

# Chapter 8

## Clustering and Visualization of Multivariate Time Series

**Alfredo Vellido**

*Universidad Politécnic de Cataluña, Spain*

**Iván Olier**

*Universidad Politécnic de Cataluña, Spain*

### ABSTRACT

*The exploratory investigation of multivariate time series (MTS) may become extremely difficult, if not impossible, for high dimensional datasets. Paradoxically, to date, little research has been conducted on the exploration of MTS through unsupervised clustering and visualization. In this chapter, the authors describe generative topographic mapping through time (GTM-TT), a model with foundations in probability theory that performs such tasks. The standard version of this model has several limitations that limit its applicability. Here, the authors reformulate it within a Bayesian approach using variational techniques. The resulting variational Bayesian GTM-TT, described in some detail, is shown to behave very robustly in the presence of noise in the MTS, helping to avert the problem of data overfitting.*

### INTRODUCTION

The analysis of MTS is an established research area, and methods to carry it out have stemmed both from traditional statistics and from the Machine Learning and Computational Intelligence fields. In this chapter, we are mostly interested in the latter, but considering a mixed approach that can be ascribed to Statistical Machine Learning.

MTS are often analyzed for prediction and forecasting and, therefore, the problem is considered to be supervised. In comparison, little research has been conducted on the problem of unsupervised clustering for the exploration of the dynamics of multivariate time series (Liao, 2005). It is sensible to assume that, in many problems involving MTS, the states of a process may be reproduced or *revisited* over time; therefore, clustering structure is likely to be found in the series. Furthermore, for exploratory purposes, it

DOI: 10.4018/978-1-60566-766-9.ch008

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.

It has been argued (Keogh and Lin, 2005) that, in some cases, time series clustering is a meaningless endeavour. It is generally agreed, though, that this might only apply to certain types of time series clustering methods, such as those resorting 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, clustering using Self-Organizing Maps (SOM: Kohonen, 2001) 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 the statistically principled models we present in this chapter, over the latent space of states) should be significantly different. In this chapter we analyze 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 (which is the case).

One of the most commonly used Machine Learning methods for the clustering of MTS is indeed SOM, in general without accounting for the violation of the independent identically distributed (i.i.d.) condition. Several extensions of SOM have been developed to explicitly accommodate time series through recurrent connectivity (Chappell & Taylor, 1993; Strickert & Hammer, 2005; Tiño, Farkas & van Mourik, 2006; Voetglin, 2002). 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., Kostiainen & Lampinen, 2002; Yin & Allinson, 2001), it has mostly retained its heuristic definition, which is at the origin of some of its limitations. Generative Topographic Mapping (GTM: Bishop, Svensén & Williams, 1998a) is a nonlinear latent variable model that was originally devised as a probabilistic alternative to SOM that aimed to overcome aforementioned limitations. Its probability theory foundations have 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-Deredey & Lisboa, 2003), discrete data modelling (Bishop, Svensén & Williams, 1998b; Girolami, 2002), and robust outlier detection and handling (Bullen, Cornford & Nabney, 2003; Vellido & Lisboa, 2006), amongst others.

The GTM Through Time (henceforth referred to as GTM-TT: Bishop, Hinton & Strachan, 1997) is one such extension of GTM for the analysis of MTS. It explicitly accounts for the violation of the i.i.d. condition, given that it is defined as a constrained hidden Markov model (HMM: Rabiner, 1989). Its original formulation has several limitations, including the lack of adequate regularization procedures, which makes it prone to be negatively affected by the presence of uninformative noise in the MTS. In its basic formulation, the GTM is trained within the Maximum Likelihood (ML) framework using the Expectation-Maximization (EM) algorithm, and overfitting may occur unless regularization methods are applied. In (Bishop, Svensén & Williams, 1998b; Vellido, El-Deredey & Lisboa, 2003), regularization strategies based on Bayesian evidence approaches were introduced. Unfortunately, they require a number of modelling assumptions and approximations to be made. An alternative for the formulation of GTM that confers the model with regularization capabilities, while avoiding such approximations, is that of using variational techniques (Beal, 2003). A Variational GTM model based on the GTM with a Gaussian Process (GP) prior with added Bayesian estimation of its parameters was recently described in (Olier & Vellido, 2008). This Variational GTM was shown to limit the negative effect of data overfitting, while retaining the data visualization capabilities of the model. In this chapter, we describe the extension of the variational approach to the analysis of MTS, defining the Variational Bayesian GTM-TT.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/clustering-visualization-multivariate-time-series/36985](http://www.igi-global.com/chapter/clustering-visualization-multivariate-time-series/36985)

## Related Content

---

### Rough and Soft Set Approaches for Attributes Selection of Traditional Malay Musical Instrument Sounds Classification

Norhalina Senan, Rosziati Ibrahim, Nazri Mohd Nawi, Iwan Tri Riyadi Yantoand Tutut Herawan (2012). *International Journal of Software Science and Computational Intelligence* (pp. 14-40).

[www.irma-international.org/article/rough-soft-set-approaches-attributes/72878](http://www.irma-international.org/article/rough-soft-set-approaches-attributes/72878)

### Role of Artificial Intelligence in Water Management

Tarak Nandyand Sananda Bhattacharyya (2026). *Computational Intelligence and Optimization Methods for Sustainable Water Management* (pp. 163-208).

[www.irma-international.org/chapter/role-of-artificial-intelligence-in-water-management/392462](http://www.irma-international.org/chapter/role-of-artificial-intelligence-in-water-management/392462)

### Sitting Posture Recognition and Location Estimation for Human-Aware Environment

Yusuke Manabeand Kenji Sugawara (2011). *International Journal of Software Science and Computational Intelligence* (pp. 34-49).

[www.irma-international.org/article/sitting-posture-recognition-location-estimation/53161](http://www.irma-international.org/article/sitting-posture-recognition-location-estimation/53161)

### Fuzzy-DSS Human Health Risk Assessment Under Uncertain Environment

Palash Dutta (2018). *Handbook of Research on Investigations in Artificial Life Research and Development* (pp. 316-347).

[www.irma-international.org/chapter/fuzzy-dss-human-health-risk-assessment-under-uncertain-environment/207210](http://www.irma-international.org/chapter/fuzzy-dss-human-health-risk-assessment-under-uncertain-environment/207210)

### Detection of Microcalcifications on Mammograms

Rachida Touami, Karima KIESand Nacéra Benamrane (2020). *International Journal of Software Science and Computational Intelligence* (pp. 68-79).

[www.irma-international.org/article/detection-of-microcalcifications-on-mammograms/250861](http://www.irma-international.org/article/detection-of-microcalcifications-on-mammograms/250861)