captured through the coupling of these latent points. Furthermore, the emission probabilities are controlled by the GTM mixture distribution.

The probability of an observation sequence  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ , given a fixed state sequence  $Q = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n, \dots, \mathbf{q}_N\}$  is defined as:

$$p(X | Q) = \prod_{n=1}^N p(\mathbf{x}_n | \mathbf{q}_n = \mathbf{u}_i), \quad [7]$$

where  $\mathbf{q}_n$  is the state  $\mathbf{u}_i$  at time  $n$ . The likelihood of the GTM-TT is defined as the sum of probabilities of the observation sequence given of all possible state sequences, or paths:

$$p(X, \lambda) = \sum_{\text{all } Q} \pi_{\mathbf{q}_1} \prod_{n=1}^{N-1} P_{\mathbf{q}_n \mathbf{q}_{n+1}} \prod_{n=1}^N p_{\mathbf{q}_n}(\mathbf{x}_n), \quad [8]$$

where  $\pi_{\mathbf{q}_1}$  defines the initial state probability of  $Q$ ;  $P_{\mathbf{q}_n \mathbf{q}_{n+1}} = P[\mathbf{q}_{n+1} = \mathbf{u}_j | \mathbf{q}_n = \mathbf{u}_i]$  is the probability of transition from one hidden state to another (and therefore captures the temporal dependencies);  $p_{\mathbf{q}_n}(\mathbf{x}_n) = p(\mathbf{x}_n | \mathbf{q}_n = \mathbf{u}_i)$  is the probability found in Eq.4; and  $\lambda = \{\{\pi_i\}, \{P_{ij}\}, \mathbf{W}, \beta\}$  is the set of model parameters.

As in HMM, the likelihood defined above can be efficiently calculated using the *forward-backward procedure* (Baum & Egon, 1967). The probability of being in the state  $\mathbf{u}_i$  at time  $n$ , given the observation sequence and the model, also known as *responsibility*  $\mathbf{R}_m$  is calculated as:

$$R_m = P(\mathbf{q}_n = \mathbf{u}_i | X, \lambda) = \frac{A_{in} B_{in}}{P(X | \lambda)}. \quad [9]$$

The forward variable  $A_{in}$  is the joint probability of the past subsequence  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  and the state  $\mathbf{q}_n = \mathbf{u}_i$ , i.e.  $A_{in} = P(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}, \mathbf{q}_n = \mathbf{u}_i | \lambda)$ , given by the following recursive equation:

$$A_{in} = \left( \sum_{k=1}^M A_{k,n-1} P_{ki} \right) p_i(\mathbf{x}_n) \quad [10]$$

where  $A_{i,1} = \pi_i p_i(\mathbf{x}_1)$ . The backward variable  $B_{in}$ , which is the probability of the future subsequence  $\mathbf{x}_{n+1}, \mathbf{x}_{n+2}, \dots, \mathbf{x}_N$  given hidden state  $\mathbf{q}_n = \mathbf{u}_i$ , i.e.

$B_{in} = P(\{\mathbf{x}_{n+1}, \mathbf{x}_{n+2}, \dots, \mathbf{x}_N\} | \mathbf{q}_n = \mathbf{u}_i, \lambda)$ , is calculated using the following recursive equation:

$$B_{in} = \sum_{k=1}^M P_{ki} p_k(\mathbf{x}_{n+1}) B_{k,n+1} \quad [11]$$

where  $B_{iN} = 1$ .

In addition to parameters  $(\mathbf{W}, \beta)$ , which can be obtained in the M-step of the EM algorithm as for the standard GTM, GTM-TT modelling entails the estimation of the initial state probabilities  $\{\pi_i\}$  and the state transition probabilities  $\{p_{ij}\}$ . In order to describe the procedure for the re-estimation of this parameters, we first define  $\xi_n(i, j)$ : the joint probability of hidden state  $\mathbf{u}_i$  at time  $n$  and hidden state  $\mathbf{u}_j$  at time  $n+1$ , given the data  $X$  and the model. Then, the re-estimation formulae are defined as follows:

$$\hat{\pi}_i = R_{i1}, \quad [12]$$

$$\hat{P}_{ij} = \frac{\sum_{n=1}^{N-1} \xi_n(i, j)}{\sum_{n=1}^{N-1} R_{in}}. \quad [13]$$

As mentioned in the introduction, the GTM is embodied with visualization capabilities that are akin to those of the SOM. The clusters of multivariate time series points can be summarily visualised in the low-dimensional latent space (in 1 or 2 dimensions) of GTM-TT by means of the

posterior-mode projection (Bishop, Svensén & Williams, 1998a), defined as:

$$\mathbf{q}_n^* = \arg \max_{1 \leq i \leq M} R_m \quad [14]$$

The optimum path over the space of states is defined by the state sequence  $Q^* = \{\mathbf{q}_1^*, \mathbf{q}_2^*, \dots, \mathbf{q}_N^*\}$ .

Beyond the posterior mode, the distribution of the *responsibility* over the latent space of states can also be directly visualized in full. Both of these possibilities will be used in section 3 for reporting the results of all the experiments.

### 2.3 Time series relevance determination using GTM-TT

For a time series clustering solution to be considered useful in practical applications, it has to be reasonably interpretable, and this interpretability would improve if clusters could be described using only the time series that are most relevant for the definition of the cluster structure. Therefore, the development of an unsupervised method for TSRD should sensibly increase the GTM-TT model interpretability and, as a result, its usefulness.

Recently, a method for feature selection in unsupervised model-based clustering with mixture models was presented in (Law, Figueiredo & Jain, 2004) and extended to the GTM for static data in (Vellido, 2005; Vellido, Lisboa & Vicente, 2006). This method calculates an unsupervised feature saliency as part of the EM algorithm. Such saliency measures the relevance of a feature on the definition of the cluster structure defined by the model. Here, we extend the method to TSRD for the GTM-TT.

Formally, the saliency of a time series  $d$  is defined as  $\rho_d = P(\eta_d = 1)$ , where  $\eta = \{\eta_1, \dots, \eta_D\}$  is a set of binary indicators that can be integrated in the GTM-TT optimization algorithm as missing (latent) labels. A value of  $\eta_d = 1$  would indicate the full relevance of time series  $d$ . According to this definition, the distribution of  $\mathbf{x}_n$  can be written as

$$p(\mathbf{x}_n | \mathbf{q}_n = \mathbf{u}_i, \mathbf{W}, \beta, \mathbf{w}_o, \beta_o, \boldsymbol{\rho}) = \prod_{d=1}^D \{ \rho_d p(x_d | \mathbf{q}_n = \mathbf{u}_i, \mathbf{w}_d, \beta) + (1 - \rho_d) q(x_d | \mathbf{u}_o, w_{o,d}, \beta_{o,d}) \} \quad [15]$$

where  $\mathbf{w}_d$  is the vector of  $\mathbf{W}$  corresponding to time series  $d$  and  $\boldsymbol{\rho} = \{\rho_1, \dots, \rho_D\}$ . The relevance of a given time series  $d$  is defined by  $\rho_d$ ; correspondingly, a time series  $d$  is irrelevant if it follows a density  $q(x_d | w_{o,d}, \beta_{o,d})$ , common to all the states-components of the mixture, which should reflect any prior knowledge we might have regarding irrelevant features, or otherwise take the form of a general, uninformative distribution. The common component requires the definition of two extra adaptive parameters  $\mathbf{w}_o = \{w_{o,1}, \dots, w_{o,D}\}$  and  $\beta_o = \{\beta_{o,1}, \dots, \beta_{o,D}\}$ . Notice that Eq. 15 entails the assumption that time series are conditionally independent, which is equivalent to the assumption of a diagonal covariance matrix. Given the common variance constraint of the GTM-TT, the model complies by definition with such assumption.

