Bayesian time-varying autoregressions: Theory, methods and applications

Raquel Prado[*], Gabriel Huerta[†] and Mike West[‡]

Abstract

We review the class of time-varying autoregressive (TVAR) models and a range of related recent developments of Bayesian time series modelling. Beginning with TVAR models in a Bayesian dynamic linear modelling framework, we review aspects of latent structure analysis, including time-domain decomposition methods that provide inferences on the structure underlying non-stationary time series, and that are now central tools in the time series analyst's toolkit. Recent model extensions that deal with model order uncertainty, and are enabled using efficient Markov Chain Monte Carlo simulation methods, are discussed, as are novel approaches to sequential filtering and smoothing using particulate filtering methods. We emphasize the relevance of TVAR modelling in a range of applied contexts, including biomedical signal processing and communications, and highlight some of the central developments via examples arising in studies of multiple electroencephalographic (EEG) traces in neurophysiology. We conclude with comments about current research frontiers.

Keywords: Bayesian analysis; Dynamic linear models; Model uncertainty; MCMC simulation; Time series decompositions; TVAR models

[*]Departamento de Cómputo Científico y Estadística and CESMa, Universidad Simón Bolívar, Apartado 89000, Caracas, Venezuela. *E-mail:* raquel@cesma.usb.ve.

[†]Departamento de Probabilidad y Estadística, Centro de Investigación en Matemáticas, Apartado 402, Guanajuato, Gto., México. *E-mail: ghuerta@cimat.mx*

[‡]Institute of Statistics and Decision Sciences, Duke University, USA. *E-mail: mw@isds.duke.edu*

1 Introduction

Time varying autoregressive (TVAR) models have provided useful empirical representations of non-stationary time series in various applied fields. Since the early 1980s, W. Gersch, G. Kitagawa and coauthors, have demonstrated the flexibility of high-order TVAR models to describe changes in the stochastic structure of series with marked and time-varying periodicities. These authors have focussed on issues of changes in the instantaneous power spectra implied by the TVAR models, and questions of feedback and time lags in multiple time series in the analysis of seismic and electroencephalographic data (e.g., Kitagawa 1983, Gersch 1987, Kitagawa and Gersch 1996).

The late 1990s saw a range of novel developments with TVAR and related models that led to new methods of time series decomposition and analysis with broad applicability. Core results underlying this development were introduced in West (1997) with a focus on exploring latent quasi-periodic components in standard AR models. These methods provide useful insights into the latent structure of observed time series that often have physical interpretations. In connection with these developments, Huerta and West (1999) proposed a novel class of priors for model parameters and model order, such priors characterising the number and structure of latent underlying components in AR processes. The basic decomposition ideas were developed in important classes of TVAR models in West *et al.* (1999), Prado and West (1997) and Prado (1998), which cover a range of methodological and practical areas, including issues of model fitting and resulting inferences for component structure underlying non-stationary time series. These works have proven the great utility of such decomposition methods in applied areas, with a major highlight being in the studies of electroencephalogram (EEG) recordings on human subjects (Krystal *et al.* 1999).

Standard TVAR models and decompositions are easily implemented using sequential updating and filtering/smoothing algorithms for dynamic linear models (West and Harrison, 1997). However, due to the developments of Markov Chain Monte Carlo (MCMC) methods, efficient algorithms are available for implementation of more flexible and sophisticated time series models (e.g. Carter and Kohn 1994, de Jong and Shephard 1995). In particular, Prado and Huerta (1999) deal with model order uncertainty within the TVAR modelling framework via MCMC methods. This work focuses on issues of how inference on latent structure is affected when uncertainty on model order is considered. In connection with the use of simulation methods for model implementation, a key area of current research interest in time series is focussed on adapting simulation methods to a sequential analysis context via particle filtering. Good illustrations of the methodology in this area are Pitt and Shephard (1999) and Liu and West (2000), and the set of papers in Doucet *et al.* (2000) that provide a broad overview of the field. Additionally, methods for performing smoothing in non-linear non-Gaussian dynamic models when analysis is based on sequential simulation methods, and with central applications to TVAR models in communications applications, appear in Godsill *et al.* (2000a, b) and Doucet *et al.* (2000).

The purpose of this paper is to review these main areas of recent development in theory, computation and application of TVAR models and related methods. In Section 2 we introduce canonical TVAR models and describe the decomposition structure for the univariate and multivariate cases. In Section 3 we consider interesting extensions involving time-varying relationships between multiple series. Section 4 discusses developments of TVAR methodology to address model order uncertainty. Various analyses of non-stationary EEG series illustrate the methodology and implied practical aspects of the theory described in Sections 2, 3 and 4. Section 5 highlights other recent developments, especially in sequential computation for on-line filtering and smoothing in TVAR models, with additional comments on current research frontiers and future directions.

