Abstract—Most of the existing research on time series concerns supervised forecasting problems. In comparison, little research has been devoted to unsupervised methods for the visual exploration of multivariate time series. In this paper, the capabilities of the Generative Topographic Mapping Through Time, a model with solid foundations in probability theory that performs simultaneous time series data clustering and visualization, are assessed in detail in several experiments. The focus is placed on the detection of atypical data, the visualization of the evolution of signal regimes, and the exploration of sudden transitions, for which a novel identification index is defined.

Index Terms—Generative Topographic Mapping; Topology-constrained hidden Markov models; Multivariate time series analysis; Data visualization; Clustering

I. INTRODUCTION

Multivariate time series analysis has long ago become an established research area. 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 [1]. All 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 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; as a result, 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, 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 interesting time series clustering results have been obtained with Kohonen’s Self-Organizing Map (SOM: [2]) neural networks even without accounting for the violation of the i.i.d. condition. Nevertheless, some extensions of SOM have been developed to explicitly accommodate time series, mostly through recurrent connectivity (Chappell and Taylor, [3]; Strickert and Hammer, [4]). Despite attempts to fit this model into a probabilistic framework (e.g., Yin and Allinson, [5]; Kostiainen and Lampinen, [6]), it has retained its heuristic definition, which is at the origin of some of its limitations, such as the lack of a proper error function to optimize, and the lack of adaptive optimization of the model parameters. On the contrary, the Generative Topographic Mapping (GTM: Bishop et al., [7]) is a stochastic model 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 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, [8]).

The GTM Through Time (henceforth referred to as GTM-TT: Bishop et al, [9]) is one of the many possible extensions of the standard GTM allowed by its probabilistic definition. It was defined to deal with time series, but its capabilities for exploratory analysis through visualization have never been assessed in detail. In this brief paper we 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.

The rest of the paper is structured as follows: First, in section 2, an introduction to the GTM as a constrained mixture of Gaussians is provided. This is followed, in section 3, by a description of the GTM-TT. Several experiments for the assessment of the GTM-TT performance are described, and their results presented and discussed, in section 4. The paper wraps up with a brief conclusion section.

II. THE STANDARD GTM 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 \(\mathbb{R}^L\) (with \(L\) being usually 1 or 2 for visualization purposes) onto a manifold embedded in the \(\mathbb{R}^D\) space, where the observed data reside. The mapping that generates the manifold takes the form:

\[
y = W\Phi(u)
\]

(1)

where \(y \in \mathbb{R}^D\), \(u \in \mathbb{R}^L\), \(W\) is the matrix that generates the mapping, and \(\Phi\) is a set of \(S\) basis functions \(\phi_s\) (radially symmetric Gaussians in the standard model). To achieve computational tractability, the prior distribution of \(u\) 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:

\[ p(u) = \frac{1}{M} \sum_{i=1}^{M} \delta(u - u_i) \]

(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 \( x_n \) are independent, identically distributed (i.i.d.), leads to the definition of a complete log-likelihood in the form:

\[ l(W, \beta) = \sum_{n=1}^{N} \ln p(x_n | W, \beta) \]

= \sum_{n=1}^{N} \left( \frac{1}{M} \sum_{i=1}^{M} p(x_n | u_i, W, \beta) \right) \]

(3)

From Eq. 3, the adaptive parameters of the model, which are \( W \) and the common inverse variance of the Gaussian components, \( \beta \), can be optimized using the EM algorithm. Details are provided in [7].

III. GENERATIVE TOPOGRAPHIC MAPPING FOR TIME SERIES: THE GTM-TT

A. GTM for time series: The GTM-TT

Multivariate time series are not i.i.d. data and, therefore, the standard definition of the GTM in section II can only provide a rough approximation to their proper modelling. A variation on the standard model, called the GTM Through Time, was defined in Bishop et al. [9] as a topology-constrained hidden Markov model to deal explicitly with time series.