The likelihood for GTM-TT (Eq. 8) is redefined as follows:

$$p(X, \lambda) = \sum_{\text{all } Q} \pi_{\mathbf{q}_1} \prod_{n=1}^{N-1} P_{\mathbf{q}_n \mathbf{q}_{n+1}} \prod_{n=1}^N \prod_{d=1}^D (a_{i,n,d} + b_{n,d}) \quad [16]$$

where

$$a_{i,n,d} = \rho_d \left( \frac{\beta}{2\pi} \right)^{1/2} \exp \left[ -\frac{\beta}{2} \left( x_{n,d} - \sum_m \phi_m(\mathbf{u}_i) w_{m,d} \right)^2 \right] \quad [17]$$

and

$$b_{n,d} = (1 - \rho_d) \left( \frac{\beta_{o,d}}{2\pi} \right)^{1/2} \exp \left[ -\frac{\beta_{o,d}}{2} (x_{n,d} - \phi_o(\mathbf{u}_o) \mathbf{w}_o)^2 \right] \quad [18]$$

The likelihood defined above can be efficiently calculated as in Eqs. 10 and 11. The parameters  $\pi_i$  and  $P_{ij}$  can also be calculated as for the standard GTM-TT (Eqs. 12 and 13). The maximization of the expected likelihood yields the following update formulae for the model

parameters:

$$\rho_d^{new} = \frac{1}{N} \sum_{i,n} R_{in} u_{i,n,d} \quad [19]$$

where

$$u_{i,n,d} = \frac{a_{i,n,d}}{a_{i,n,d} + b_{n,d}}; \quad [20]$$

$$\beta^{new} = \frac{\sum_{i,n} R_{in} \sum_d u_{i,n,d}}{\sum_{i,n} R_{in} \sum_d u_{i,n,d} \left( \sum_m \phi_m(\mathbf{u}_i) w_{md} - x_{nd} \right)^2} \quad [21]$$

$$\beta_{o,d}^{new} = \frac{\sum_{i,n} R_{in} v_{i,n,d}}{\sum_{i,n} R_{in} v_{i,n,d} \left( \phi_o(\mathbf{u}_o) w_{o,d} - x_{nd} \right)^2}, \quad [22]$$

where

$$v_{i,n,d} = \frac{b_{n,d}}{a_{i,n,d} + b_{n,d}}. \quad [23]$$

The elements of matrix  $\mathbf{W}^{new}$ , for each time series  $d$ , are obtained as the solution of the following system of equations:

$$\Phi^T \mathbf{G}^* \Phi \mathbf{W}_d^{new} - \Phi^T \mathbf{R}^* \mathbf{X}_d = 0 \quad [24]$$

where  $\mathbf{R}^*$  has elements  $R_{in}^* = u_{i,n,d} R_{in}$  for a given time series  $d^*$  with  $R_{in}$  given by Eq. 9, and  $\mathbf{G}^*$

has elements  $g_{ii'}^* = \begin{cases} \sum_{n=1}^N R_{in}^* & i = i' \\ 0 & i \neq i' \end{cases}$ . Similarly, we obtain  $\mathbf{w}_o^{new}$ , for each time series, as the

solution of:

$$\phi_o^T \mathbf{g}^* \phi_o \mathbf{w}_{o,d}^{new} - \phi_o^T \mathbf{r}^* \mathbf{X}_d = 0 \quad [25]$$

where  $\mathbf{r}^*$  has elements  $r_n^* = \sum_i R_{in}^* = \sum_i v_{i,n,d} R_{in}$  for a given time series  $d^*$ , and  $\mathbf{g}^* = \sum_{i,n} R_{in}^*$ .

### 3. Experiments

Two sets of experiments were devised; the first concerned the assessment of the suitability of the GTM-TT model for the exploratory analysis of multivariate time series through clustering and visualization. The second concerned the evaluation of the unsupervised TSRD method defined in section 2.3.

#### 3.1 Assessment of GTM-TT as a time series clustering and visualization tool

Several experiments were devised in order to assess the suitability of the GTM-TT model for the analysis, assisted by visualization, of multivariate time series. These experiments were organized according to three different main objectives: First, we aimed to compare the different results yielded by the standard GTM and the GTM-TT when clustering multivariate time series. This way, the advantages of using the latter model would be highlighted. Secondly, we aimed to illustrate how sudden transitions (also referred to as change points (Yamanishi & Takeuchi, 2002)) and, correspondingly, low-variability periods are reflected on the GTM-TT latent space of the states. Finally, the third objective was to illustrate, using the evolution of the data responsibilities over time, how regimes and their transitions are reflected in the latent space of states.

##### 3.1.1 Experimental data sets

Three publicly available real data sets and a fourth synthetically generated one were used for the experiments outlined in the previous paragraph. They are now summarily described:

(1) *Artificial\_data*: 3-variate time series consisting of 80 data points were artificially generated to

simulate different regimes and their transitions.

(2) *Shuttle\_data*: 6-variate time series consisting of 1000 data points obtained from various inertial sensors from Space Shuttle mission STS-57<sup>1</sup>. These time series are particularly appropriate for the planned experiments for they contain subsequences of little variability followed by sudden transition periods. This data set has previously been used for cluster detection in (Lin, Vlachos, Keogh & Gunopulos, 2004).

(3) *System\_data*: 9-variate time series consisting of 1908 samples that describe the operation of a workstation in a networking environment during one week. These data contain long periods of low activity with interspersed short bursts of high activity. They have been used in (Vesanto & Hollmén, 2001) for cluster detection and in (Tikka & Hollmén, 2004) for linear dependency detection. Due to their incompleteness, missing values were imputed prior to analysis through a variant of GTM whose performance was assessed in (Olier & Vellido, 2005), showing robust performance.

(4) *Physio\_data*: 3-variate time series consisting of 3400 samples of physiological data, used in the Santa Fe Competition<sup>2</sup> in 1991. They are non-stationary and consist of three physiological variables measured in a subject while sleeping. Importantly, these time series contain clearly atypical subsequences due to a measurement error (failure in a sensor). These data have also been used in (Strickert & Hammer, 2005) to assess the performance of MSOM, a variant of the standard SOM for time series.

---

<sup>1</sup> Available from: [www.cs.ucr.edu/~eamonn/](http://www.cs.ucr.edu/~eamonn/)

<sup>2</sup> Available from: [www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html](http://www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html)

### 3.1.2 Capturing the dynamics of time series through visualization: differences between GTM and GTM-TT

It has been argued that, in some cases, time series clustering is a meaningless endeavour (Keogh & Lin, 2005). It is generally agreed, though, that this might only apply to certain types of time series clustering methods: mostly those that resort to subsequence clustering. Nevertheless, it has recently been shown (Simon, Lee & Verleysen, 2006) that, for univariate time series and with an adequate pre-processing based on embedding techniques, SOM clustering of time series can indeed be meaningful. This must be understood in the sense that the distribution of two different time series over the SOM map (or, in the case of GTM-TT here, over the latent space of states) should be significantly different. Note that in this study we are dealing with multivariate time series but, equally, clustering would be meaningful only if the distribution of two multivariate time series over the latent space of states clearly differed.

Let us illustrate this with a comparison between the results yielded by the standard GTM and the GTM-TT, focusing on the illustration of the main differences in the visualization of the time series clusters. *Artificial\_data* and *System\_data* are used for these first experiments. The *Artificial\_data* series are displayed in Fig. 1, and their posterior-mode projection from Eq. 14 is represented in the cluster membership map of Fig. 2. In this map, the GTM latent states are represented by squares that are scaled according to the ratio of data points that the model understands as being generated by the corresponding mixture component.

The representation corresponding to the GTM-TT is characteristically more compact than the one for the standard GTM. This is due to the fact that periods of little data variability are assigned by GTM-TT to the same hidden state. On the contrary, the standard GTM lacks information related to the time sequence context of each point, which results in a much more dispersed representation, even for almost flat signal periods. As shown in Fig. 3, this is even more obvious for the *System\_data*, as they consist of a combination of idle periods and sudden outbursts.

GTM-TT consistently unifies all idle activity within a single state, leaving a few surrounding states to represent the narrow activity periods. Again, the standard GTM representation is much sparser. This effect is explained by the different meaning of the responsibility matrix in the standard GTM and in GTM-TT: each element of  $\mathbf{R}_{in} = p(\mathbf{u}_i | \mathbf{x}_n)$  represents, in the standard GTM, the probability of the hidden state  $\mathbf{u}_i$  given  $\mathbf{x}_n$ , independently of other data. Instead, in GTM-TT,  $R_{in} = P(\mathbf{q}_n = \mathbf{u}_i | X)$  defines the probability of being in the state  $\mathbf{u}_i$  at time step  $n$  of a data sequence (see Eq.9). Therefore,  $R_{in}$  in GTM-TT contains contextual information for each data point. As a result, even if  $\mathbf{x}_{n_1}$  and  $\mathbf{x}_{n_2}$  were identical data vectors, they might still have different posterior-mode projections in the latent space of states of GTM-TT, due to the inclusion of contextual information.

Let us go back at this point to the problem of clustering meaningfulness. Different series should be represented in the latent space by different distributions and trajectories. In order to illustrate this characteristic of GTM-TT in more detail, a subsequence of *Artificial\_data* (since  $n = 21$  until  $n = 40$ ) was attached at the end of the series with its data points randomly shuffled. As shown in Fig. 4, these individually identical data points are represented by rather different state trajectories in the GTM-TT membership map, indicating that the model is actually identifying the difference between series. On the contrary, the standard GTM would represent both data subsequences in exactly the same way.