2 TVAR models and decomposition theory

2.1 Model specification

A univariate time series x_t, follows a time-varying autoregressive model of fixed order p, or TVAR(p), if

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

where $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p})'$ is the time-varying vector of coefficients and ϵ_t are zero-mean independent innovations assumed Gaussian with possibly time-varying variances σ_t^2. No explicit stationarity constraints are imposed on the AR parameters at each time t. However, if such parameters lie in the stationary region, as it is the case in many applications, 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 structure for $\boldsymbol{\phi}_t$ and σ_t^2. Here we assume that the AR parameters evolve according to a random walk, however, more elaborate evolutions may be specified, as in Godsill *et al.* (2000b), for example. A random walk evolution structure on $\boldsymbol{\phi}_t$, that is, $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1} + \boldsymbol{\xi}_t$ with $\boldsymbol{\xi}_t \sim N(\mathbf{0}, \mathbf{W}_t)$, provides adaptation to the changing structure of the series over time without anticipating specific directions of changes (West and Harrison 1997, chapter 3). The variation in time of $\boldsymbol{\phi}_t$ is controlled via standard discount factor methods (West and Harrison 1997). A single discount factor $\beta \in (0, 1]$ leads to values of each $\mathbf{W}_t$ such that low values of β imply high variability of the $\boldsymbol{\phi}_t$ sequence, while high values, in the range 0.9-0.999, are typically considered in practice. 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 $\boldsymbol{\xi}_t$, and with $\eta_t \sim Beta(a_t, b_t)$. The parameters a_t and b_t are defined at each t by a discount factor $\delta \in (0, 1]$ analogous to β. Suitable values for the discount factors and the model order p may be obtained via marginal likelihoods, mean square errors or mean absolute deviations as discussed in West *et al.* (1999) and Prado (1998). Given p, β and δ, the TVAR model can be framed as a dynamic linear regression with model coefficients $\boldsymbol{\phi}_t$. The equations for sequential updating and retrospective filtering/smoothing of general dynamic linear models (West and Harrison 1997, chapters 4 and 10) lead to posterior inferences on $\boldsymbol{\phi}_t$ and σ_t^2.

2.2 Time series decompositions

In recent years, applied interests in a variety of fields have stimulated Bayesian time series research focussed on latent time-frequency structure analysis. In particular, decomposition methods and related theory and analysis for TVAR models have been recently developed in

West *et al.* (1999), Prado and West (1997) and Prado (1998). This section reviews the key points of these developments.

Consider a general dynamic linear model (DLM) in which the observed scalar time series y_t, $t = 1, 2, \ldots,$ is modelled as

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

where x_t is the latent signal, ν_t is an observation error, $\boldsymbol{\theta}_t$ is a $d \times 1$ state vector, $\mathbf{F}$ is a column d-vector, $\mathbf{G}_t$ is the $d \times d$ state evolution matrix and $\boldsymbol{\omega}_t$ is the d-vector of state innovations. The state evolution matrix $\mathbf{G}_t$ may depend on uncertain, possibly time-varying parameters. The noise terms ν_t and $\boldsymbol{\omega}_t$ are mutually uncorrelated white noise, though more complex structures may be considered to handle measurement error and outlier components.

The decomposition results summarised below are based on standard theory of model structure and similar models (West and Harrison 1997, chapter 5). Assume that at each time t, the state matrix $\mathbf{G}_t$ in (2) has exactly d different eigenvalues, some of which could be complex and in such case they will appear in conjugate pairs. The number of complex and real eigenvalues may vary over time but, for the sake of simplicity, assume that at each time t there are c pairs of complex eigenvalues denoted by $r_{t,j} \exp(\pm i\omega_{t,j})$ for $j = 1, \ldots, c$, and $r = d - 2c$ real eigenvalues denoted by $r_{t,j}$ for $j = 2c+1, \ldots, d$. Then $\mathbf{G}_t = \mathbf{E}_t\mathbf{A}_t\mathbf{E}_t^{-1}$ where $\mathbf{A}_t$ is the diagonal matrix of eigenvalues in arbitrary but fixed order, and $\mathbf{E}_t$ is a $d \times d$ matrix whose columns correspond to the eigenvectors appearing in the order given by the eigenvalues. For each t define $\mathbf{H}_t = \mathrm{diag}(\mathbf{E}_t'\mathbf{F})\mathbf{E}_t^{-1}$ and linearly transform the state parameter vector $\boldsymbol{\theta}_t$ to $\boldsymbol{\gamma}_t = \mathbf{H}_t\boldsymbol{\theta}_t$. Rewriting (2) we have

