Time-varying autoregressions with model order uncertainty

Raquel Prado* and Gabriel Huerta†

Abstract

We explore some aspects of the analysis of latent component structure in non-stationary time series based on time-varying autoregressive (TVAR) models that incorporate uncertainty about model order. Our approach assumes that at each time, model order evolves according to a discrete random walk and can take values up to a specified upper bound. Priors that are conjugate-normal are considered for the autoregressive coefficients and the evolution of such coefficients over time is specified through a random walk. Simulation from the posterior distribution of the AR coefficients and model order can be obtained via a Gibbs sampling format based on *Forward Filtering Backward Simulation* algorithms. Aspects of model implementation and inference on decompositions and latent structure are discussed for some synthetic data and for electroencephalogram (EEG) traces previously analysed with fixed order TVAR models.

Keywords: Dynamic linear models; Time-varying autoregressions; Model uncertainty; Time series decompositions; MCMC

*Departamento de Cómputo Científico y Estadística & Centro de Estadística y Software Matemático (CESMa), Universidad Simón Bolívar, Apartado 89000, Caracas, Venezuela. *E-mail:* raquel@cesma.usb.ve.

†Department of Statistics, Northwestern University, 2006 Sheridan Road, Evanston, Illinois, 60208 USA. *E-mail:* gabriel@bayes.stats.nwu.edu.

1 Introduction

Modern statistical analysis recognises the relevance of model uncertainty in parameter estimation and inference. From a Bayesian perspective, model uncertainty can be handled and summarised through the computation of posterior model probabilities. Additionally, in a time series context, it is also crucial to study how prediction on future values and inference on latent structure is affected by the propagation of uncertainty. This is the main goal of our paper within the framework of time-varying parameter autoregressive (TVAR) models and using Markov Chain Monte Carlo (MCMC) methods.

There is a vast literature of time series models that incorporate model uncertainty with MCMC methods. For instance, in the case that the models in consideration are all linear autoregressive (AR) processes, Barnett *et al.* (1996) presented a MCMC, based on a stochastic variable search approach, that deals with model order uncertainty through priors on the partial autocorrelations that restrict to stationarity. On the same line, Barbieri and O'Hagan (1997) developed a MCMC using a similiar parameterisation, but based on the reversible jump (Green, 1995) to produce posterior inference on model order. Troughton and Godsill (1997) developed an efficient MCMC method that uses the reversible jump to explore the posterior distributions for model order, model coefficients and innovation variance. Their priors are defined on the standard AR coefficients rather than on the autocorrelation function. More recently, Huerta and West (1999) incorporated model order uncertainty in the linear AR framework with emphasis on prior specification for latent structure. This leads to a novel class of prior distributions on the characteristic reciprocal roots of the process. In terms of posterior simulation, their MCMC appeals both, to a stochastic variable selection and reversible jump ideas. These references illustrate how model uncertainty may be incorporated with MCMC methods in the context of linear and/or stationary time series models.

In the case of more general and non-stationary Dynamic Linear Models (DLM), West and Harrison (1997), chapter 12 and following Harrison and Stevens (1976), presents an approach to include model uncertainty through mixtures when each of the models in consideration is a conjugate DLM. This approach is known as Multi-Process models for which $p(k|D_t)$, the probability of model k given all the information up to time t, denoted by D_t, is available in analytic form. The class of DLMs is broad and quite flexible; time series models that incorporate cyclical patterns and trends can be expressed in this form. It is worth noting that when some of the DLMs in consideration are not conjugate but conditionally conjugate, the *Forward Filtering Backward Simulation* algorithms in Carter and Kohn (1994) and Frühwirth-Schnatter (1994) can be applied to compute the posterior probabilities of each model. In the framework where there is no conditional conjugacy, posterior probabilities may be obtained via *particle filter* methods as reported by Pitt and Shephard (1999). Andrieu *et al.* (1999) constitutes a recent reference in this direction; model order uncertainty and sequential updating are addressed for an autoregressive process with observational noise. The algorithm is based on *particle filters* selected with Bayesian importance sampling and MCMC reversible jump steps.

In this paper, we deal with model uncertainty restricting the potential models to the class of time-varying autoregressive models or TVAR. This class of models has been successful in studying underlying structure of non-stationary time series in many applications, in particular, in analysis of various kinds of electroencephalographic (EEG) signals. Some of

the references in this area are Gersch (1987), Kitagawa and Gersch (1996), Prado and West (1997), West *et al.* (1999) and Krystal *et al.* (1999). Specification of TVAR models and decomposition theory of non-stationary time series, based on flexible DLM representations, are developed in West *et al.* (1999). Here, we review these references and related theory of decomposition and underlying structure analysis in Section 2. In Section 3, we extend and discuss how to address model order uncertainty in TVAR models based on standard conjugate priors for model coefficients, variances and a discrete uniform prior on model order. Furthermore, issues of time series decompositions and latent structure are discussed within this class of TVAR models. In Sections 4 and 5, we apply the models and methodology developed in Section 3, to study latent structure in synthetic data and EEG traces. Concluding remarks and possible extensions of the modelling framework are discussed in Section 6.

2 The class of TVAR models and decompositions

In this Section we summarise the model specification and decomposition results developed in West *et al.* (1999) for the class of time-varying autoregressions. A time-varying autoregression of order p or TVAR(p) is described by,

$$x_t = \sum_{j=1}^{p} \phi_{t,j} x_{t-j} + \epsilon_t, \tag{1}$$

where $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p})'$ is the time-varying parameter vector and ϵ_t are zero-mean independent innovations assumed Gaussian with possibly time-varying variances $\epsilon_t \sim N(0, \sigma_t^2)$. No explicit stationarity constraints are imposed on the AR parameters at each time t. However, if such parameters lie in the stationarity region the series can be thought as locally stationary and the changes in the parameters over time represent global non-stationarities. The model is completed by specifying the evolution components for the time-varying AR parameters $\boldsymbol{\phi}_t$ and the innovations variance σ_t^2. The evolution on the AR parameters is taken as a random walk $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1} + \boldsymbol{\xi}_t$, with zero mean innovations $\boldsymbol{\xi}_t$ that are uncorrelated and normal $\boldsymbol{\xi}_t \sim N(0, \mathbf{W}_t)$. Similarly, the changes in time of σ_t^2 are modelled with a multiplicative random walk $\sigma_t^2 = \sigma_{t-1}^2(\delta/\eta_t)$ where η_t are mutually independent and independent of ϵ_t and $\boldsymbol{\xi}_t$, with $\boldsymbol{\xi}_t \sim Be(a_t, b_t)$ (West and Harrison, 1997 chapter 10). Time variation in the AR coefficients and the innovations variance is controlled via standard discount factors β and $\delta \in (0, 1)$. These discount factors measure the degree of variation in time of $\boldsymbol{\phi}_t$ and σ_t^2 respectively. Low values of the discount factors are consistent with high variability of the parameters over time, but values closer to one are typically more relevant in practice. Sequential updating and retrospective filtering/smoothing algorithms (West and Harrison, 1997) can be applied to obtain posterior distributions for the AR parameters $\boldsymbol{\phi}_t$, and the evolution variances σ_t^2.

The basic decomposition result for the class of TVAR models (West *et al.*, 1999), arises from writing the TVAR model in a particular DLM form and then using standard theory of model structure and the notions of similar models in linear stochastic systems (West and Harrison 1997, chapter 5). The TVAR model in (1) has a DLM form,

$$x_t = \mathbf{F}'\boldsymbol{\theta}_t, \quad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \tag{2}$$

where $\mathbf{F} = (1, 0, \ldots, 0)'$, $\mathbf{x}_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})'$, $\boldsymbol{\omega}_t = \epsilon_t \mathbf{F}$ and

$$\mathbf{G}_t \equiv \mathbf{G}(\boldsymbol{\phi}_t) = \begin{pmatrix} \phi_{t,1} & \phi_{t,2} & \ldots & \phi_{t,p-1} & \phi_{t,p} \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \tag{3}$$

for each t. Note that the eigenvalues of $\mathbf{G}_t$ are the reciprocals of the roots of the autoregressive characteristic equation at each time t. Suppose that, at each time t, $\mathbf{G}_t$ has p distinct eigenvalues, with c pairs of complex eigenvalues denoted by $r_{t,j} \exp(\pm i\omega_{t,j})$ for $j = 1, \ldots, c$, and $r = p - 2c$ real eigenvalues denoted by $r_{t,j}$ for $j = 2c+1, \ldots, p$. Then, the basic decomposition result for the class of TVAR states that

$$x_t = \sum_{j=1}^{c} z_{t,j} + \sum_{j=2c+1}^{p} y_{t,j}, \tag{4}$$

where the $z_{t,j}$ processes are defined through the complex eigenvalues and the $y_{t,j}$ through the real eigenvalues. In particular, for the standard AR(p) process, $\boldsymbol{\phi}_t = \boldsymbol{\phi}$, $\mathbf{G}_t = \mathbf{G}$, and the eigenvalues of $\mathbf{G}$ are the reciprocals of the roots of the usual AR characteristic equation. In this case $r_{t,j} = r_j$ for $j = 1, \ldots, p$ and $\omega_{t,j} = \omega_j$ for $j = 1, \ldots, c$. Furthermore, each $y_{t,j}$ follows a standard AR(1) process with AR parameter r_j, and each $z_{t,j}$ follows an ARMA(2,1) whose AR(2) component is quasi-periodic with time-varying amplitude and phase and constant characteristic frequency and modulus ω_j and r_j respectively. In the general TVAR case, the interpretation of the latent processes may be slightly different. In many problems of applied interest (West *et al.* 1999, Krystal *et al.* 1999) $\boldsymbol{\phi}_t$ changes slowly in time and the eigenvectors of $\mathbf{G}_t$ and $\mathbf{G}_{t-1}$ are very similar. As a result, the processes $y_{t,j}$ approximately follow a TVARMA(1) with time-varying AR parameter $r_{t,j}$ and the $z_{t,j}$ processes approximately follow a TVARMA(2,1) with time-varying characteristic frequency $\omega_{t,j}$ and modulus $r_{t,j}$. Now, if the changes in time of the $\boldsymbol{\phi}_t$ parameters, and therefore the changes in the eigenstructure of $\mathbf{G}_t$, are more variable, the stochastic structure of the latent processes is not represented by TVAR(1) and TVARMA(2,1) components (details appear in Prado, 1998).