### 3.1.3 Sudden transitions and little variability periods

According to the previously described characteristics of the GTM-TT, we might expect the model to facilitate the visualization of sudden transitions and of long periods of little data variability. In more detail, sudden data transitions might be expected to correspond to sudden jumps between usually distant, and possibly scarcely populated, map states. Instead, subsequences of little variability might

be expected to clump in few, possibly highly populated, map states. This effect can be clearly appreciated in Fig. 5, where, as in Fig. 3, the *System\_data* have been used. These data consist mostly of long idle signal periods, followed by sudden system activity bursts. On the left-hand side map, a brief period of sudden change (subsequence A) is represented as a succession of brief jumps between states covering a wide map area. On the right-hand side map, a mostly idle period (subsequence B) is mostly captured in a single state. In conclusion, the GTM-TT seems to capture these dynamics quite adequately, in an intuitive and interpretable way.

A further experiment was carried out using the *Shuttle\_data*, which are displayed in Fig. 6. Five non-overlapping periods (A to E) are considered; all but period B seem to contain little variability, although they are separated by sudden transitions. Once again, this is clearly reflected in the GTM-TT map of Fig. 7. Confirming the results of the previous experiment, the low variability subsequences bundle up in a few highly populated states, with neat transitions of quick state-to-state jumps. Period B, of intermediate variability, is, on the contrary, represented as a more gradual and less jittery evolution over states.

In order to make this acquired knowledge operative in the detection of sudden transitions or change points, the availability of a quantitative measure of sudden variation would be beneficial. Here we define one such measure by assuming that, as in biological learning (Friston, 2003), novel evidence steps up the learning rate. We might therefore expect sudden data transitions to be accompanied by sudden increases of the model likelihood. Consequently, the weighted mean of the emission probabilities can be used as a measure of the suddenness of transitions in multivariate time series, in the form:

$$p(\mathbf{x}_n) = \sum_{i=1}^M R_{in} p_{u_i}(\mathbf{x}_n) \quad [26]$$

However, logarithms of the probabilities might be more useful. This way, an easily interpretable *index of variability*, denoted  $IV_n$ , can be defined as:

$$IV_n = -\sum_{i=1}^M R_{in} \log[p_{u_i}(\mathbf{x}_n)] \quad [27]$$

Fig. 8 displays  $IV_n$  for the *Shuttle\_data* series. Little variability intervals A, C, D, and E, are shown to have a  $IV_n$  value close to zero. However, sudden transitions between these intervals show high relative values which are proportional to the intensity of the transitions. Interval B is represented by quasi-periodic changes of  $IV_n$ , as expected.

An alternative approach to sudden change detection in time series, based on the concept of wavelet footprints, was recently defined in (Sharifzadeh, Azmoodeh & Shahabi, 2005).

#### 3.1.4 Visualization of regimes and regime transitions

A regime in multivariate time series, in a loose sense, can be described as a subsequence with differentiated interpretation (examples of their analysis can be found, for instance, in Bishop, Hinton & Strachan, 1997; Kabán & Girolami, 2002). In this study, the *Physio\_data* set is used to illustrate how regimes and transitions between regimes are visualized through GTM-TT. First, some data preprocessing, following (Strickert & Hammer, 2005), was carried out to remove trends. Furthermore, the atypical subsequences described in the previous subsection were fully removed.

The top plot of Fig. 9 shows these data after pre-processing, where a possible transition between regimes (A) and a regime interval (B) are highlighted and considered for further analysis. The GTM-TT cluster membership maps corresponding to subsequences A and B are displayed in the bottom plots. As shown there, a regime concentrates in a well defined area of the membership map, whereas a transition between regimes is likely to involve states from past and future regimes in two distinct areas. This is explored in more detail in Figs. 10 and 11.

Fig. 10 shows the responsibilities at four consecutive time steps of the regime transition A. The evolution of the posterior distribution in the latent space of states is clearly observed, with a

gradual transference of responsibility from one area to another that includes intermediate multimodalities. In turn, Fig. 11 shows the responsibilities at four non-consecutive time steps of regime B. In this case, the evolution of the posterior distribution is concentrated in a single area of the membership map.

### 3.2 Time series relevance determination

A further set of experiments was designed to evaluate the proposed multivariate TSRD method defined in section 2.3. The first goal was the generation of time series relevance rankings that, in practical applications, might be used as a basis for time series selection. For that, and depending on the application requirements, different saliency thresholds might be set, so that time series with a saliency below the given threshold would be considered not to correspond to the underlying cluster structure of the series and, therefore, of not enough interest for cluster interpretation through exploratory visualization.

#### 3.2.1 Experimental data sets

Two data sets were used for the TSRD method evaluation experiments:

- (1) *Artificial\_data2*: This new artificial data set consists of three series obtained from a subset of *Shuttle\_data*, which include two brisk transitions between relatively stable periods (plotted in Fig. 12), to which three randomly generated series are added. The latter are less likely to have cluster structure (and, therefore, are less likely to be relevant) than the former.
- (2) *Shuttle\_data*: the same data set described in section 3.1.1 is also used for this second set of experiments. These 6-variate time series consist of 1000 data points and are once again deemed suitable because their combination of subsequences of little variability followed by sudden transition periods is likely to yield a clear cluster structure.

### 3.2.2 Experimental results and discussion

In a first experiment, we aimed to test whether the TSRD method for GTM-TT was capable of gauging relevance properly in *Artificial\_data2*. The model was run 20 times with different random parameter initializations. The resulting saliencies, obtained from Eq. 19, are plotted in Fig. 13.

As expected, the three time series including brisk transitions between relatively stable periods yielded the highest saliencies ( $\bar{p}_d > 0.8$ ) and they are, therefore, the most relevant or, in other words, those which the model considers to be generating most of the data cluster structure. On the contrary, the three extra series without clear clustering pattern yielded very low saliencies in the area of 0.1, as expected. It is also interesting to note that the smaller bars for the three time series including brisk transitions suggest that the model is more certain of the relevance of these series. These results are reinforced by the GTM-TT maps in Fig. 14. The map on the left, corresponding to the standard GTM-TT, suggests the presence of three groups of states, which are of course generated by the three relevant time series with the two brisk transitions, although two of these groups are not clearly differentiated. We would expect that the application of the TSRD method generated a sharper separation between the three groups of states. This is the case, as evidenced by the map on the right hand side plot of Fig. 14, corresponding to the GTM-TT with TSRD, which shows not only wider gaps between the three main groups of states than the map on the left hand side, but also a higher concentration of time points in only a few individual states. Exploring such effect was the second goal of these experiments, and the results can be easily explained: The TSRD method is not only providing a ranking of the time series, but it is also actively limiting the effect of the less relevant series during the data fitting process, while enhancing the effect of the most relevant ones. As a result, the final cluster structure mainly reflects what enhances it most.

Another important effect of the application of TSRD to the clustering of time series using GTM-TT can be observed in the calculation of the *index of variability* ( $IV_n$ ) introduced in

subsection 3.1.3. Irrelevant time series that generate little cluster structure could mask the existing change points and make their detection difficult. The use of TSRD alleviates this potential problem by minimizing their negative effect. This is illustrated by the results shown in Fig. 15 for *Artificial\_data2*: the standard GTM-TT does not reveal clearly defined change points, whereas GTM-TT with TSRD clearly identifies two periods of rapid change that correspond precisely to the transitions shown in Fig. 12.

These relevance determination experiments are now repeated with a real data set, namely *Shuttle\_data*. The time series saliency results are reported in Fig. 16. In this case, the differences in saliency between the six available series are much smaller, and all series are attributed high relevance ( $\bar{\rho}_d > 0.85$ ) in defining the cluster structure of the data. Interestingly though, the series with highest saliency, all with  $\bar{\rho}_d > 0.95$ , namely 1, 4 and 5, correspond to those with the neatest and somehow simplest combinations of quasi-stationary periods and broad magnitude shifts, as seen in Fig. 17.

#### 4. Conclusions

Although visual exploration is an important stage in data mining processes, little research has been devoted to unsupervised methods for the visual exploration of multivariate time series. In this paper, the capabilities of the GTM-TT, a stochastic latent model of the manifold learning family that performs simultaneous time series data clustering and visualization, have been assessed through experiments concerning visualizations of sudden transitions and low variability periods, and of regimes and regime transitions. The GTM-TT has been shown to produce more faithful visual representations of multivariate time series than the standard GTM model. It has also been shown that GTM-TT facilitates the visualization of sudden transitions and long periods of little data variability in a multivariate sense. Furthermore, a novel index for the detection of sudden transitions

or change points has been defined and tested with successful results.