$$y_t = x_t + \nu_t, \;\; x_t = \mathbf{1}'\boldsymbol{\gamma}_t, \;\; \boldsymbol{\gamma}_t = \mathbf{A}_t\mathbf{K}_t\boldsymbol{\gamma}_{t-1} + \boldsymbol{\delta}_t, \tag{3}$$

where $\mathbf{1} = (1, 1, \ldots, 1)'$, $\boldsymbol{\delta}_t = \mathbf{H}_t\boldsymbol{\omega}_t$ is a zero-mean normal innovation with a structured and singular variance matrix and $\mathbf{K}_t = \mathbf{H}_t\mathbf{H}_{t-1}^{-1}$. Then (3) implies that x_t is the sum of the individual components of $\boldsymbol{\gamma}_t = (\gamma_{t,1}, \ldots, \gamma_{t,d})'$. The final r elements of $\boldsymbol{\gamma}_t$ are real-valued processes, corresponding to the real eigenvalues $r_{t,j}$. Rename these processes $y_{t,j}$. The initial $2c$ elements of $\boldsymbol{\gamma}_t$ appear in complex pairs and therefore $z_{t,j} = \gamma_{t,2j-1} + \gamma_{t,2j}$ is also a real-valued process. The basic decomposition result for the class of models that can be expressed in the form (2) is simply

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

Given known, estimated or simulated values of $\mathbf{F}$, $\mathbf{G}_t$ and $\boldsymbol{\theta}_t$ at each time t, the processes $z_{t,j}$ and $y_{t,j}$ can be evaluated over time by computing the eigenstructure of the evolution matrix $\mathbf{G}_t$ and the linear transformations described above. We now explore the structure of the processes $z_{t,j}$ and $y_{t,j}$ for the class of TVAR and vector AR models.

2.2.1 Decompositions for TVAR models

The TVAR model (1) can be expressed in a DLM or state-space model form (2), with $d = p$, $\nu_t = 0$, $\mathbf{F} = (1, 0, \ldots, 0)'$, $\boldsymbol{\theta}_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})'$, $\boldsymbol{\omega}_t = \epsilon_t \mathbf{F}$ and

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

The eigenvalues of $\mathbf{G}_t$ are the reciprocals roots of the instantaneous autoregressive characteristic equation at time t, $\boldsymbol{\phi}_t(u) = (1 - \phi_{t,1}u - \ldots - \phi_{t,p}u^p)$. In particular, for the standard AR(p) process $\mathbf{G}_t = \mathbf{G}$, therefore $r_{t,j} = r_j$ for $j = 1, \ldots, p$ and $\omega_{t,j} = \omega_j$ for $j = 1, \ldots, c$. Furthermore, it is easy to see that 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 constant characteristic frequency ω_j (or wavelength $2\pi/\omega_j$) and modulus r_j (West *et al.* 1999). In this case and assuming stationarity, the decomposition is essentially that derived by the partial fractions decomposition of an AR(p) process (Box and Jenkins 1976).

In the general TVAR case each $y_{t,j}$ is dominated by a TVAR(1) with time-varying AR parameter $r_{t,j}$, while each $z_{t,j}$ is dominated by a TVARMA(2,1) with time-varying characteristic frequency $\omega_{t,j}$ and modulus $r_{t,j}$. The stochastic structure of $y_{t,j}$ and $z_{t,j}$ is not exactly represented by TVAR(1) and TVARMA(2,1) components, since there is an element of linear mixing of the latent processes through time. However, such mixing components are negligible in most practical applications. The main point for this result is that the matrix $\mathbf{K}_t$ in equation (3) for a TVAR model, will generally not be equal to the identity, a key feature for the special latent structure of a constant AR model. $\mathbf{K}_t$ will be close to the identity when $\mathbf{G}_t$ and $\mathbf{G}_{t-1}$ are similar, i.e. in cases when $\boldsymbol{\phi}_t$ changes slowly in time. When the $\mathbf{K}_t$ matrices are very close to identity matrices the component processes in the decomposition have a structure almost completely dominated by TVAR(1) and TVARMA(2,1) processes. A detailed discussion on this appears in Prado (1998) and West *et al.* (1999).

2.2.2 Multivariate decompositions

The decompositions presented above have a direct extension to the multivariate framework. Details of the results summarised here can be found in Prado (1998).