We present the above decomposition for an EEG series taken from West *et al.* (1999). Figure 1 displays the data and estimated components in the decomposition of the series, based on a TVAR(12) model with discount factors $\delta = 1.0$ and $\beta = 0.997$. The EEG trace was recorded on a patient who received a moderate level of ECT stimulus intensity. Components (1), (2), and (3) in the decomposition are the highest amplitude components lying in the delta (0 to 4 Hz) and theta (4 to 8 Hz) frequency bands. These components are individual processes dominated by a TVARMA(2,1) quasi-periodic structure. For instance, process (1) has a TVARMA(2,1) quasi-periodic structure with time-varying characteristic frequency lying in a delta band that starts around 4 Hz and gradually decays towards the end of the seizure. The time-varying characteristic modulus of process (1) is consistently high, with values higher than 0.95 over the seizure course. Components (4), (5) and (6) are low amplitude components representing neural and experimental noise. The end of the seizure occurs at around $t = 1800$. Clearly, the contribution of components (2), (3), (4), (5)

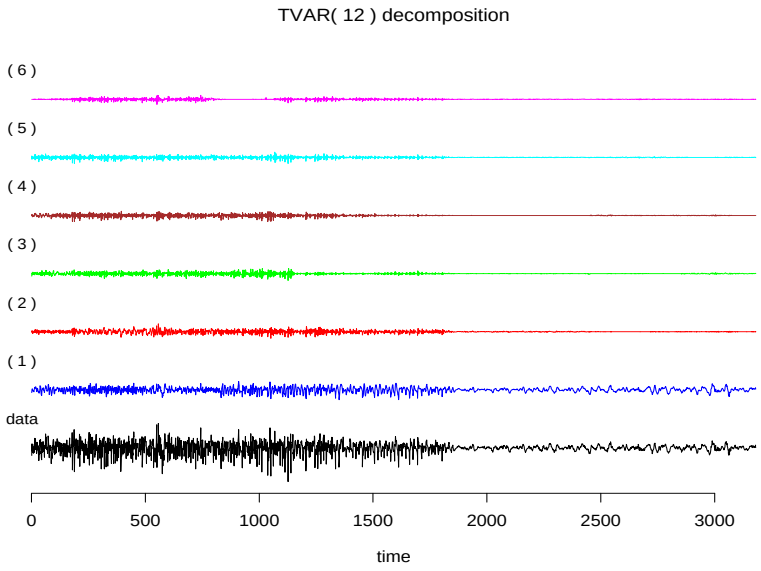

Figure 1: Data and estimated components in the decomposition of an EEG series based on a TVAR(12). From the bottom up, the graph displays the series followed by estimated components in order of decreasing amplitude.

and (6) is practically negligible after $t = 1800$. The structure of the EEG signal is more complex while the seizure starts and matures than when it begins to decay and eventually dies off. Thus, it seems reasonable to consider a TVAR model with a higher model order during the beginning and middle parts of the seizure than towards the end of the seizure.

3 Autoregressions with time-variation on the AR coefficients and model order

A time-varying autoregression with time-varying order p_t or TVAR(p_t) is described by,

$$x_t = \sum_{j=1}^{p_t} \phi_{t,j} x_{t-j} + \epsilon_t, \tag{5}$$

where the autoregressive coefficients evolve according to a random walk as defined in Section 2 for the fixed order TVAR, and $\epsilon_t \sim N(0, \sigma^2)$. We assume that p_t, the model order at time t, is an integer that can only take values from a lower bound $p_{\min}$, up to a fixed upper bound $p_{\max}$ both assumed fixed. Equivalently, we may write the TVAR(p_t) via a set of indicator variables,

$$x_t = \sum_{j=1}^{p_{\max}} \mathit{1}_{t,j} \phi_{t,j} x_{t-j} + \epsilon_t, \tag{6}$$

where $\mathit{1}_{t,j}$ can only take the values 0 or 1 and the vector of coefficients $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p_{\max}})'$ has always dimension $p_{\max}$. In addition, we define $\mathit{1}_t = (\mathit{1}_{t,1}, \mathit{1}_{t,2}, \ldots, \mathit{1}_{t,p_{\max}})$, the vector of

indicator variables at time t. Notice that each of the possible $2^{(p_{\min}-p_{\max})}$ vectors $\mathbf{1}_t$ indexes a particular model order. Furthermore, we reduce the model space so that for every time t, if an autoregression of lag j is significant, i.e. $\phi_{t,j} \neq 0$, any smaller lag is also included in the model, i.e. $\phi_{t,s} \neq 0$ for $s < j$. This imposes the following restriction on $\mathbf{1}_t$: if $\mathbf{1}_{t,j} = 0$ for a particular j, then $\mathbf{1}_{t,s} = 0$ for all values of the index s greater than j. In consequence, whenever $\sum_{j=1}^{p_{\max}} \mathbf{1}_{t,j} = p_t$, $\mathbf{1}_t$ can only be a vector whose first p_t entrances are ones and the remaining are zero.

Model completion requires specification of an initial prior $p(\boldsymbol{\phi}_1, \sigma_1^2 | D_0)$ where D_0 denotes initial information and all details of the model structure. For simplicity, we assume $\mathbf{1}_1 = (1, 1, 1, \ldots, 1)$ so that $p_1 = p_{\max}$, and the standard conjugate framework based on normal/inverse gamma distributions with a relative diffuse initial prior. The evolution of p_t is defined through a first order random walk whose transition probabilities are denoted by $P[p_t = i | p_{t-1} = j]$ and the possible values for i and j range from $p_{\min}$ to $p_{\max}$. Particular forms for this transition matrix will be specified in the applications.

In this framework, estimation of the TVAR(p_t) follows a two-stage Gibbs sampling format. Conditional on the sequence p_t, or equivalently on the sequence of vectors $\mathbf{1}_t$ over $t = 1, \ldots n$, the standard sequential updating and retrospective filtering/smoothing algorithms for DLMs apply. Based on all the observed information $D_n = \{D_0, x_1, x_2, \ldots, x_n\}$ the sequences $\boldsymbol{\phi}_t$ and σ^2 are simulated from joint normal/gamma distributions that depend on p_t; $t = 1, 2, \ldots, n$. On the other hand, given $\boldsymbol{\phi}_t$ for all t and σ^2, we sample from the conditional posterior distribution of p_t via the filtering/smoothing algorithm for discrete random variables described in Carter and Kohn (1994). For clarity of exposition, we include in D_t the conditioning sequences $\boldsymbol{\phi}_t$ and σ^2. For each $t = 1, 2, \ldots, n$, we compute the conditional prior probabilities for p_t, $P[p_t = i | D_{t-1}]$; for $i = p_{\min}, \ldots, p_{\max}$ by marginalising the joint probability distribution $P[p_t = i, p_{t-1} = j | D_{t-1}]$ over the values of j. Through Bayes theorem, we compute the posterior probabilities for p_t, $P[p_t = i | D_t]$ and keep the values for both, the prior and posterior probabilities at each time t. At this point, we generate a value p_n from the discrete distribution $P[p_n = i | D_n]$. Then, for $t = n-1, n-2, \ldots, 1$, we compute $P[p_t = i | p_{t+1}, D_t]$ for all $i = p_{\min}, \ldots, p_{\max}$; where p_{t+1} is the sampled value for model order at time $t+1$. We generate a value p_t from this discrete distribution and continue in this way until we reach $t = 1$. The collection of values $p_n, p_{n-1}, p_{n-2}, \ldots, p_1$ conform a sample from the conditional posterior distribution for model orders. Full details of the Gibbs sampling scheme are presented in the Appendix.