For real applications of time series analysis in areas such as, for instance, natural sciences, medicine, and industry, we would like the clustering results and their visualization to be both interpretable and actionable. Unfortunately, the interpretation of the GTM-TT clustering results through exploratory visualization might become difficult for data sets consisting of a large number of time series, and the data analyst would benefit from a time series relevance ranking method and from a selection method based on it. In this paper, we have defined one such TSRD method that is an integral part of the GTM-TT model data fitting process. It has been tested using several artificial and real data sets and the results show not only that it can properly gauge the relative relevance of individual time series, but also that it can produce neater clustering results by effectively minimizing the negative impact of the least relevant series on the clustering process itself. Furthermore, the TSRD method has been shown to improve the efficiency of the detection of change points through the index of variability also defined in this paper.

Future research on the application of GTM-TT to the exploratory analysis of multivariate time series will be devoted to the visual identification and characterization of atypical data subsequences. This research should find a straightforward application in anomaly detection tasks in the form of visually intuitive anomaly early warning methods.

**References**

- Baum, L.E. & Egon, J.A. (1967). An inequality with applications to statistical estimation for probabilistic functions for a Markov process and to a model for ecology. *Bulletin of the American Mathematical Society*, **73**, 360-363.
- Bengio, Y. & Chapados, N. (2003). Extensions to metric-based model selection. *Journal of Machine Learning Research*, **3**, 1209-1227.
- Bishop, C., Hinton, G. & Strachan, I. (1997). GTM through time. In proceedings of *The IEE Fifth International Conference on Artificial Neural Networks*, Cambridge, U.K., 111-116.
- Bishop, C., Svensén, M. & Williams C. (1998a). GTM: The Generative Topographic Mapping. *Neural Computation*, **10** (1), 215-234.
- Bishop, C., Svensén, M. & Williams C. (1998b). Developments of the Generative Topographic Mapping. *Neurocomputing*, **21** (1), 203-224.
- Bullen, R.J., Cornford, D. & Nabney, I.T., (2003). Outlier detection in scatterometer data: neural network approaches. *Neural Networks*, **16** (3-4), 419-426.
- Carreira-Perpiñan, M.A. (2000). Reconstruction of sequential data with probabilistic models and continuity constraints. In S. Solla, T. Leen & K.R. Muller (Eds.), *Advances in Neural Information Processing Systems 12*, NIPS'1999, (pp. 414-420).
- Chappell G. & Taylor, J. (1993). The temporal Kohonen map. *Neural Networks*, **6**, 441-445.
- Chatfield, C. (2000). *Time Series Forecasting*. Chapman & Hall/CRC Press.
- Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, **16**, 1325-1352.
- Girolami, M. (2002). Latent variable models for the topographic organisation of discrete and strictly positive data. *Neurocomputing*, **48**, 185-198.
- Grabmeier, J. & Rudolph, A. (2002). Techniques of cluster algorithms in data mining. *Data Mining*

and *Knowledge Discovery*, **6**, 303-360.

Kabán, A. & Girolami, M. (2002). A dynamic probabilistic model to visualise topic evolution in text streams. *Journal of Intelligence Information Systems*, **18** (2-3), 107-125.

Keogh, E. & Lin, J. (2005). Clustering of time series subsequences is meaningless: Implications for previous and future research. *Knowledge and Information Systems*, **8**(2), 154-177.

Kohonen, T. (2001). *Self-Organizing Maps* (3<sup>rd</sup> edition). Berlin: Springer-Verlag.

Kostiainen, T. & Lampinen, J. (2002). On the generative probability density model in the self-organizing map. *Neurocomputing*, **48** (1-4), 217-228.

Law, M.H.C., Figueiredo, M.A.T. & Jain A.K. (2004). Simultaneous feature selection and clustering using mixture models. *IEEE Transactions on Pattern Analysis*, **26** (9), 1154-1166.

Lin, J., Vlachos, M. Keogh, E. & Gunopulos, D. (2004). Iterative incremental clustering of time series. In E. Bertino et al. (Eds.), *Advances in Database Technology – EDBT 2004 Proceedings*. LNCS **2992**, (pp.106-122).

Olier, I. & Vellido, A. (2005). Comparative assessment of the robustness of missing data imputation through Generative Topographic Mapping. In J. Cabestany, A. Prieto & F. Sandoval (Eds.), *Computational Intelligence and Bioinspired Systems, IWANN 2005 Proceedings*. LNCS **3512**, 787-794.

Sharifzadeh, M., Azmoodeh, F. & Shahabi, C. (2005). Change detection in time series using wavelet footprints. In C.B. Medeiros, M. Egenhofer & E. Bertino (Eds.), *Advances in Spatial and Temporal Databases, SSTD 2005 Proceedings*. LNCS **3633**, 127-144.

Simon, G., Lee, J.A. & Verleysen, M. (2006). Unfolding preprocessing for meaningful time series clustering. *Neural Networks*, **19** (6-7), 877-888.

Strickert, M. & Hammer, B. (2005). Merge SOM for temporal data. *Neurocomputing*, **64**, 39-71.

- Tikka, J. & Hollmén, J. (2004). Learning linear dependency trees from multivariate time-series data. In proceedings of the IEEE International Conference on Data Mining (ICDM 2004) Workshop on Temporal Data Mining: Algorithms, Theory and Applications. Brighton, UK.
- Tiño, P. & Nabney, I. (2002). Hierarchical GTM: constructing localized nonlinear projection manifolds in a principled way. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24** (5), 639-656.
- Tiño, P., Farkaš, I. & van Mourik, J. (2006). Dynamics and Topographic Organization of Recursive Self-Organizing Maps. *Neural Computation*, **18**, 2529-2567.
- Vellido, A. (2005). Preliminary theoretical results on a feature relevance determination method for Generative Topographic Mapping, Technical Report LSI-05-13-R, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.
- Vellido, A. (2006). Missing data imputation through GTM as a mixture of  $t$ -distributions. *Neural Networks*, In press.
- Vellido, A., El-Deredy, W. & Lisboa, P.J.G. (2003). Selective smoothing of the Generative Topographic Mapping. *IEEE Transactions on Neural Networks*, **14**, 847-852.
- Vellido, A., Lisboa, P.J.G. & Vicente, D. (2006). Robust analysis of MRS brain tumour data using  $t$ -GTM. *Neurocomputing*, **69** (7-9), 754-768.
- Vesanto, J. (1999). SOM-based data visualization methods. *Intelligent Data Analysis*, **3** (2), 111-126.
- Vesanto, J. & Hollmén, J. (2001). An automated report generation tool for the data understanding phase. In A. Abraham & M. Koeppen (Eds.) *Hybrid Information Systems. Proceedings of HIS'01* (pp.611-625).
- Voegtlin, T. (2002). Recursive self-organizing maps. *Neural Networks*, **15**(8-9), 979-991.

Wong, P.C. (1999). Visual data mining, *IEEE Transactions on Computer Graphics Applications*, **19** (5), 20-21.

Yamanishi, K. & Takeuchi, J.-I. (2002). A unifying framework for detecting outliers and change points from non-stationary time series data. In proceedings of the 8<sup>th</sup> *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, 676-681.

Yin, H. & Allinson, N. (2001). Self-organizing mixture networks for probability density estimation. *IEEE Transaction on Neural Networks*, **12**, 405-411.

Yoon, H., Yang, K. & Shahabi, C. (2005). Feature subset selection and feature ranking for multivariate time series. *IEEE Transactions on Knowledge and Data Engineering*, **17** (9), 1186-1198.

Zhang, G., Patuwo, B. & Hu, M. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal on Forecasting*, **14** (1), 35-62.

**Figure legends**

**Fig. 1:** *Artificial\_data* time series.

**Fig. 2:** Cluster membership maps of *Artificial\_data* using (a) standard GTM and (b) GTM-TT. An  $8 \times 8$  square latent grid was used to train both models.

**Fig. 3:** Cluster membership maps of *System\_data* using (a) standard GTM and (b) GTM-TT. A  $10 \times 10$  squared grid was used to train both models.

**Fig. 4:** (Left): A GTM-TT cluster membership map representing the subsequence of *Artificial\_data* from  $n=21$  to  $n=40$ . (Right): The same subsequence, which was attached at the end of the series with all data points randomly shuffled, yields a different GTM-TT membership map displaying a different trajectory (represented by the lines between states).