Consider an m-dimensional time series process $\mathbf{y}_t = (y_{1,t}, \ldots, y_{m,t})'$ modelled using a multivariate DLM (West and Harrison 1997, chapter 16)

$$\mathbf{y}_t = \mathbf{x}_t + \boldsymbol{\nu}_t, \quad \mathbf{x}_t = \mathbf{F}'\boldsymbol{\theta}_t, \quad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{5}$$

where $\mathbf{x}_t$ is the underlying m-dimensional signal, $\boldsymbol{\nu}_t$ is an m-dimensional vector of observation errors, $\mathbf{F}$ is a $d \times m$ matrix of constants, $\boldsymbol{\theta}_t$ is the d-dimensional state vector, $\mathbf{G}_t$ is the $d \times d$ state evolution matrix and $\boldsymbol{\omega}_t$ is a d-vector of state innovations. The noise terms $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are zero mean innovations, assumed independent and mutually independent with variance-covariance matrices $\mathbf{V}_t$ and $\mathbf{W}_t$ respectively. As in the univariate case, assume that $\mathbf{G}_t$ has exactly d distinct eigenvalues at each time t, with c pairs of complex eigenvalues $r_{t,j}\exp(\pm\omega_{t,j})$

for $j = 1, \ldots, c$, and $r = d - 2c$ real eigenvalues $r_{t,j}$ for $j = 2c+1, \ldots, d$. Define m matrices $\mathbf{H}_{i,t} = diag(\mathbf{E}_t'\mathbf{F}_i)\mathbf{E}_t^{-1}$, with $\mathbf{F}_i$ the i-th column of the matrix $\mathbf{F}$, and consider m new state vectors $\boldsymbol{\gamma}_{i,t} = \mathbf{H}_{i,t}\boldsymbol{\theta}_t$ and m new state innovation vectors $\boldsymbol{\delta}_{i,t} = \mathbf{H}_{i,t}\boldsymbol{\omega}_t$ for $i = 1, \ldots, m$. Then, we obtain m models, $\mathcal{M}_i$, one for each of the scalar components of $\mathbf{x}_t$

$$\mathcal{M}_i : \quad \begin{array}{rcl} x_{i,t} & = & \mathbf{1}'\boldsymbol{\gamma}_{i,t} \\ \gamma_{i,t} & = & \mathbf{A}_t\mathbf{K}_{i,t}\gamma_{i,t-1} \quad + \quad \boldsymbol{\delta}_{i,t}, \end{array} \tag{6}$$

with $\mathbf{K}_{i,t} = \mathbf{H}_{i,t}\mathbf{H}_{i,t-1}^{-1}$. Therefore, using the decomposition results for univariate time series, $x_{i,t}$ can be expressed as a sum of $c + r$ components

$$x_{i,t} = \sum_{j=1}^{c} z_{i,t,j} + \sum_{j=2c+1}^{d} y_{i,t,j}, \tag{7}$$

where $z_{i,t,j}$ are real-valued processes related to the pairs of complex eigenvalues for $j = 1, \ldots, c$, and $y_{i,t,j}$ are real processes related to the real eigenvalues for $j = 2c+1, \ldots, d$.

In particular, if $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,m})'$ follows an m-dimensional vector autoregressive model, VAR(p)

$$\mathbf{x}_t = \sum_{j=1}^{p} \boldsymbol{\Phi}_j \mathbf{x}_{t-j} + \boldsymbol{\epsilon}_t \tag{8}$$

where $\boldsymbol{\Phi}_j$ are $m \times m$ matrices of AR coefficients and $\boldsymbol{\epsilon}_t$ are m-dimensional zero mean innovation vectors with covariance matrix $\mathbf{V}$, it is easy to see that each $x_{t,i}$ series has a decomposition as the sum of several AR(1) and ARMA(2, 1) processes. The $z_{i,t,j}$ processes in the decomposition are quasi-periodic, following ARMA(2,1) models with characteristic frequencies and moduli ω_j and r_j for $j = 1, \ldots, c$, while the $y_{i,t,j}$ processes have an AR(1) structure with AR coefficients r_j for $j = 1, \ldots, mp$. Thus, each univariate element $x_{i,t}$ has a decomposition whose latent ARMA(2,1) and AR(1) processes are characterised by the *same* frequencies and moduli across i, though the phases and amplitudes associated with these components are specific to each univariate element $x_{i,t}$.