3.1 Decompositions for TVAR models with time-varying order

The TVAR model with time-varying order p_t at each time t, can be written in DLM form as

$$x_t = \mathbf{F}' \mathbf{x}_t, \quad \mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \boldsymbol{\omega}_t$$

where $\mathbf{F}$ is a $p_{\max} \times 1$ vector $\mathbf{F} = (1, 0, \ldots, 0)'$, $\mathbf{x}_t = (x_t, x_{t-1}, \ldots, x_{t-p_{\max}+1})'$, $\boldsymbol{\omega}_t = \epsilon_t \mathbf{F}$ and $\mathbf{G}_t$ a $p_{\max} \times p_{\max}$ matrix,

$$\mathbf{G}_t \equiv \mathbf{G}(\boldsymbol{\phi}_t) = \begin{pmatrix} \phi_{t,1} & \phi_{t,2} & \cdots & \phi_{t,p_t-1} & \phi_{t,p_t} & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & \vdots & & \ddots & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & & 1 & 0 \end{pmatrix}. \tag{7}$$

Then, at time t, only p_t coefficients appear in the matrix $\mathbf{G}_t$. Assume that $\mathbf{G}_t$ has p_t distinct non-zero eigenvalues $\alpha_{1,t}, \ldots, \alpha_{p_t,t}$ and a zero eigenvalue $\alpha = 0$ with multiplicity $p_{\max} - p_t$. The non-zero eigenvalues correspond to the reciprocal roots of the characteristic polynomial at time t, $\Phi_{p_t}(u) = (1 - \phi_{1,t}u - \ldots - \phi_{t,p_t}u^{p_t})$. Then, $\mathbf{G}_t = \mathbf{E}_t \mathbf{A}_t \mathbf{E}_t^{-1}$ with

$$\mathbf{A}_t = \text{block diag}[\mathbf{A}_{p_t}, \mathbf{J}_{(p_{\max}-p_t)}(0)], \quad \mathbf{E}_t = [\mathbf{e}_{1,t}, \ldots, \mathbf{e}_{p_t,t}, \mathbf{h}_{1,t}, \ldots, \mathbf{h}_{(p_{\max}-p_t),t}],$$

where $\mathbf{A}_{p_t} = \text{diag}(\alpha_{t,1}, \ldots, \alpha_{t,p_t})$ and $\mathbf{J}_{(p_{\max}-p_t)}(0)$ is the $(p_{\max} - p_t) \times (p_{\max} - p_t)$ Jordan block associated to the eigenvalue $\alpha = 0$,

$$\mathbf{J}_{(p_{\max}-p_t)}(0) = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

$\mathbf{E}_t$ is a $p_{\max} \times p_{\max}$ matrix whose first p_t columns correspond to the p_t eigenvectors associated to the non-zero distinct eigenvalues $\alpha_{1,t}, \ldots, \alpha_{p_t,t}$. The last $p_{\max} - p_t$ columns of the matrix $\mathbf{E}_t$ correspond to the $p_{\max} \times 1$ vectors $\mathbf{h}_{1,t}, \ldots, \mathbf{h}_{(p_{\max}-p_t),t}$, where each $\mathbf{h}_{j,t}$ is such that its last j components are ones and its first $p_{\max} - j$ components are zeros. Some of the non-zero eigenvalues could be complex and in such case they appear in pairs of complex conjugates. Assume that at each time t, there are c_t pairs of complex eigenvalues denoted by $r_{t,j} \exp(\pm i\omega_{t,j})$ for $j = 1, \ldots, c_t$ and $r_t = p_t - 2c_t$ non-zero real and distinct eigenvalues denoted by $r_{t,j}$ for $j = 2c_t + 1, \ldots, p_t$. For each time t, define the matrix $\mathbf{H}_t = \mathbf{D}_t \mathbf{E}_t^{-1}$, with $\mathbf{D}_t = \text{diag}(\mathbf{E}_t'\mathbf{F})\mathbf{E}_t^{-1}$, and linearly transform $\mathbf{x}_t$ via $\boldsymbol{\gamma}_t = \mathbf{H}_t \mathbf{x}_t$. Then, we can write $x_t = (1, \ldots, 1)'\boldsymbol{\gamma}_t$ or equivalently,

$$x_t = \sum_{j=1}^{c_t} z_{t,j} + \sum_{j=2c_t+1}^{p_t} y_{t,j}, \tag{8}$$

with $z_{t,j} = \gamma_{t,2j-1} + \gamma_{t,2j}$ for $j = 1, \ldots, c_t$, and $y_{t,j} = \gamma_{t,j}$ for $j = 2c_t + 1, \ldots, p_t$. Notice that the decomposition result is analogous to (4) but now the number of components depend on time-varying c_t and r_t.

In order to gain insight on the structure of the latent processes $z_{t,j}$ and $y_{t,j}$, consider the models $\mathcal{M}_i$ for $i = p_{\min}, \ldots, p_{\max}$,

$$\mathcal{M}_i : \quad x_t = \mathbf{F}'\mathbf{x}_t, \quad \mathbf{x}_t = \mathbf{G}_t^{(i)}\mathbf{x}_{t-1} + \boldsymbol{\omega}_t$$

with

$$\mathbf{G}_t^{(i)} = \begin{pmatrix} \phi_{t,1} & \phi_{t,2} & \ldots & \phi_{t,i-1} & \phi_{t,i} & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix},$$

i.e. model $\mathcal{M}_i$ is a TVAR(i). For each model $\mathcal{M}_i$ reparameterise $\mathbf{x}_t$ and $\boldsymbol{\omega}_t$ via $\boldsymbol{\gamma}_t^{(i)} = \mathbf{H}_t^{(i)}\mathbf{x}_t$ and $\boldsymbol{\delta}_t^{(i)} = \mathbf{H}_t^{(i)}\boldsymbol{\omega}_t$ with $\mathbf{H}_t^{(i)} = \mathbf{D}_t^{(i)}(\mathbf{E}_t^{(i)})^{-1}$. Then we have,

$$x_t = \mathbf{1}'\boldsymbol{\gamma}_t^{(i)} \quad \boldsymbol{\gamma}_t^{(i)} = \mathbf{A}_t^{(i)}\mathbf{K}_t^{(i)}\boldsymbol{\gamma}_{t-1}^{(i)} + \boldsymbol{\delta}_t^{(i)},$$

with $\mathbf{K}_t^{(i)} = \mathbf{D}_t^{(i)}(\mathbf{E}_t^{(i)})^{-1}\mathbf{E}_{t-1}^{(i)}\mathbf{D}_{t-1}^{(i)*} = \mathbf{H}_t^{(i)}\mathbf{H}_{t-1}^{(i)*}$. $\mathbf{D}_t^{(i)*}$, $\mathbf{H}_{t-1}^{(i)*}$ are generalised inverse matrices of $\mathbf{D}_t^{(i)}$ and $\mathbf{H}_{t-1}^{(i)}$ respectively. Given known, estimated or simulated values of $\mathbf{F}$, $\mathbf{G}_t^{(i)}$ and $\mathbf{x}_t$ for each t, we can obtain a decomposition for x_t based on model $\mathcal{M}_i$ for each $i = p_{\min}, \ldots, p_{\max}$,

$$x_t = \sum_{j=1}^{c^{(i)}} z_{t,j}^{(i)} + \sum_{j=2c^{(i)}+1}^{i} y_{t,j}^{(i)},$$

such that $r^{(i)} = i - 2c^{(i)}$. The value $c^{(i)}$ is the number of pairs of complex non-zero eigenvalues of $\mathbf{G}_t^{(i)}$, denoted by $r_{t,j}^{(i)} \exp(\pm i\omega_{t,j}^{(i)})$ for $j = 1, \ldots, c^{(i)}$, and $r^{(i)}$ is the number of real eigenvalues. As in the standard TVAR case and for simplicity, we are assuming that $c^{(i)}$ and $r^{(i)}$ are fixed in time (see West *et al.* 1999). In cases where the AR coefficients change slowly in time, $\mathbf{K}_t^{(i)} \approx \text{blockdiag}[\mathbf{I}_{i\times i}, \mathbf{0}_{(p_{\max}-i)\times(p_{\max}-i)}]$. Then, each $z_{t,j}^{(i)}$ is dominated by a TVARMA(2,1) process with time-varying modulus $r_{t,j}^{(i)}$ and frequency $\omega_{t,j}^{(i)}$, and each $y_{t,j}^{(i)}$ is dominated by a TVAR(1) with time-varying modulus $r_{t,j}^{(i)}$. On the other hand, if the changes in the AR coefficients in time are highly variable between times t and $t-1$, the latent processes $z_{t,j}^{(i)}$ and $y_{t,j}^{(i)}$ are "mixed" via $\mathbf{K}_t^{(i)}$ and the stochastic latent structure is not exactly represented by the TVAR(1) and TVARMA$(2,1)$.

The general decomposition result (8) for TVAR models with time-varying model order p_t, is such that at a particular time t, $z_{t,j} = z_{t,j}^{(p_t)}$, $y_{t,j} = y_{t,j}^{(p_t)}$, $r_{t,j} = r_{t,j}^{(p_t)}$ and $\omega_{t,j} = \omega_{t,j}^{(p_t)}$. Then, each $z_{t,j}$ in (8) coincides instantaneously, specifically at time t, with a process that, conditionally on smooth changes in the AR parameters over time, is dominated by a TVARMA(2,1) and has instantaneous modulus and frequency $r_{t,j}^{(p_t)}$ and $\omega_{t,j}^{(p_t)}$. Similarly, each $y_{t,j}$ coincides at time t with a process that is dominated by a TVAR(1) with time-varying modulus $r_{t,j}^{(p_t)}$.