**Fig. 5:** The top plot represents one of the series of *System\_data*, which measures idle CPU time of the workstation. On the bottom plot, the GTM-TT cluster membership map on the left-hand side corresponds to a short data subsequence of the complete *System\_data*, from  $n=400$  to sample  $n=440$ , with sudden transitions associated to bursts of system activity (subsequence A). The map on the right-hand side corresponds to a longer idle period of the system, from  $n=1000$  to  $n=1150$  (subsequence B). Note that the size of the squares is relative to the number of data points represented and, therefore, state-squares of the same size in the left and right-hand side plots do not indicate equal number of data points assigned to them; the distribution of the data over the states is nevertheless the same.

**Fig. 6:** Four periods of relatively little variability in *Shuttle\_data*, separated by sudden transitions, are singled out as A, C, D and E. A fifth period B of higher variability, also delimited by sudden transitions, spans from the end of A to the beginning of C.

**Fig. 7:** The subsequences A to E of the *Shuttle\_data* series are visualized in the GTM-TT membership map. The latent space consisted of a squared  $10 \times 10$  grid of states. The latent states representing the little variability periods are encircled, and sudden transitory intervals are represented by dotted arrows. The state transitions of period B are represented by continuous arrows.

**Fig. 8:** Index of variability ( $IV_n$ ) results for *Shuttle\_data*.

**Fig. 9:** The processed *Physio\_data* time series are displayed on the top plot. A possible regime transition A and a possible regime B are highlighted. The GTM-TT membership maps are displayed on the bottom plots (A: left) corresponding to the regime transition A, between time step  $n=700$  and time step  $n=703$ ; and (B: right) corresponding to regime B at four non-consecutive time steps  $\{n=2360; n=2530; n=2620; n=2740\}$ . In both cases, the latent space of states is a  $10 \times 10$  grid.

**Fig. 10:** Evolution of the posterior distributions across transition A, since  $n=700$  (a) until  $n=703$  (d). A gradual transition of responsibility between two regimes in distinct map regions can be observed.

**Fig. 11:** Evolution of the posterior distributions across regime B, at four non-consecutive time steps  $\{n=2360; n=2530; n=2620; n=2740\}$ . They move across a narrowly delimited region with no trace of multimodality.

**Fig. 12:** The three series from *Artificial\_data2* with highest expected relevance.

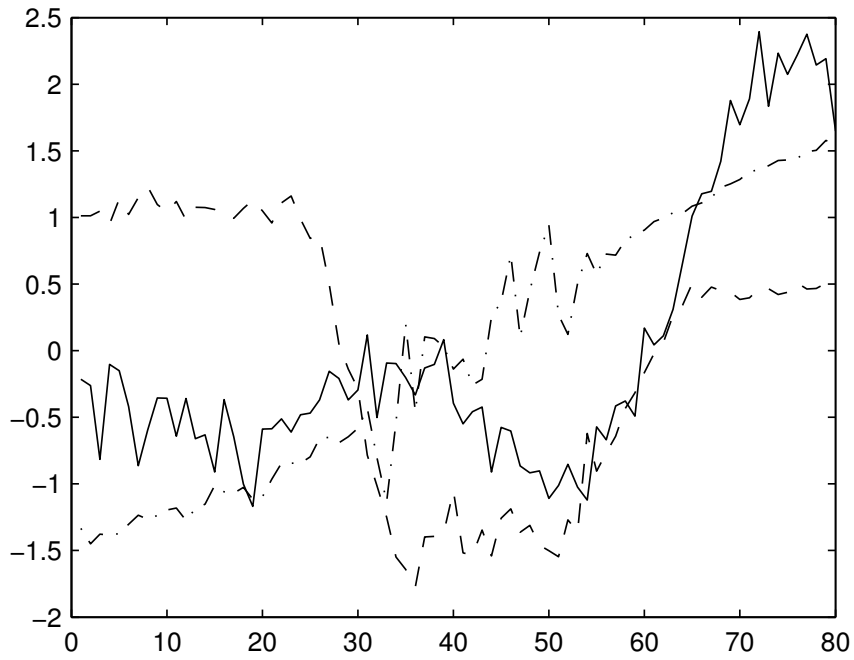
**Fig. 13:** Estimated values (represented by their means, over 20 runs of the algorithm, plus and minus one standard deviation) of the saliencies  $\rho$  for all features from *Artificial\_data2*, using the random varying initialization strategy described in the main text.

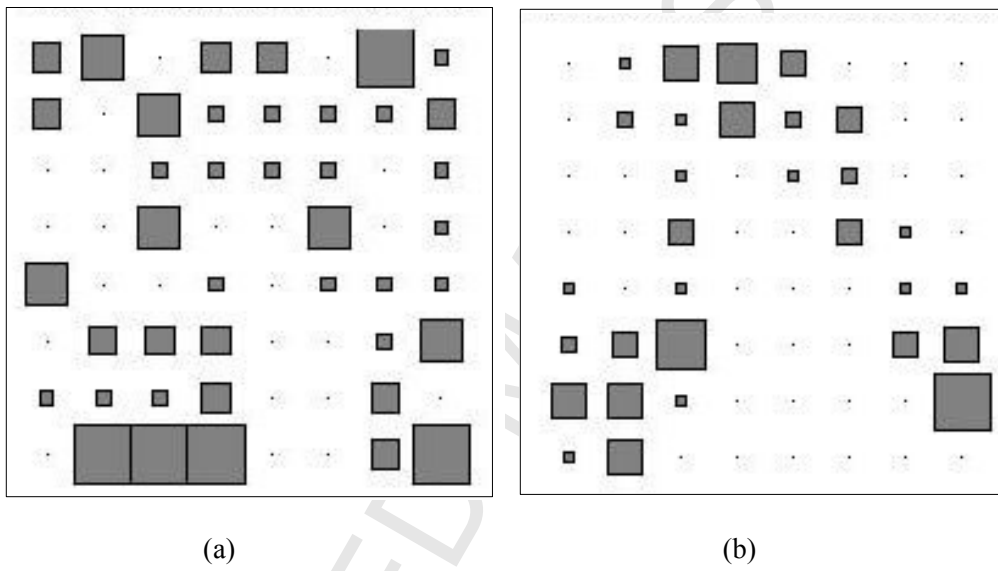
**Fig. 14:** Cluster membership maps for the *Artificial\_data2*. (Left) map corresponding to the standard GTM-TT described in section 2.2. (Right) map corresponding to the GTM-TT with time series relevance determination, described in section 2.3. The latent space consisted of a squared  $6 \times 6$  grid of states.

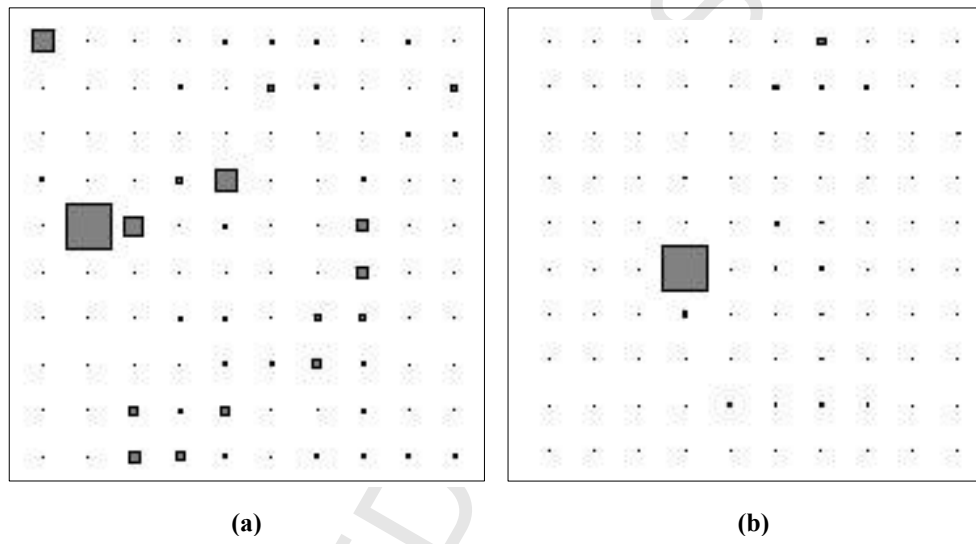
**Fig. 15:**  $IV_n$  for *Artificial\_data2* using the standard GTM-TT (Top) and the GTM-TT with TSRD (Bottom). As expected, the detection of change points is far more efficient when TSRD is used.

**Fig. 16:** Estimated values (represented by their means, over 20 runs of the algorithm, plus and minus one standard deviation) of the saliencies  $\rho$  for all features from *Shuttle\_data*, using the random varying initialization strategy described in the main text.