2.3 Latent structure in multiple EEG traces

Various applied studies have been generated recently in the area of EEG analysis (Prado and West 1997, Krystal *et al.* 1999, Prado *et al.* 2000). We illustrate the use of TVAR models and decompositions in the analysis of an EEG trace from a dataset previously studied in Prado and West (1997) and Prado *et al.* (2000). The EEG series analysed here is one of 19 traces recorded at different scalp locations during a patient seizure, elicited by electroconvulsive therapy (ECT) as antidepressant treatment. Details of the analyses of the full dataset via TVAR models and related decomposition theory can be found in Prado and West (1997), and further developments, including estimation of time-varying lag/lead structure among the 19 channels, appear in Prado *et al.* (2000). The purpose of these studies is to explore differences and commonalities in latent structure across the 19 traces in order to characterise aspects of the spatio-temporal dynamics that improve the understanding of the physiology driving the antidepressant effectiveness of ECT.

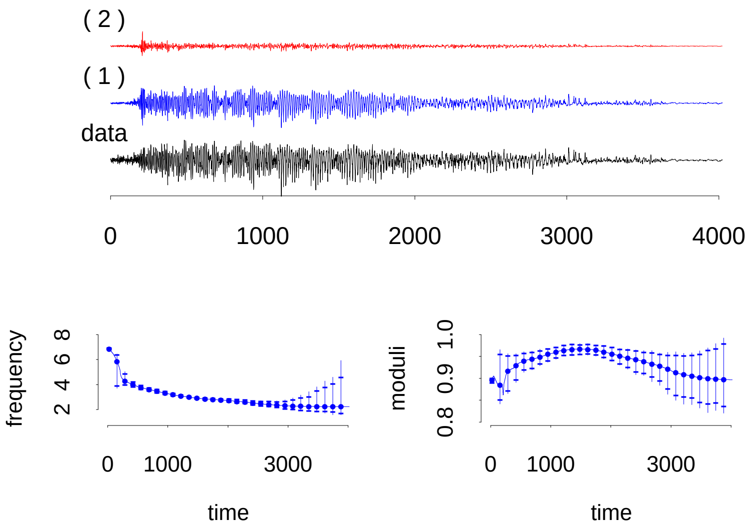
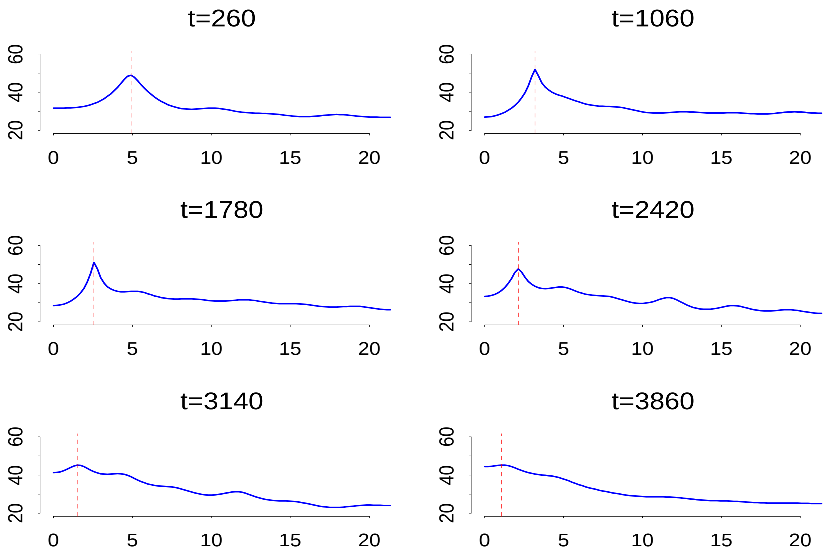


Figure 1: *top frame:* data and estimated components in the decomposition of EEG series **Fz** based on a TVAR(12) model. From the bottom up, the graph displays the time series followed by two estimated components in order of increasing characteristic frequency. *Bottom frames:* trajectories and 95% posterior bands of the estimated characteristic frequency and modulus of the lowest frequency component in series **Fz**.

The top frames of Figures 1 and 5 display a section of an EEG series recorded at a channel located in the central frontal cortical region of a patient scalp, named **Fz** in EEG nomenclature. The series displays high frequency oscillations at the beginning that slowly decay into lower frequencies, accompanied by an increase in the amplitude of the signal, relative to the amplitude observed at initial stages, until it finally decreases towards the end of the seizure episode. Figure 1 (top frame) displays the data and two of the estimated latent components in the decomposition of the series, based on a TVAR(12) model with constant observational variance $\sigma_t^2 = \sigma^2$, and discount factor $\beta = 0.996$ controlling the variability of $\boldsymbol{\phi}_t$. Components (1) and (2) correspond to 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 TVARMA(2,1) quasi-periodic structures. Process (1) is dominated by a TVARMA(2,1) with a time-varying characteristic frequency that gradually decays in time, as shown in the left bottom frame of Figure 1. This component, characteristic of slow-waves that usually appear in middle and late phases of effective ECT seizures (Weiner and Krystal 1993), also dominates in amplitude, having moduli values higher than 0.8 during most of the seizure course (see right bottom frame of Figure 1). Component (2) lies in the theta frequency band and is much lower in amplitude and modulus than component (1). Higher frequency components also appear in the decomposition having much lower amplitudes than the lower frequency components that really characterise the seizure episode.