3.1.1 Example

Consider a time series x_t, $t = 1, \ldots, 500$ such that for $t = 1, \ldots, 250$ follows a TVAR(2) and for $t = 251, \ldots, 500$ follows a TVAR(3). Figure 2 (c) displays a sequence of 500 observations generated from a TVAR(3) with a pair of complex reciprocal roots with time-varying modulus and wavelength denoted by $r_{1,t}$ and $\lambda_{1,t}$, and a real characteristic reciprocal root with modulus $r_{2,t}$. Figures 2 (d), (g) and (e) display the trajectories in time of $r_{1,t}$, $\lambda_{1,t}$ and $r_{2,t}$ respectively. Note that $r_{t,2}$ is zero for $t = 1, \ldots, 250$ indicating that we have a TVAR(p_t) with $p_t = 2$ for $t = 1, \ldots, 250$ and $p_t = 3$ for $t = 251, \ldots, 500$. This particular TVAR(p_t) can be written in DLM form with $\mathbf{F} = (1, 0, 0)'$, $\mathbf{x}_t = (x_t, x_{t-1}, x_{t-2})'$, $\boldsymbol{\omega}_t = (\epsilon_t, 0, 0)$,

$$\mathbf{G}_t = \begin{pmatrix} \phi_{1,t} & \phi_{2,t} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{for} \quad t = 1, \ldots, 250,$$

and

$$\mathbf{G}_t = \begin{pmatrix} \phi_{t,1} & \phi_{t,2} & \phi_{t,3} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{for} \quad t = 251, \ldots, 500.$$

Then, for $t = 1, \ldots, 250$, $\mathbf{G}_t = \mathbf{E}_t \mathbf{A}_t \mathbf{E}_t^{-1}$ with,

$$\mathbf{A}_t = \begin{pmatrix} r_{1,t} \exp(i2\pi/\lambda_{t,1}) & 0 & 0 \\ 0 & r_{1,t} \exp(-i2\pi/\lambda_{t,1}) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and $\mathbf{E}_t = [\mathbf{e}_{t,1}, \bar{\mathbf{e}}_{t,1}, \mathbf{h}_{t,1}]$ where $\mathbf{e}_{t,1}$ and $\bar{\mathbf{e}}_{t,1}$ are a pair of complex conjugate eigenvectors associated to the complex characteristic roots $r_{1,t} \exp(\pm i2\pi/\lambda_{t,1})$ and $\mathbf{h}_{1,t} = (0, 0, 1)'$. Similarly, for $t = 251, \ldots, 500$ we have $\mathbf{G}_t = \mathbf{E}_t \mathbf{A}_t \mathbf{E}_t^{-1}$ with

$$\mathbf{A}_t = \begin{pmatrix} r_{1,t} \exp(i2\pi/\lambda_{t,1}) & 0 & 0 \\ 0 & r_{1,t} \exp(-i2\pi/\lambda_{t,1}) & 0 \\ 0 & 0 & r_{2,t} \end{pmatrix},$$

and $\mathbf{E}_t = [\mathbf{e}_{t,1}, \bar{\mathbf{e}}_{t,1}, \mathbf{e}_{t,2}]$ with $\mathbf{e}_{t,1}, \bar{\mathbf{e}}_{t,1}, \mathbf{e}_{t,2}$ the eigenvalues associated to $r_{1,t} \exp(\pm i2\pi/\lambda_{t,1})$ and $r_{t,2}$. The decomposition for x_t is

$$x_t = \begin{cases} z_{1,t} & \text{for} \quad t = 1, \ldots, 250 \\ z_{1,t} + y_{1,t} & \text{for} \quad t = 251, \ldots, 500. \end{cases} \tag{9}$$

Figures 2 (a) and (b) display the processes $z_{1,t}$ and $y_{t,1}$ over time. Note that the component $y_{t,1}$ only appears for $t = 251, \ldots, 500$ and that its contribution to the decomposition of the series is practically negligible due to the fact that its modulus is never higher than 0.14. Now, in order to explore the structure of the latent processes in the decomposition we consider $\mathcal{M}_2$ and $\mathcal{M}_3$,

$$\mathcal{M}_2 : x_t = \mathbf{F}' \mathbf{x}_t, \quad \mathbf{x}_t = \mathbf{G}_t^{(2)} \mathbf{x}_{t-1} + \boldsymbol{\omega}_t, \quad \text{for all} \quad t = 1, \ldots, 500$$

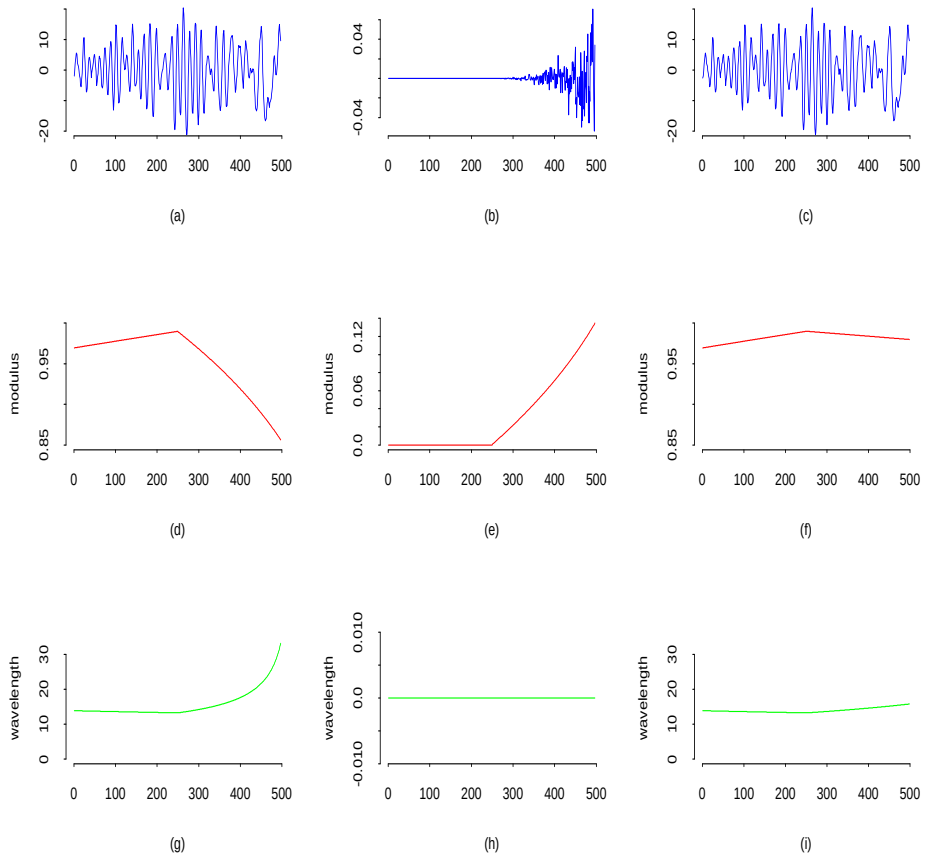

Figure 2: (a) Latent process $z_{t,1}^{(3)}$ in the decomposition of x_t based on the actual values of the TVAR(3) model parameters. (b) Latent process $y_{1,t}^{(3)}$ in the decomposition of the x_t based on the actual TVAR(3) model parameters. (c) Latent process $z_{i,1}^{(2)}$ based on a TVAR(2) process; this process also corresponds to the actual data series. (d) Trajectory of $r_{1,t}^{(3)}$. (e) Trajectory of $r_{2,t}^{(3)}$. (f) Trajectory of $r_{1,t}^{(2)}$. (g) Trajectory of $\lambda_{1,t}^{(3)}$. (h) Trajectory of $\lambda_{2,t}^{(3)}$. (i) Trajectory of $\lambda_{1,t}^{(2)}$.

and

$$\mathcal{M}_3 : x_t = \mathbf{F}'\mathbf{x}_t, \quad \mathbf{x}_t = \mathbf{G}_t^{(3)}\mathbf{x}_{t-1} + \boldsymbol{\omega}_t, \quad \text{for all} \quad t = 1, \ldots, 500,$$

with

$$\mathbf{G}_t^{(2)} = \begin{pmatrix} \phi_{1,t} & \phi_{t,2} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{G}_t^{(3)} = \begin{pmatrix} \phi_{1,t} & \phi_{t,2} & \phi_{t,3} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then, we obtain two decompositions for x_t, one based on model $\mathcal{M}_2$, and another one based on model $\mathcal{M}_3$,

$$x_t = z_{1,t}^{(2)} \qquad \text{and} \qquad x_t = z_{1,t}^{(3)} + y_{1,t}^{(3)},$$

where $z_{1,t}^{(2)}$ is a TVARMA(2,1) process with time-varying characteristic modulus $r_{1,t}^{(2)}$ and wavelength $\lambda_{1,t}^{(2)}$ (or frequency $2\pi/\lambda_{1,t}^{(2)}$). Figure 2 (c) shows the process $z_{1,t}^{(2)}$ (this process obviously corresponds to the actual data series x_t) and Figures 2 (f) and (i) display the trajectories in time of the chacteristic modulus and wavelength, $r_{1,t}^{(2)}$ and $\lambda_{1,t}^{(2)}$. Similarly, $z_{1,t}^{(3)}$ is dominated by a TVARMA(2,1) process with time-varying modulus and wavelength $r_{1,t}^{(3)}$ and $\lambda_{1,t}^{(3)}$ and $y_{1,t}$ is dominated by a TVAR(1) process with time-varying modulus $r_{2,t}^{(3)}$. Figures 2 (a), (b), (d), (e), (g) and (h) display the processes $z_{1,t}^{(3)}$, $y_{1,t}^{(3)}$ and the trajectories of $r_{1,t}^{(3)}$, $r_{2,t}^{(3)}$, $\lambda_{1,t}^{(3)}$ and $\lambda_{2,t}^{(3)}$ respectively. The decomposition in (9) is such that $z_{1,t} = z_{1,t}^{(2)}$ for $t = 1, \ldots, 250$ and $z_{1,t} = z_{2,t}^{(3)}$, $y_{1,t} = y_{1,t}^{(3)}$ for $t = 251, \ldots, 500$. In this particular case $x_{1,t}^{(2)} = x_{1,t}^{(3)}$ for $t = 1, \ldots, 250$.