In GTM-TT, points in latent space are considered as hidden states and temporal dependencies are captured through their coupling. Furthermore, the emission probabilities are controlled by the GTM mixture distribution. The joint probability distribution of the multivariate time data \( X \), and hidden states \( U = \{u_1, u_2, \ldots, u_n, \ldots, u_N\} \) takes the form:

\[ p(U, X) = \pi_i \prod_{n=2}^{N} p_{i_{n-1}i_n} \prod_{n=1}^{N} p(x_n | u_n) \]

(4)

where \( \pi_i \) defines the initial state probability of \( U \); \( p_{i_{n-1}i_n} = p(u_i | u_{i_{n-1}}) \) is the probability of transition from one hidden state to another (and therefore captures the temporal dependencies); and \( p(x_n | u_n) \) is the probability found on the second line of Eq. 3. This leads to the definition of a likelihood for the GTM-TT model:

\[ L = p(X) = \sum_{U} p(U, X) \]

(5)

which can be efficiently calculated using the forward-backward procedure [10]. The probability of being in the state \( i \) at time \( n \), given the data and the model, or responsibility \( R_{in} \), is calculated as:

\[ p(u_i | X) = R_{in} = \frac{\alpha_n(i) \beta_n(i)}{L} \]

(6)

The forward variable \( \alpha_n(i) \) is the joint probability of the past subsequence \( x_1, x_2, \ldots, x_n \) and the state \( i_n \), given by the following recursive equation:

\[ \alpha_n(i) = \left( \sum_{k=1}^{M} \alpha_{n-1}(k) p_{ki} \right) p(x_n | u_i) \]

(7)

where \( \alpha_1(i) = \pi_i p(x_1 | u_i) \). The backward variable \( \beta_n(i) \), which is the probability of the future subsequence \( x_{n+1}, x_{n+2}, \ldots, x_N \) given hidden state \( i_n \), is given by the following recursive equation:

\[ \beta_n(i) = \sum_{k=1}^{M} p_k p(x_{n+1} | u_{n+1}) \beta_{n+1}(k) \]

(8)

where \( \beta_N(i) = 1 \).

B. Optimization of the adaptive parameters

In addition to parameters \( W, \beta \), 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 \( i \) at time \( n \) and hidden state \( j \) at time \( n + 1 \), given the data \( X \) and the model. In this way, the re-estimation formulas are defined as follows:

\[ \hat{\pi}_i = R_{i1} \]

(9)

\[ \hat{p}_{ij} = \sum_{n=1}^{N-1} \xi_n(i,j) \sum_{n=1}^{N-1} R_{in} \]

(10)

C. Visualization of multivariate time data series

As mentioned in the introduction, the GTM is embodied with visualization capabilities that are akin to those of the SOM. Multivariate time series can be summarily visualised in the low-dimensional latent space (1 or 2 dimensions) of GTM-TT by means of the posterior-mode projection [7], defined as

\[ i_n^{\text{map}} = \arg \max_{\{i\}} R_{in} \]

(11)

The distribution of the responsibility over the latent space of states can also be directly visualized. Both of these possibilities will be used in the next section for reporting the results of all the experiments.

IV. EXPERIMENTS

Several experiments were designed to assess the suitability of the GTM-TT model for the analysis, assisted by visualization, of multivariate time series. They are organized according to four different objectives: First, we aim to compare the different results yielded by the standard GTM and the GTM-TT when dealing with time series. This way, the advantages of using the latter model will be highlighted. Secondly, we aim to illustrate how sudden transitions (also referred to as
change points [11]) and low-variability periods are reflected on the GTM-TT latent space of states. The third goal is the exploration of the model capability of detecting anomalous data sequences (also known as surprise pattern detection or novelty detection [12]). Finally, the fourth objective is to illustrate, using the evolution over time of the data responsibilities, how regimes and their transitions are reflected in the latent space of states.