**Fig. 17:** Individual time series of *Shuttle\_data* (1<sup>st</sup> to 6<sup>th</sup> from top to bottom and left to right)

**Figure 1**

**Figure 2**

**Figure 3**

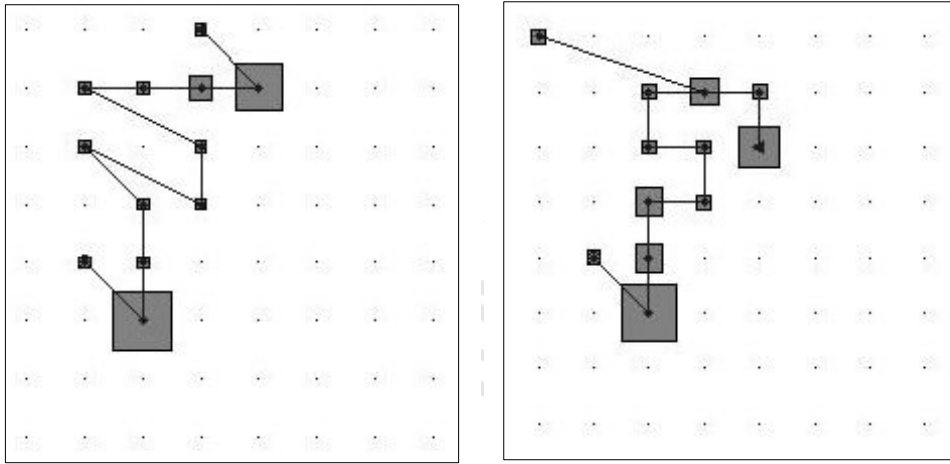
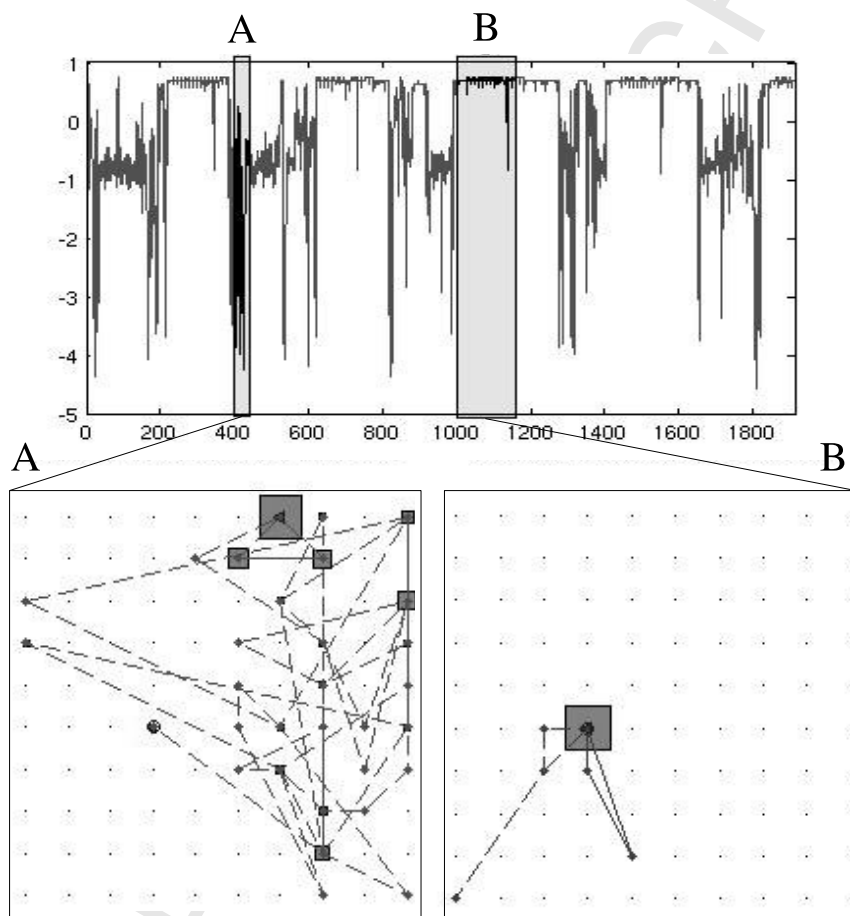
**Figure 4**