In order to illustrate other aspects of the computation and interpretation of the latent processes in the decomposition, assume that a TVAR(p_t) was fitted to the data and that we want to compute the decomposition based on estimated values of the model parameters $\hat{\boldsymbol{\phi}}_t$ and $\hat{p}_t$. Furthermore, assume that the estimated values of the time-varying AR coefficients are exactly the true values, i.e. $\hat{\boldsymbol{\phi}}_t = \boldsymbol{\phi}_t = (\phi_{1,t}, \phi_{2,t}, \phi_{3,t})'$, but that $\hat{p}_t = 2$ for $t = 1, \ldots, 350$ and $\hat{p}_t = 3$ for $t = 351, \ldots, 500$. Then, based on the estimated values of the parameters, we would have the following decomposition,

$$x_t = \begin{cases} z_{1,t} & \text{for} \quad t = 1, \ldots, 350 \\ z_{1,t} + y_{1,t} & \text{for} \quad t = 351, \ldots, 500, \end{cases} \tag{10}$$

with $z_{1,t} = z_{1,t}^{(2)}$ for $t = 1, \ldots, 350$ and $z_{1,t} = z_{1,t}^{(3)}$, $y_{1,t} = y_{1,t}^{(3)}$ for $t = 351, \ldots, 500$. In other words, the decomposition based on the estimated model is such that $z_{1,t}$ is a process with a TVARMA$(2,1)$ structure with characteristic modulus $r_{1,t}^{(2)}$ and wavelength $\lambda_{1,t}^{(2)}$ for $t = 1, \ldots, 350$ and a process dominated by a TVARMA$(2,1)$ structure with characteristic modulus $r_{1,t}^{(3)}$ and wavelength $\lambda_{1,t}^{(3)}$ for $t = 351, \ldots, 500$.

4 Study of synthetic data

In this Section we present the analysis of a synthetic data set via TVAR models that consider model order uncertainty. Figure 3 (a) displays the synthetic data at the bottom and the

two main quasi-periodic components ordered in terms of decreasing wavelength from the bottom up, i.e. component (1) corresponds to the higher wavelength component. The latent processes shown in the graph were computed using the actual values of the $\boldsymbol{\phi}_t$ parameters for $t = 1, \ldots, 2000$. The first 500 data points were generated from a TVAR(2) while the last 1500 observations were generated from a TVAR(4). The trajectories in time of the characteristic wavelengths, moduli and amplitudes associated to the actual time-varying AR parameters are displayed in Figures 3 (b), (c) and (d) respectively. We analyse the series with various TVAR(p_t) models, with p_t taking values from $p_{\min} = 2$ up to $p_{\max}$ and $p_{\max} = 4, 5, 6, 7, 8$. Relatively high values of the discount factor for the AR coefficients, in the range $0.99 - 0.998$ are considered. The transition probabilities that describe the evolution of the model order in time are defined as follows,

$$P[p_t = i | p_{t-1} = j] = \begin{cases} q_{ii} & j = i \\ q_{i-1,i} & j = i-1, \quad i \leq p_{\max} \\ q_{i-2,i} & j = i-2, \quad i \leq p_{\max} \\ q_{i+1,i} & j = i+1, \quad i \geq p_{\min} \\ q_{i+2,i} & j = i+2, \quad i \geq p_{\min} \\ 0 & \text{otherwise,} \end{cases}$$

with high values of q_{ii}, usually in the range $0.9 - 0.99$. A discrete uniform prior $P(p_1 = i) = 1/(p_{\max} - p_{\min} + 1)$ for $i = p_{\min}, \ldots, p_{\max}$ is set on the model order while relatively diffuse and conjugate normal/inverse-gamma priors are considered for the AR coefficients and the innovations variance.

Figure 4 shows the estimated posterior mean, the estimated posterior median (solid lines) and 95% posterior bands (dotted lines) for the model order at each time t. These results were obtained using $p_{\min} = 2$, $p_{\max} = 8$, $\beta = 0.995$, and a transition probability matrix

$$Q = \begin{pmatrix} 0.97 & 0.015 & 0.015 & 0 & 0 & 0 & 0 \\ 0.011 & 0.97 & 0.011 & 0.008 & 0 & 0 & 0 \\ 0.005 & 0.01 & 0.97 & 0.01 & 0.005 & 0 & 0 \\ 0 & 0.005 & 0.01 & 0.97 & 0.01 & 0.005 & 0 \\ 0 & 0 & 0.005 & 0.01 & 0.97 & 0.01 & 0.005 \\ 0 & 0 & 0 & 0.008 & 0.011 & 0.97 & 0.011 \\ 0 & 0 & 0 & 0 & 0.015 & 0.015 & 0.97 \end{pmatrix},$$

and are based on 2000 samples of the posterior distribution taken from 8000 iterations of the Gibbs sapler after a burn-in of 3000 iterations. The graph shows that the model order order roughly oscillates between 2 and 3 for the first 500 observations, with a median $\hat{p}_t = 2$, in agreement with the actual value of the model order between $t = 1$ and $t = 500$. After $t = 500$ and up to approximately $t = 1250$, the model order takes values between 2 and 5, with a median $\hat{p}_t = 3$. From about $t = 1250$ until the end of the series, p_t oscillates between 3 and 5 with a median $\hat{p}_t = 4$ for most of the time interval. The posterior median for p_t, starts around $\hat{p}_t = 2$, and it increases up to roughly $\hat{p}_t^* = 4$ towards the end of the series.

Figures 5 (a) and (b) show the estimated latent processes, ordered by decreasing wavelength from the bottom up, and the trajectories of the characteristic wavelengths for a TVAR(4) model. The highest amplitude characteristic wavelength displays a similar trajectory in time to the trajectory of the actual characteristic wavelength up to roughly $t = 1000$.

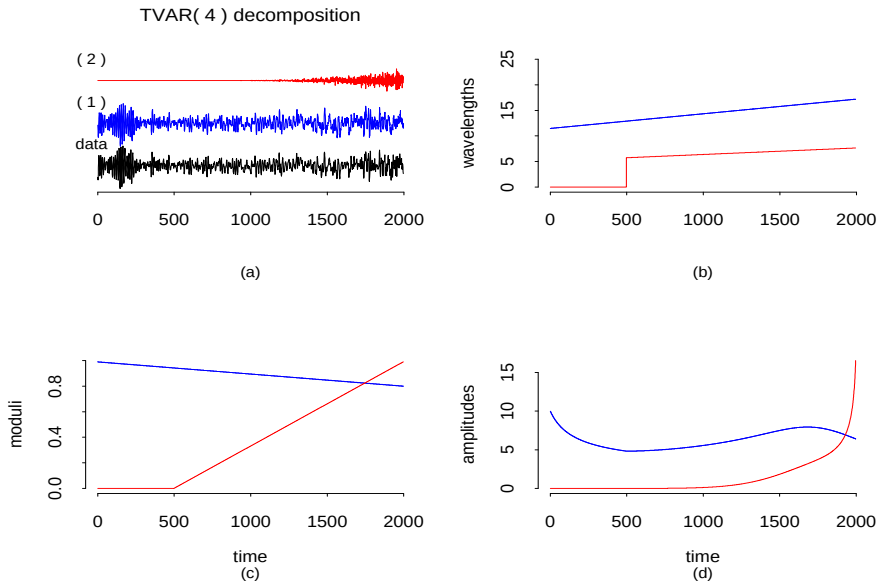

Figure 3: (a) Simulated process and the two main quasi-cyclical components ordered by wavelength. (b) Traces of values for wavelengths corresponding to the 2 main complex roots ordered by wavelength. (c) Traces of values for moduli corresponding to the 2 main complex roots ordered by wavelength. (d) Traces of values for amplitude corresponding to the 2 main complex roots ordered by wavelength.