The time trajectories of the characteristic frequency and modulus of the latent processes in the decomposition have an equivalent frequency-domain interpretation. In cases where the stationarity conditions are satisfied, i.e. if $|r_{t,j}| < 1$, the instantaneous spectral density of each latent quasi-periodic process $z_{t,j}$ is peaked around its characteristic frequency $\omega_{t,j}$ and the sharpness of the peak is an increasing function of its characteristic modulus $r_{t,j}$. Then, the spectrum of the full signal is time-varying, given at each time t as the product of the instantaneous spectra of the $y_{t,j}$ and of the AR part of $z_{t,j}$. The top frames of Figure 2 display six instantaneous spectra, computed at posterior mean estimates of the AR parameter $\boldsymbol{\phi}_t$ at different times during the seizure course. The vertical dotted lines indicate the value of the frequency (in Hz or cycles/sec) with the highest peak in each spectrum. The bottom frame of the figure displays the evolution of the instantaneous spectra computed at estimated posterior means of the AR parameter $\boldsymbol{\phi}_t$ at 50 equally spaced time points over the seizure course. The dotted lines correspond to the spectra shown in the top frames. As seen previously in the time-domain graphs displayed in Figure 1, the estimated spectra show that the EEG signal is dominated by the quasi-periodic process with the lowest characteristic frequency. The frequency is time-varying, having estimated values higher that 5 Hz at the beginning of the seizure that gradually decay towards the end (see dotted vertical lines). The degree of sharpness in the estimated spectra also varies over time, being sharpest at early-central portions of the seizure. This result is consistent with the estimated modulus trajectory in time of the latent process (1) graphed in Figure 1.



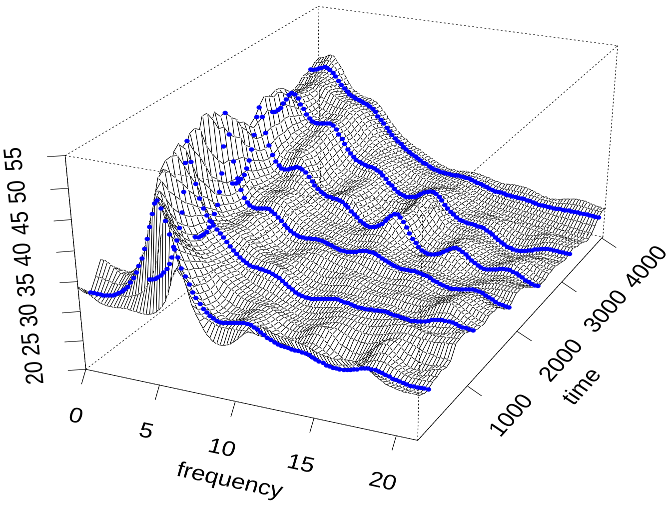


Figure 2: *top frame:* instantaneous estimated AR spectra for channel **Fz** computed at times $t = 260, 1060, 1780, 2420, 3140, 3860$. *Bottom frame:* evolution of the instantaneous spectra computed at estimated posterior mean values of ϕ_t at 50 equally spaced points along the seizure course.

3 Time-varying lag/lead structure in multiple TVAR models applied to EEG signals

There are ranges of possible model developments that might be of interest in relating time-varying characteristics of multiple time series through time. Motivated by the EEG analysis context above, novel models involving time-varying lag/lead structures among series were introduced in Prado *et al.*, (2000). The univariate TVAR decomposition analysis exemplified above is easily applied across collections of related series, as a starting point for exploring cross-series relationships. In the EEG context above, univariate TVAR(12) analyses and related decompositions yield to similar inferences across the full set of 19 EEG traces (Prado and West 1997). The instantaneous AR characteristic polynomials exhibit and maintain at least two pairs of complex conjugate roots across the 19 series, one of which corresponds to the dominant "seizure" latent process that lies in the delta frequency band. The range of values taken by the characteristic frequencies and moduli of the lowest frequency components over the seizure course, is consistent across the 19 EEG channels. Such common patterns suggest the notion of modelling the multiple traces via *latent factor models*, with one or two quasi-periodic latent processes or *factors* driving the behaviour of the series. As the factors may have a different impact on channels located at different sites on the scalp, the influence of the factors on each EEG series would be then weighted by individual regression coefficients or *factor weights*.