Figure 5



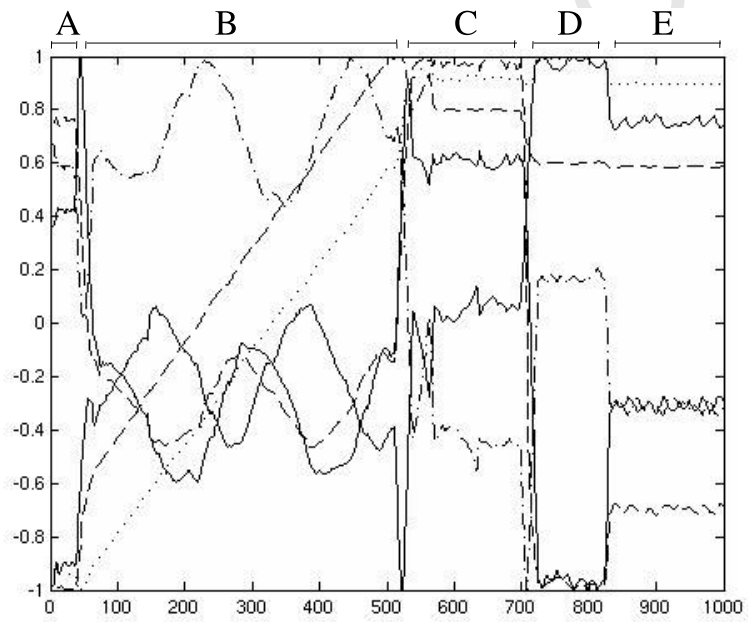
**Figure 6**

Figure 7

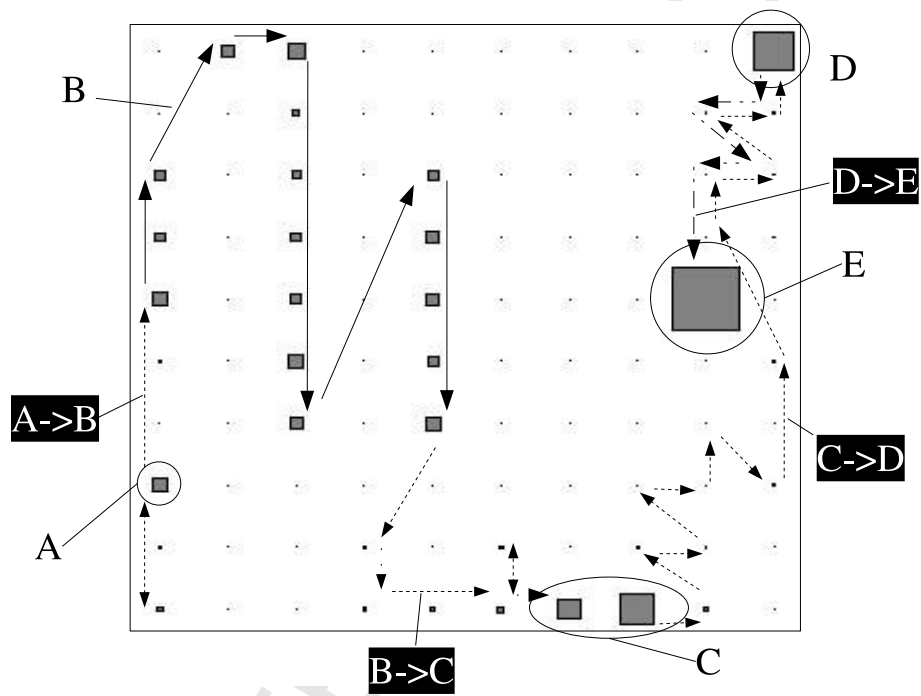


Figure 8

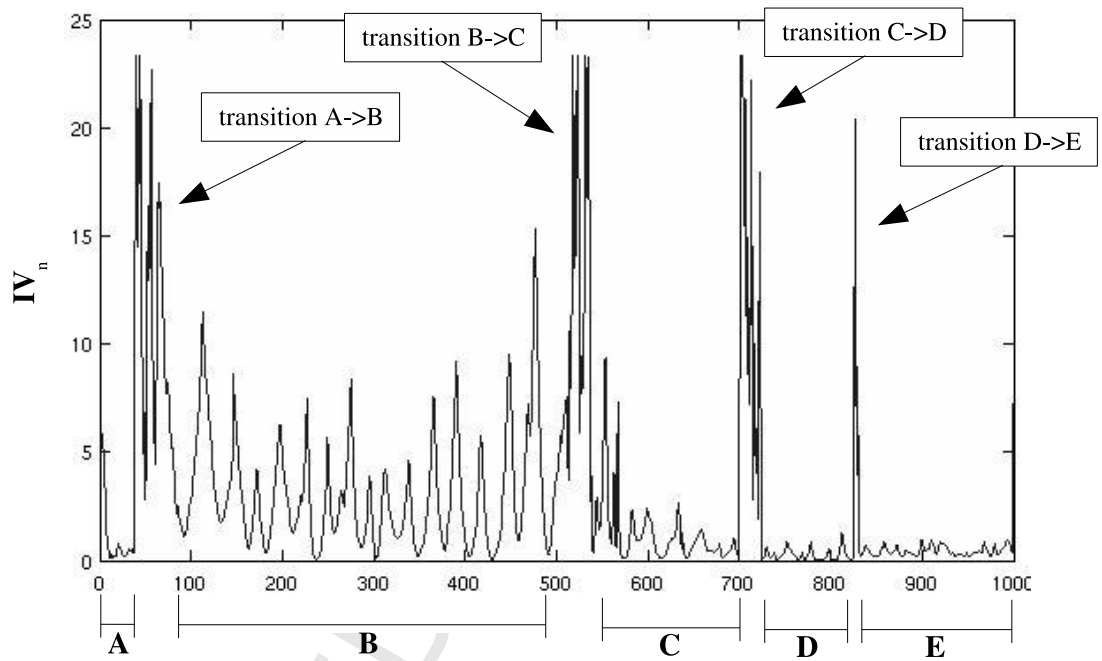
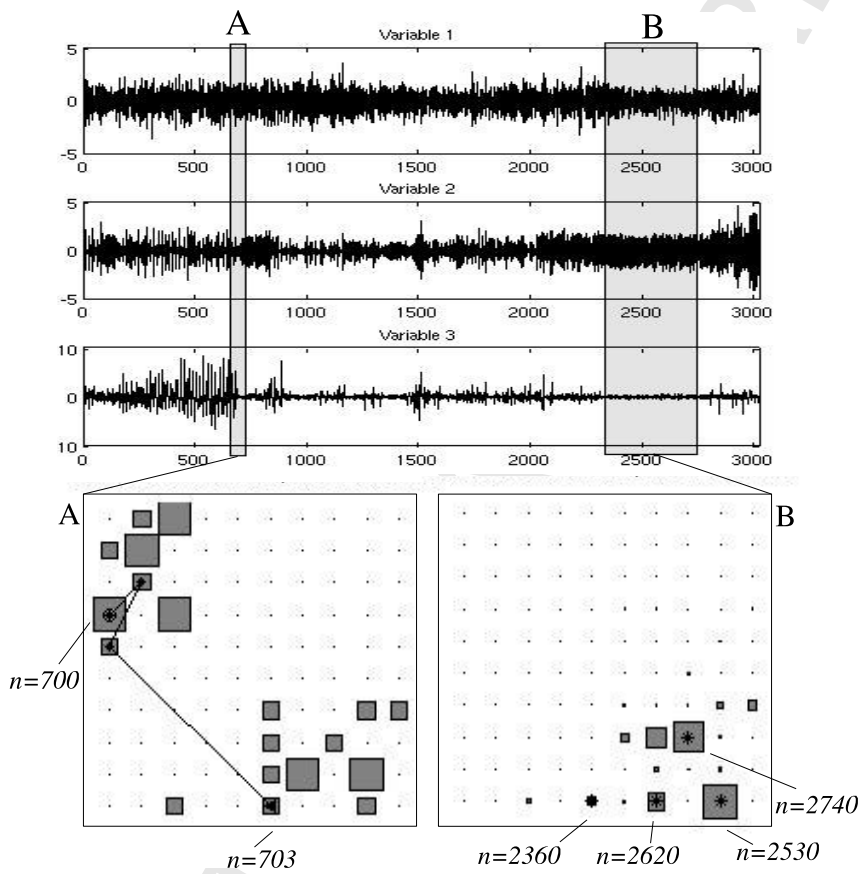
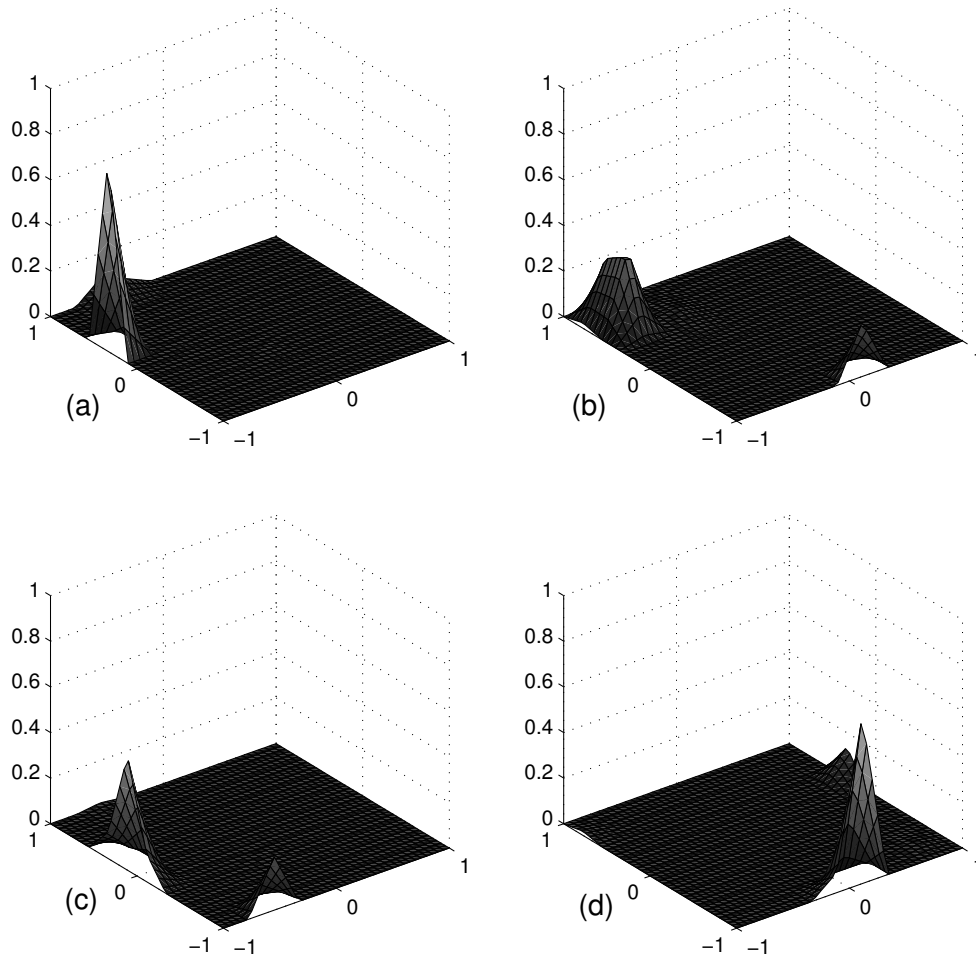
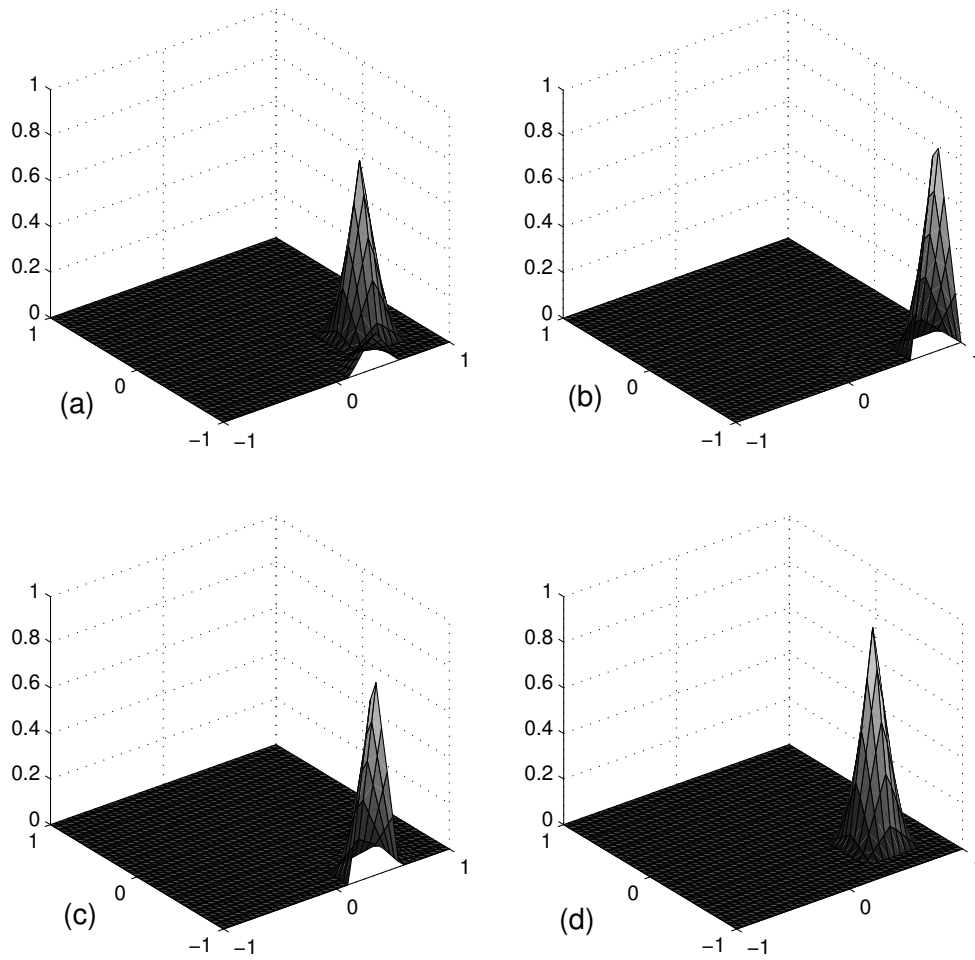
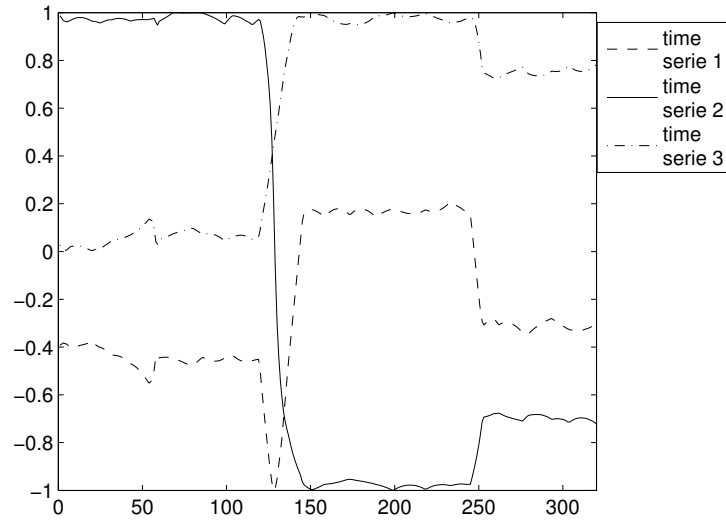