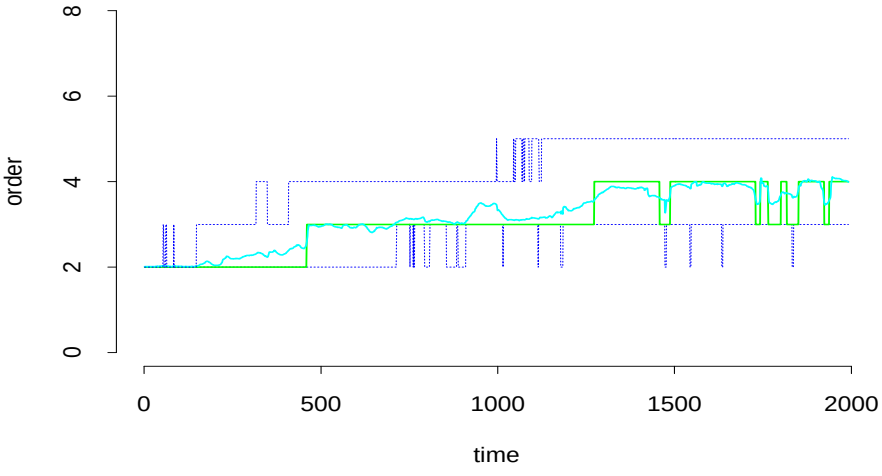

Figure 4: Simulated series: the estimated posterior mean and posterior median for model order at each time t with 95% posterior bands.

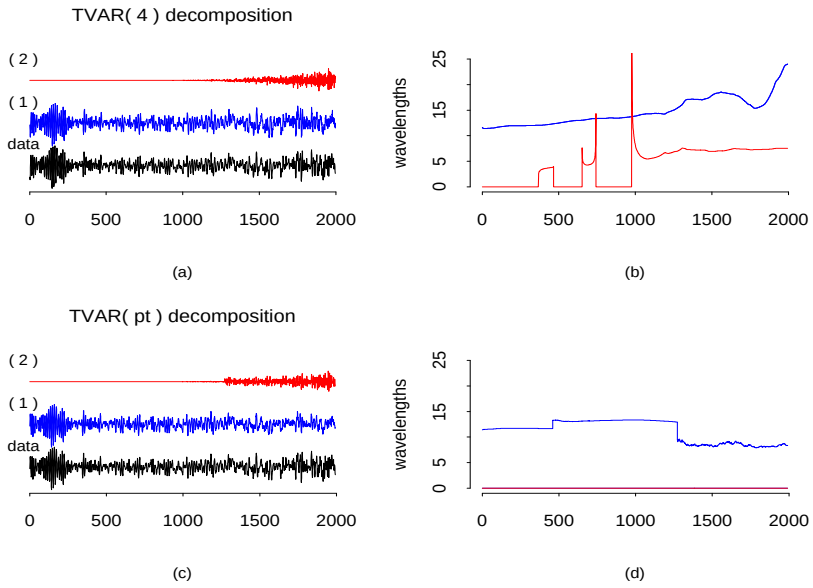

Figure 5: (a) Fixed order TVAR: simulated process and estimated quasi-cyclical components ordered by decreasing wavelength from the bottom up. (b) Fixed order TVAR: Traces of estimated values for wavelengths. (c) TVAR(p_t): simulated process and the main two estimated quasi-cyclical components ordered by wavelength. (d) TVAR(p_t): Traces of estimated values for wavelengths.

After $t = 1000$ the estimated highest characteristic wavelengths obtained with the TVAR(4) model have higher values than the actual characteristic wavelengths. The lowest amplitude characteristic wavelengths appear intermittently from $t = 1$ to roughly $t = 1000$, i.e. switches from complex to real characteristic components are experienced at this time period, and the trajectory of the estimated characteristic wavelength is similar to the trajectory of the actual characteristic wavelength (see Figure 3 (b)). Figures 5 (c) and (d) show the estimated latent processes in the decomposition, again ordered by decreasing amplitude from the bottom up, and the trajectories of the characteristic wavelengths for the TVAR(p_t) model described previously, based on the estimated posterior median for the model order at each time t, $\hat{p}_t$. The highest characteristic wavelength displays similar values to the actual highest characteristic wavelength from $t = 1$ up to aproximately $t = 1280$, and values similar to the actual lowest characteristic wavelength after $t = 1280$.

5 An application: analysis of an EEG trace via TVAR models with uncertainty on the model order

Consider again the EEG series displayed at the bottom of Figure 1. The latent components in the decomposition of the series shown in the graph, were computed using estimated posterior means for the AR coefficients and the innovations variance of a TVAR(12) model. In this Section, we model the series with a TVAR(p_t) whose model order at time t, p_t, can take

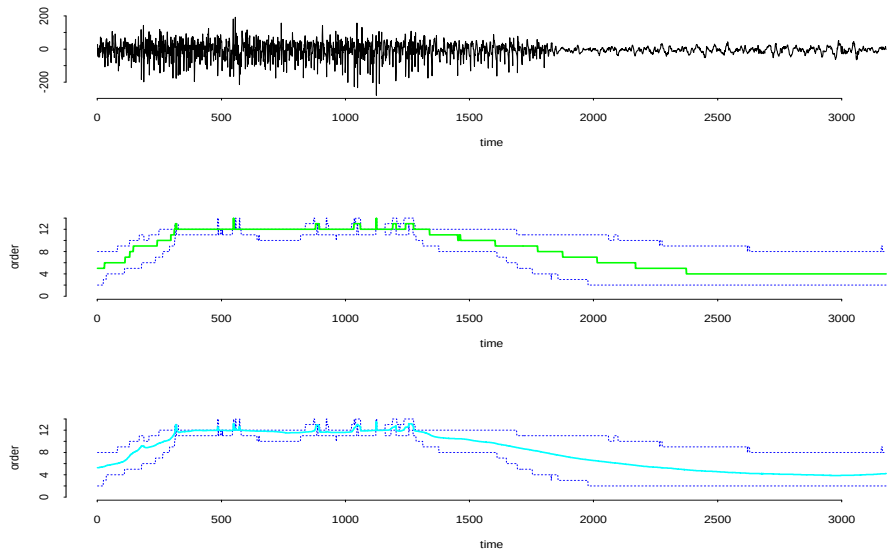

Figure 6: From the top down we have the EEG data, the estimated posterior mean for model order and each time t with 95% posterior bands and the estimated posterior median for model order at each time t with 95% posterior bands

values from $p_{\min} = 0$ up to $p_{\max} = 14$. Higher values of $p_{\min}$ and $p_{\max}$ were also considered, obtaining similar inferences in terms of the latent structure. The transition probability structure considered here is similar to the structure described in the previous Section for the synthetic data example. In particular, high values of q_{ii}, usually in the range 0.9 - 0.9999, are taken for $i = 0, \ldots, 14$, i.e. abrupt changes in the model order from time $t-1$ to time t are not permitted. Similarly, high values of the discount factors for the AR coefficients, in the range 0.99 - 0.999, are used. A discrete uniform prior $P(p_1 = i) = 1/15$ for $i = 0, \ldots, 14$ is set on the model order and relatively diffuse and conjugate normal/inverse-gamma priors are used for the AR coefficients and the innovations variance.

Figure 6 displays from the top down, the EEG data, the trajectory in time of the estimated posterior mean for model order with 95% posterior bands (center panel) and the trajectory of the estimated posterior median for model order with 95% posterior bands. The model was specified with a discount factor $\beta = 0.997$ for the AR coefficients, and transition probability matrix defined by, $q_{ii} = 0.99$ for all i; $q_{i,i+1} = q_{i,i-1} = 0.004$ and $q_{i,i+2} = q_{i,i-2} = 0.001$ for $2 \leq i \leq 12$; $q_{0,1} = q_{0,2} = q_{14,13} = q_{14,12} = 0.005$; $q_{1,0} = q_{1,2} = q_{13,14} = q_{13,12} = 0.004$ and $q_{1,3} = q_{13,11} = 0.002$. The instantaneous posterior means, medians and 95% posterior bands for model order displayed in Figure 6, are based on 5000 samples of the posterior distribution taken from 15000 iterations of the Gibbs sampler after a burn-in of 3000 iterations. The graphs show that the model order is higher in middle parts of the seizure - roughly from $t = 350$ until $t = 1300$ - than at the beginning and towards the end of the seizure. The posterior median for model order increases from $\hat{p}_t = 4$ up to $\hat{p}_t = 12$ at the beginning of the seizure and it decreases from $\hat{p}_t = 12$, at approximately $t = 1300$, to $\hat{p}_t = 4$ at approximately $t = 2400$. The patterns observed in the trajectories of the estimated model order posterior