A. 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: These 6-variate time series consist of 1000 data points obtained from various inertial sensors from Space Shuttle mission STS-571. These data are particularly appropriate for the planned experiments for they contain subsequences of little variability followed by sudden transition periods. They were used for cluster detection in [13].

(3) System_data: 9-variate time series consisting of 1908 samples that describe the operation of a workstation in a networking environment over one week. These data contain long periods of low activity with interspersed short bursts of high activity, and they were used in [14] for cluster detection. Due to their incompleteness, missing values were imputed through a variant of GTM whose performance was validated in [15].

(4) Physio_data: 3-variate time series consisting of 3400 samples of physiological data, used in the Santa Fe Competition2 in 1991. They consist of three physiological variables measured in a subject while sleeping, and contain clearly atypical subsequences due to a measurement error (failure in a sensor). These data were also used in [4] to assess the performance of a variant of the standard SOM for time series.

B. Capturing the dynamics of time series through visualization: differences between GTM and GTM-TT

As stated in previous paragraphs, a comparison between the results yielded by the standard GTM and the GTM-TT is intended, with a focus on the illustration of the main differences in the visualization of the time series. Artificial_data and System_data are used for these first experiments. The Artificial_data series are displayed in Fig.1 together with their GTM posterior-mode projection, from Eq. 11, onto the latent space. In this map, the latent states in this map, the GTM latent states are represented by squares that are scaled in size according to the ratio of data points that the model understands as being generated by the associated mixture component. The mapping 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 sequence context of each point, resulting in a more disperse representation, even for almost flat signal periods. As shown in Fig.2, 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 in a single state, leaving a few surrounding states to represent the narrow activity periods. Again, the standard GTM representation is much sparser. This effect can be explained by the different meaning of the responsibility matrix in the standard GTM and in GTM-TT: each element of \( R_{in} = p(\mathbf{u}_i|\mathbf{x}_n) \) represents, in the former, the probability of the hidden state \( \mathbf{u}_i \) given \( \mathbf{x}_n \), independently of other data. Instead, in the latter, \( R_{in} = p(\mathbf{u}_i|\mathbf{X}) \) defines the probability of being in the state \( i \) at time step \( n \) of a data sequence (see Eq.6). Therefore, \( R_{in} \) in GTM-TT contains contextual information for each data point. As a result, in GTM-TT, 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.

---

1 [www.cs.uchicago.edu/~eamonn](www.cs.uchicago.edu/~eamonn)
2 [www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html](www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html)
standard GTM would represent both data subsequences in exactly the same way.

![Fig. 3. A GTM-TT membership map representing the subsequence of Artificial_data from \( n = 21 \) to \( n = 40 \) is shown in (a). The same subsequence, attached at the end of the series with all data points randomly shuffled, yields, as shown in (b), a different GTM-TT membership map displaying a different trajectory (represented by the lines between states).](image)

### C. 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. 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.4, where, as in Fig.2, 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 A is represented as a succession of brief jumps between states covering a wide map area. On the right-hand side map, a rather idle period B is mostly captured in a single state. In conclusion, the GTM-TT seems to capture these dynamics, in an intuitive and interpretable way.

![Fig. 4. (Top plot): A variable of System_data that measures idle CPU time, with two periods, corresponding to sudden transitions associated to bursts of system activity (A) and idle time (B) highlighted. (Bottom plot, left): GTM-TT membership map corresponding to period A. (Bottom plot, right): Map of the longer period 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 same number of data points; the distribution of the data over the states is nevertheless the same.](image)

A further experiment was carried out using the Shuttle_data, displayed in Fig.5. Five non-overlapping periods (A to E) were considered; all but period B contain little variability, although they are separated by sudden transitions. Once again, this is clearly reflected in the GTM-TT map of Fig.6. Confirming the results of the previous experiment, the low variability subsequences bundle up in a few highly-populated states, with 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.