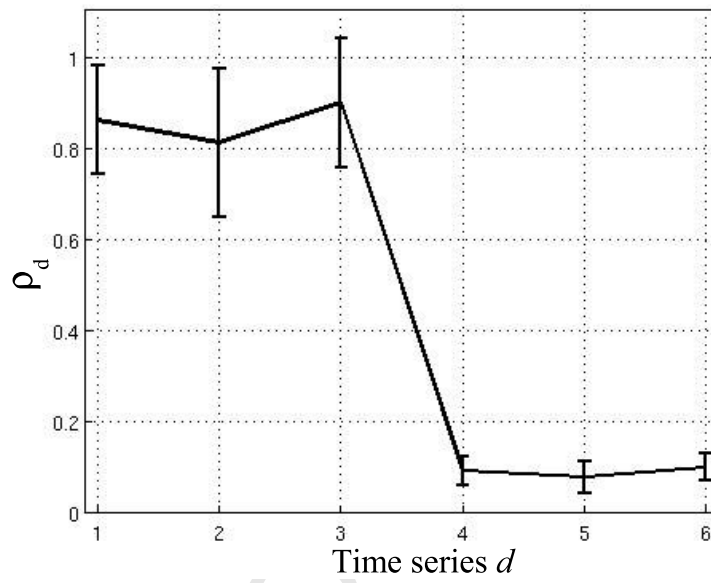
Figure 9

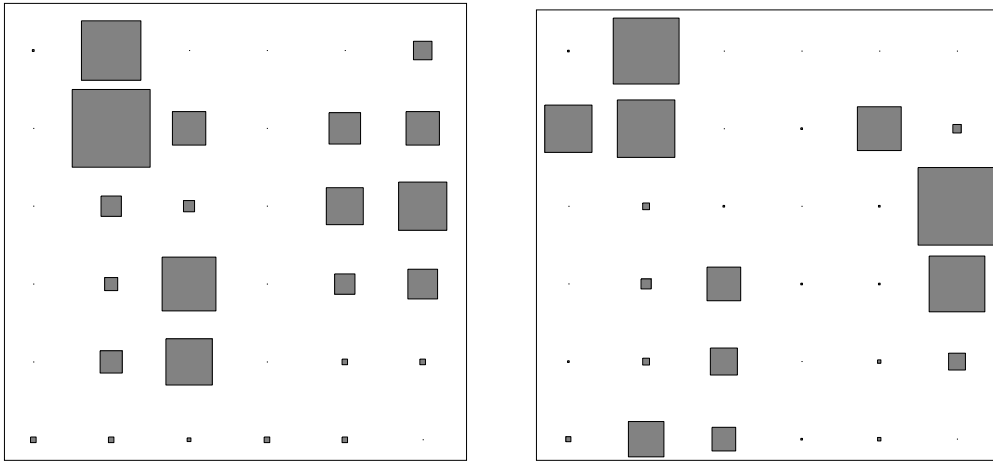


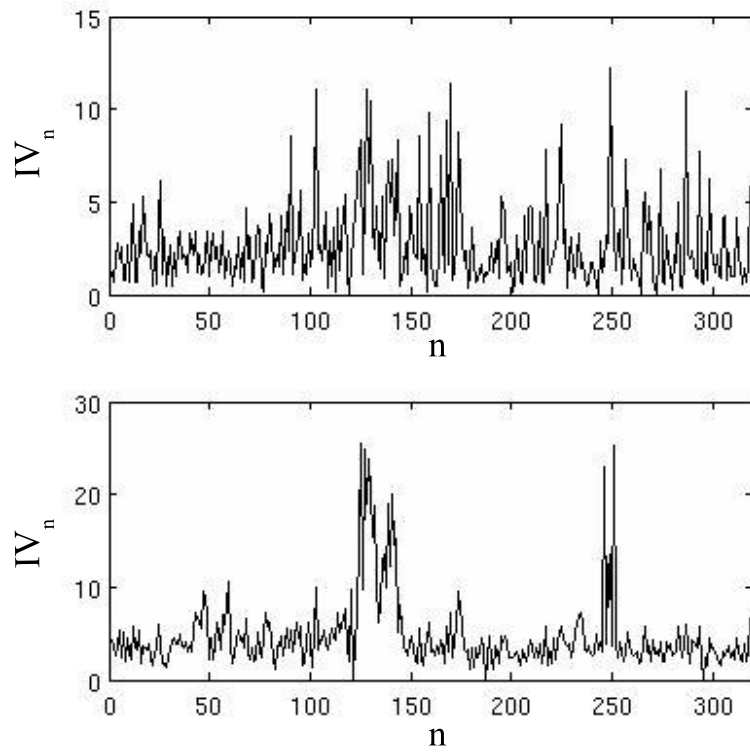
**Figure 10**

**Figure 11**

**Figure 12**

**Figure 13**

**Figure 14**

**Figure 15**

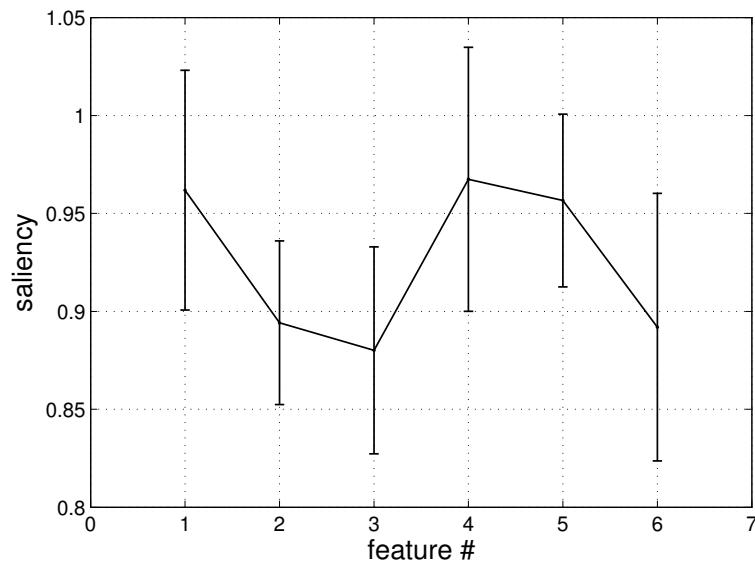
**Figure 16**

Figure 17