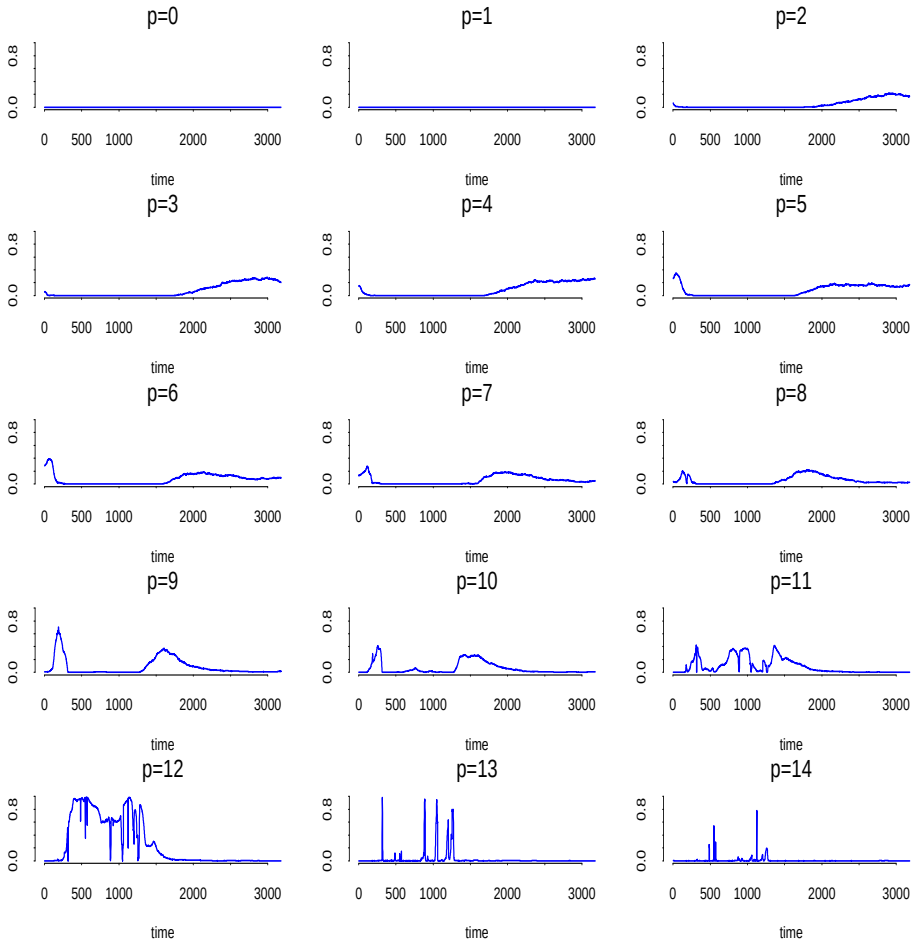

Figure 7: Estimated posterior probabilities for model order at each time t

mean and median over time, indicate that the complexity of the latent structure is higher at middle parts of the seizure than at the beginning and once the seizure is over. Furthermore, the complexity of the data structure measured as a function of model order starts to decrease prior to the seizure dissipation, just before $t = 1500$.

Figure 7 displays estimated posterior probabilities for model order at each time t. Model orders $11, 12, 13$ and 14 have higher posterior probabilities at the middle parts of the seizure, while lower model orders are preferred for starting and late periods of the series.

Figure 8 shows the decomposition of the EEG series based on estimated posterior means of the AR coefficients and posterior medians for model order at each time t. From the bottom up, the graph displays the data followed by estimated latent processes in order of decreasing amplitude. Component (1) has the highest amplitude and it is related to a pair of complex conjugate characteristic roots from approximately $t = 1$ to $t = 2200$. This component is also associated with the highest characteristic wavelength for almost all t in the time interval that spans from $t = 1$ to $t = 2200$. Trajectories of the wavelengths related to the quasi-periodic processes, based on estimated posterior means of the AR coefficients truncated at the posterior median for model order, appear in Figure 9 (a). Component (1) has a characteristic

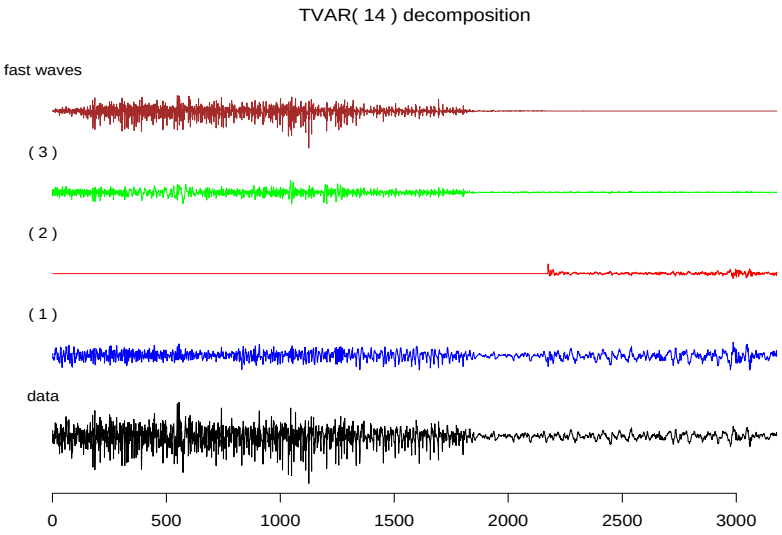

Figure 8: Data and estimated components in the decomposition of an EEG series based on a TVAR(p_t). From the bottom up, the graph displays the series followed by estimated components in order of decreasing amplitude.

wavelength of about 10 to 8 from $t = 1$ to approximately $t = 850$ that then increases from 12, right after $t = 850$, up to roughly 20 at $t = 2200$ (see Figure 9 (a)). Equivalently, in terms of frequency in the original sampling scale (see discussion on West *et al.* 1999 about subsampling of the original data series), we have a dominant frequency of about 4.35 to 5.33 Hz during the first 850 observations and a frequency that decreases from approximately 3.56 Hz right after $t = 850$ to 2.13 Hz at $t = 2200$. These findings are consistent with the results presented in West *et al.* (1999) obtained with a fixed order model, but now we are taking model order uncertainty into account. After $t = 2200$ the latent process (1) is associated to a real characteristic root with relatively high time-varying modulus. Component (2) in the decomposition corresponds to a real characteristic root and component (3) is a quasi-periodic process related to the second highest wavelength. Component (4), denoted in the Figure as "fast waves", is the sum of all remaining low amplitude processes that correspond to high frequency and noise components.

Figure 9 (a) sketches the frequency components present in the data taking into account uncertainty on the number of such components over time. However, these wavelength trajectories display jumps and artifacts that are simply the result of using the first $\hat{p}_t$ elements of the estimated $\boldsymbol{\phi}_t$ vector at each time t. A graph of the approximate wavelengths trajectories can be obtained by considering the maximum order model. Figure 9 (b) displays the wavelengths trajectories of the first four quasi-periodic components computed at the estimated posterior mean of $\boldsymbol{\phi}_t$ with $p_t = p_{\max} = 14$ for all t. The highest wavelength also corresponds to the highest amplitude component in the decomposition for $t = 1$ to roughly $t = 2100$. This is indicated in the Figure by the dark segment on the trajectory. After $t = 2100$ this

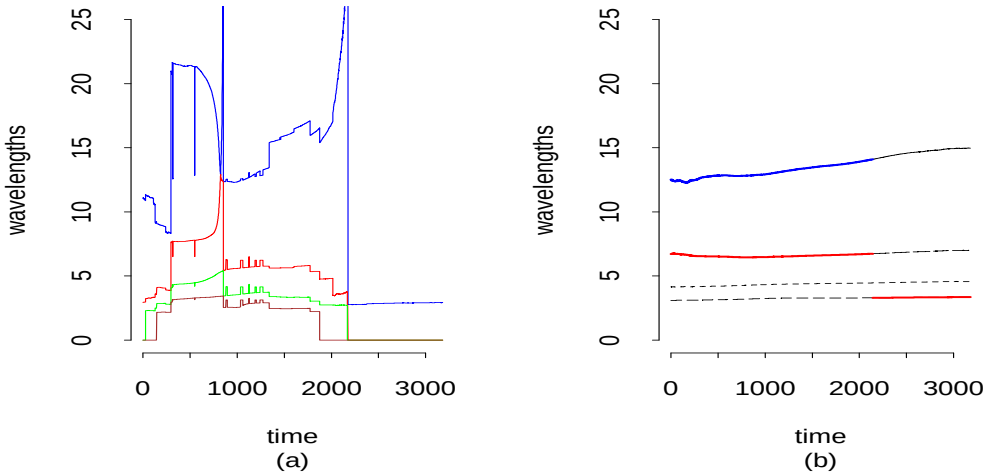

Figure 9: (a) Traces of wavelengths corresponding to the first four quasi-cyclical components computed at instantaneous model order medians. (b) Traces of wavelengths corresponding to the first four quasi-cyclical components ordered by wavelength computed at the maximum order for all t.

quasi-periodic process dissipates being a real component the one with the highest amplitude. The second highest and the lowest wavelengths in the Figure correspond to the second highest amplitude components. Again, the darker segments indicate the portions in which each component is dominant.

As mentioned before, inferences obtained by including model order uncertainty as a modelling component are consistent with the results obtained in previous analyses of the same series (West *et al.*, 1999; Krystal *et al.*, 1999). However, additional insight in terms of complexity of the latent structure is gained using a TVAR(p_t) model approach. For instance, the decrease in model order observed just prior to the end of the seizure might be relevant in connection with assessing the clinical efficacy of the treatment.

6 Remarks, conclusions and future directions

In this paper we presented a TVAR model that fully incorporates uncertainty on model order with a first order random walk. The model can be decomposed in terms of latent processes, perhaps of quasi-cyclical nature, which generalises the decomposition results for fixed order TVAR presented in West *et al.* (1999). Based on MCMC methods, we explored the performance of the model for a synthetic series and an EEG series. In both cases, we studied the impact of discount factor selection and specification of transition probabilities for model order in terms of inference on latent structure and on the model order p_t. Since model fitting relies on simulation, the TVAR(p_t) requires a higher computational demand than a fixed order TVAR. Additionally, and as seen in the EEG series analysis of Section 5, the TVAR(p_t) may need more thoughtful interpretations for the estimated latent components and trajectories

of wavelengths, moduli and amplitudes corresponding to the different characteristic roots of the process.