![Fig. 5. 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.](image)

![Fig. 6. The subsequences A to E conforming 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 discontinuous oriented lines. The state transitions of period B are represented by a continuous oriented line.](image)

In order to make this acquired knowledge operative in trend-change and anomaly detection tasks, 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 [16], 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 difference between the probabilities of the subsequences \( X_n \) and \( X_{n-1} \), where \( X_n = \{x_1, \ldots, x_n\} \), can be used as a measure of the suddenness of transitions in multivariate time series, in the form:

\[
P(X_n) - P(X_{n-1}) = \sum_{i=1}^{M} \alpha_n (i) - \sum_{i=1}^{M} \alpha_{n-1} (i) \tag{12}\]

An easily interpretable relative index of variability, denoted \( RIV_n \), can be defined, using logarithmic differences, as:
\[ RIV_n = \frac{-\{\log P(X_n) - \log P(X_{n-1})\}}{\sum_n \{\log P(X_n) - \log P(X_{n-1})\}} \]  

(13)

According to this equation, the suddenness of the transitions will be proportional to \( RIV_n \), with a lower limit of 0.

Fig. 7 displays \( RIV_n \) for the Shuttle data series. Little variability intervals A, C, D, and E, are shown to have a \( RIV_n \) 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 small changes of \( RIV_n \), as expected.

**D. Isolating atypical subsequences**

The problem of detecting atypical subsequences in time series has attracted much attention [12] in recent times. Despite the lack of a standard definition, an atypical time series subsequence might be considered as that with a pattern differing substantially from what would be expected according to the evidence provided by the rest of the time series. Therefore, atypical subsequences might be expected to reside in rather isolated areas of the GTM-TT membership map, distinctly separated from the rest of the data. We test this hypothesis using the Physio data set. From their description, the beginning and the end of the series might contain atypical subsequences due to measurement errors caused by a system failure. The data are shown in Fig. 8, together with the corresponding GTM-TT membership map, in which the atypical subsequences have been highlighted. These maps confirm the hypothesis: most of the non-atypical data are clustered within the centre of the map, whereas the atypical data are pushed in group towards the map limits.

**E. 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 [17], [9]). 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 and Hammer [4], 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 preprocessing, where a possible transition between regimes (A) and a regime interval (B) is highlighted and considered for further analysis. The GTM-TT 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 responsabilities 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 responsabilities 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.

**F. Limitations of the GTM Through Time**

So far, the different capabilities of the GTM-TT for visual time series data exploration have been demonstrated in some detail. In spite of this, some limitations of the model must be acknowledged:

- GTM-TT requires the optimization of more parameters than the standard GTM. If probabilities were calculated for all possible hidden state transitions, the total number of parameters to estimate would become computationally intractable. Bishop et al. [9] work around the problem by assuming a prior knowledge about the nature of the transitions between different time steps. However, this assumption would introduce additional free parameters and might not work out for certain time series. Alternatively, in [17] a prior normal distribution over the transitions, not optimized throughout the training, is defined. This assumption would be suitable for locally homogeneous time series.
The model have also been summarized. Some limitations of subsequences, and of regimes and regime transition periods, of atypical subsequences, and of regimes and regime transition maps regions is clearly observed. A novel index for the detection of sudden transitions has been provided and discussed. Furthermore, a scaling procedure \cite{18} is required, which entails additional CPU time.

\begin{itemize}
  \item For long multivariate time series, calculations of $\alpha_n(i)$ and $\beta_n(i)$ (Eqs. 7 and 8) might tend exponentially towards zero, making the results of the algorithm unreliable. Therefore, a scaling procedure \cite{18} is required, which entails additional CPU time.
\end{itemize}

\section{V. Conclusions}

The capabilities of the GTM-TT model for exploratory analysis of multivariate time series have been assessed in some detail. Visualizations of sudden transitions and low variability periods, of atypical subsequences, and of regimes and regime transitions have been provided and discussed. Furthermore, a novel index for the detection of sudden transitions has been defined and successfully applied. Some limitations of the model have also been summarized.