This direction was anticipated in Prado and West (1997) and further developed in Prado (1998) and Prado *et al.* (2000). Single factor model analyses of the multiple series reveal a spatial structure across the 19 EEG traces that univariate TVAR models are not able to capture: channels located closer on the scalp display similar estimated values of the factor weights. However, as discussed in Prado *et al.* (2000), cross-correlograms of the residuals of these models exhibit time dependent phase delays between some of the channels, evidencing substantial remaining structure across the 19 series. This motivates the use of dynamic regression models with time-varying lag/lead structures. We now describe such models following Prado *et al.*, (2000).

Let $y_{i,t}$ be the observation recorded at time t on channel i and consider the model

$$\begin{aligned} y_{i,t} &= \beta_{(i,t)} x_{i-l_{i,t}} + \nu_{i,t} \\ \beta_{(i,t)} &= \beta_{(i,t-1)} + \omega_{i,t}, \end{aligned} \tag{9}$$

where x_t is an underlying process assumed known at each time t; $l_{i,t}$ is the lag/lead that $y_{i,t}$ displays with respect to x_t, with $l_{i,t} \in \{-k_0, \ldots, 0, \ldots, k_1\}$ and k_0, k_1 known; $\beta_{(i,t)}$ is the dynamic regression coefficient of x_t for channel i; $\nu_{i,t}$ and $\omega_{i,t}$ are independent and mutually independent zero mean innovations with variances v_i and $s_{i,t}$. The changes in lag/lead structure over time are described via a one-step Markov chain model with known transition probabilities $p(l_{i,t} = k | l_{i,t} = m)$, $k, m \in \{-k_0, \ldots, 0, \ldots, k_1\}$, while a random walk is adopted to model the evolution of $\beta_{(i,t)}$. We also assume that $\nu_{t,i}$ and $\omega_{t,i}$ are independent across channels so that the equations (9) describe a collection of univariate models rather than a multivariate m-dimensional model. The specification of the evolution variances $s_{i,t}$ is handled via standard discount factor methods. Once the priors on $\beta_{(i,0)}$ and v_i are specified, posterior inference may be obtained via customised MCMC algorithms detailed in Prado *et al.* (2000).

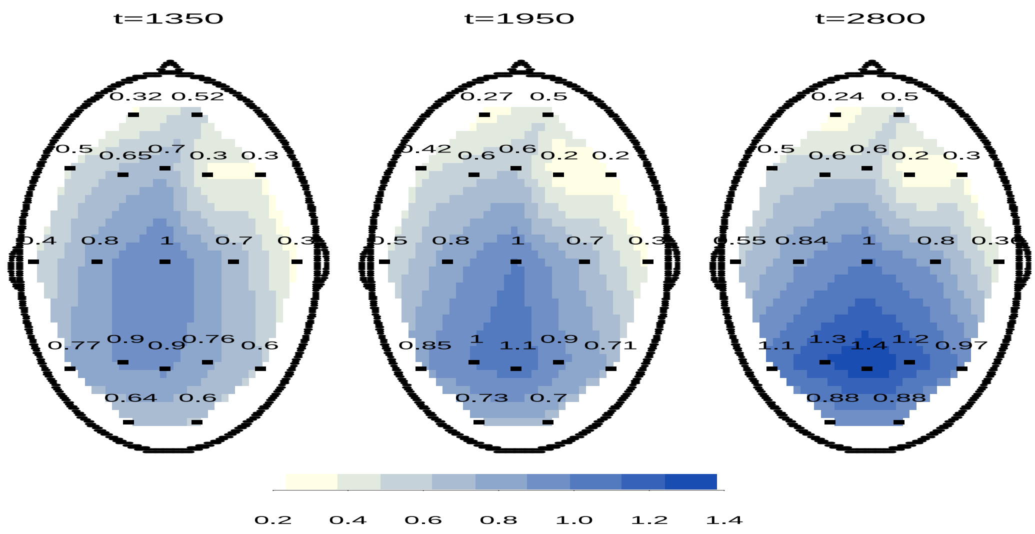


Figure 3: estimated posterior means of the dynamic factor weights at selected time points.