In connection to other mixture models, the TVAR(p_t) can be seen as a multi-process, class II mixture model (West and Harrison (1997), pp. 444-445) where each of the defining models is determined by the possible values of p_t. The limitation of this approach, is that it introduces a mixture of DLMs where each component has state vectors of different dimension. In consequence, expressions for the posterior distributions of the state vectors and model order are not available in closed form and the multi-process requires MCMC methods as presented in this paper. In contrast, an alternative for the TVAR(p_t) arises if we consider a mixture of TVARs, each with a fixed order, and assuming that one of the models holds for all time t. Such a mixture introduces a multi-process class I model with posterior distributions available in closed form. Although, this multi-process model can be handled easily, it lacks flexibility compared to the TVAR(p_t) since it does not allow instantaneous transitions on model order. Comparisons on these lines are part of current research.

In a further extension, we are considering TVAR models of time varying order but with priors on the characteristic polynomial roots following the work by Huerta and West (1999) for standard autoregressions. At each time t, the model controls for the number of complex pairs and real reciprocal roots and consequently defines model order uncertainty on p_t. Implementation issues have to explore different evolutions on the reciprocal root parameterisation and consideration of both *particle filter* methods and *Forward Filtering Backward Simulation* algorithms. We expect that this new approach will produce very interesting results and lead to challenging methodological issues.

Appendix: posterior sampling algorithm

We describe the details to simulate samples from the full posterior distribution of $(\boldsymbol{\phi}_t, \sigma^2, p_t)$ given D_n for $t = 1, \ldots, n$ with the model specifications described in Section 3. Define $\Phi = \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n\}$, with $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p_{\max}})'$, $\mathbf{X} = \{x_1, \ldots, x_n\}$, and $\mathbf{P} = \{p_1, \ldots, p_n\}$. We follow a Gibbs sampling format defined in two stages.

- Sampling from the full conditional distribution of Φ and σ^2.

  This can be done by sampling Φ from $p(\Phi|\mathbf{X}, \mathbf{P}, \sigma^2)$ and sampling σ^2 from $p(\sigma^2|\mathbf{X}, \Phi, \mathbf{P})$. Conditional on model order, we have the DLM structure

$$\begin{aligned} x_t &= \mathbf{F}_t'\boldsymbol{\phi}_t + \epsilon_t \\ \boldsymbol{\phi}_t &= \boldsymbol{\phi}_{t-1} + \boldsymbol{\xi}_t \end{aligned}$$

  with $\mathbf{F}_t = (\mathit{1}_{1,t}x_{t-1}, \ldots, \mathit{1}_{p_{\max},t}x_{t-p_{\max}})'$. Efficient generation via Forward Filtering-Backward Sampling (Carter and Kohn, 1994) can be performed conditional on $\mathbf{X}, \mathbf{P}$, and σ^2. The set of system covariance matrices $\{\mathbf{U}_t; t = 1, \ldots, n\}$ is assumed known or specified by a single discount factor following standard DLM theory (West and Harrison, 1997).

  In order to sample from $p(\sigma^2|\mathbf{X}, \Phi, \mathbf{P})$, we compute

$$e = \sum_{t=1}^{n}(x_t - \sum_{j=1}^{p_{\max}} \mathit{1}_{t,j}\phi_{t,j}x_{t-j}),$$

  and sample a distribution proportional to $p(\sigma^2)(\sigma^2)^{-(\alpha+1)}\exp\{-\beta/\sigma^2\}$ with $\alpha = n/2-1$ and $\beta = e^2/2$. We are assuming a reference prior for σ^2, $p(\sigma^2) \propto 1/\sigma^2$. Inverse gamma priors are conditionally conjugate.

- Sampling from the full conditional distribution of $\mathbf{P} = \{p_1, p_2, \ldots, p_n\}$.

  Let D_t be the information up to time t, i.e., $D_t = \{D_0, \mathbf{X}_t, \Phi_t, \sigma^2\}$ with $\mathbf{X}_t = \{x_1, \ldots, x_t\}$, $\Phi_t = \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_t\}$. $P[p_t = i|p_{t-1} = j]$ defines the transition probabilities for model order between times $t-1$ and t. For $t = 1, 2, 3, \ldots, n$ and $i = p_{\min}, \ldots, p_{\max}$, we recognise that if $p_t = i$ then $\mathit{1}_t$ is a vector whose first i entrances are ones and the remaining are zeroes. The filtering part of the algorithm requires that we compute and save the probabilities $P[p_t = i|D_t]$ and $P[p_t = i|D_{t-1}]$, for $i = p_{\min}, \ldots, p_{\max}$ and $t = 1, \ldots, n$. By Bayes theorem,

$$P[p_t = i|D_t] \propto f(x_t|\phi_t, \sigma^2, \mathit{1}_t)P[p_t = i|D_{t-1}],$$

  where $f(x_t|\phi_t, \sigma^2, \mathit{1}_t)$ is the likelihood function for the observation x_t that is easily obtained with the model definition of the TVAR. Additionally,

$$P[p_t = i|D_{t-1}] = \sum_{j=p_{\min}}^{p_{\max}} P[p_t = i|p_{t-1} = j]P[p_{t-1} = j|D_{t-1}]$$

with $P[p_{t-1} = j|D_{t-1}]$ the posterior probability evaluated at time $t-1$. For $t = 1$, $P[p_1 = j|D_0]$ is a discrete uniform distribution.

Now, to simulate each entrance of $\mathbf{P}$, at each time t, we apply the Backward Simulation Sampling algorithm step for discrete random variables presented in Carter and Kohn (1994). First, we generate a value p_n^* from the distribution $P[p_n = i|D_n]$. Then, for $t = n-1, n-2, \ldots, 1$, we compute $P[p_t = i|p_{t+1} = p_{t+1}^*, D_t]$ with the expression

$$P[p_t = i|p_{t+1} = p_{t+1}^*, D_t] = \frac{P[p_{t+1} = p_{t+1}^*|p_t = i]P[p_t = i|D_t]}{P[p_{t+1} = p_{t+1}^*|D_t]}$$

where p_{t+1}^* is a generated value of the distribution $P[p_{t+1} = i|p_{t+2} = p_{t+2}^*, D_t]$. We sample a value p_t^* from $P[p_t = i|p_{t+1} = p_{t+1}^*, D_t]$ and continue until $t = 1$. The values $p_1^*, p_2^*, \ldots, p_n^*$ are a sample from the conditional posterior distribution of $\mathbf{P}$.

Bibliography

Andrieu, C., Frietas, N. De and Doucet, A. (1999) Sequential MCMC for Bayesian model selection. In *Proc. Work. IEEE HOS'99.*

Barbieri, M.M. and O'Hagan, A. (1997) A reversible jump MCMC sampler for Bayesian analysis of ARMA time series. Technical Report. Department of Statistics Universita "La Sapienza".

Barnett, G., Kohn, R. and Sheather, S. (1996) Bayesian estimation of an autoregressive model using Markov Chain Monte Carlo. *Journal of Econometrics*, **74,** 237–254.

Carter, C.K. and Kohn, R. (1994) Gibbs sampling for state space models. *Biometrika*, **81,** 541–53.

Frühwirth-Schnatter, Sylvia (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15,** 183–102.

Gersch, W. (1987) Non-Stationary multichannel time series analysis. In *EEG Handbook, Revised Series* (ed. A. Gevins), vol. 1, pp. 261–96. New York: Academic Press.

Green, P.J. (1995) Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82,** 711–32.

Harrison, P.J. and Stevens, C.F. (1976) Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society-Series* B, **38,** 205–247.

Huerta, G. and West, M. (1999) Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society-Series* B, **61,** 1–19.

Kitagawa, G. and Gersch, W. (1996) *Smoothness Priors Analysis of Time Series,* Lecture Notes in Statistics, vol. 116. New-York: Springer-Verlag.

Krystal, A., West, M., Prado, R., Greenside, H., Zoldi, S. and Weiner, R. (1999) The EEG effects of ECT: Implications for rTMS. Technical Report 99-02. Institute of Statistics and Decision Sciences, Duke University, USA.

Pitt, M. and Shephard, N. (1999) Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, **94,** 590–599.

Prado, R. (1998) Latent structure in non-stationary time series. Ph.D. Thesis. Duke University, Durham, NC.

Prado, R. and West, M. (1997) Exploratory modelling of multiple non-stationary time series: Latent process structure and decompositions. In *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions, New York* (eds T.G. Greogoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren and R.D. Wolfinger), Lecture Notes in Statistics 12. Springer Verlag.

Troughton, P.T. and Godsill, S.J. (1997) Bayesian model selection for time series using Markov Chain Monte Carlo. Technical Report. Signal Processing and Communications Laboratory. Department of Engineering, University of Cambridge.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Linear Models (2nd Edn.)*. New York: Springer-Verlag.

West, M., Prado, R. and Krystal, A. (1999) Latent structure in non-stationary time series with application in studies of EEG traces. *Journal of the American Statistical Association*, **94,** 375–387.