Given that x_t is the same fixed underlying process for all channels, it is possible to make comparisons between channels by comparing estimated values of $\beta_{(i,t)}$ and $l_{i,t}$ across i over time. Figure 3 displays the estimated posterior means of the $\beta.$ coefficients for all the channels at selected time points during the seizure, based on a model that takes $x_t = y_{t,\mathbf{Cz}}$, that is, x_t is the actual signal recorded at the channel located centrally, at the very top of the scalp. Details on the priors, discount factors and transition probabilities considered, as well as a discussion on MCMC convergence for this model appears in Prado *et al.* (2000). The values that appear at the approximate electrode locations in the graphs correspond to the actual estimated posterior mean values. In addition, an image plot, created by linear interpolation of $\hat{\beta}_{(i,t)}$ onto a grid defined by the approximate electrode locations is displayed. Dark intensities correspond to high values of the regression coefficients while light intensities match low values. Various features of the spatio:temporal relations between channels are evident from these pictures. The graphs exhibit marked patterns of relations across neighbouring channels: a given channel shares more similarities with channels located closer to it. There is also an element of asymmetry, channels located at right-fronto temporal sites have smaller regression coefficient values than channels located at left-fronto temporal sites.

Figure 4 shows estimated lag/leads based on posterior means of the $l.$ quantities at different time points. If a given site shows the lightest intensity at time t, then the signal recorded at this site is delayed in two units of time with respect to the signal recorded at site **Cz**. Similarly, if a site shows the darkest intensity at time t the signal recorded at this site leads the signal recorded at site **Cz** in two units of time. Central portions of the seizure display intense lag/lead activity, characterised by lags in the occipital regions and leads in the frontal and pre-frontal regions with respect to channel **Cz**, while almost no lag/lead structure is apparent at the beginning and towards the end of the seizure.

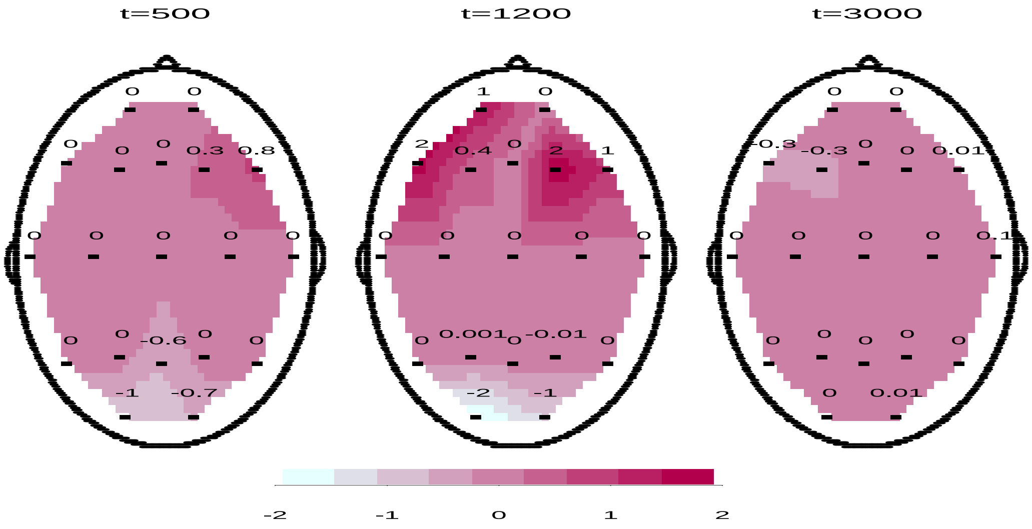


Figure 4: dynamic lag/leads based on posterior mean estimates.

4 Time-variation in model order

Time series model analyses incorporating model order uncertainty, and made implementable via Markov chain Monte Carlo (MCMC) methods, has been a growing area in recent years. Examples include methods for linear AR models, such as in Barnett *et al.* (1996), Barbieri and O'Hagan (1997) and Troughton and Godsill (1997). More recently, Huerta and West (1999) incorporated model order uncertainty in an AR(p) with emphasis on prior specification for latent structure.

For general DLMs, West and Harrison (1997, chapter 12), following Harrison and Stevens (1976), present the multi-process class of models, where model uncertainty is addressed using mixtures of conjugate DLMs. When some of the DLMs in consideration are not conjugate but conditionally conjugate, the multi-process analysis requires *Forward Filtering Backward Simulation* (FFBS) algorithms to obtain posterior model probabilities (Carter and Kohn 1994; Frühwirth-Schnatter 1994). Prado and Huerta (1999) adopt this approach to deal with model order uncertainty for TVAR models. We now review the main ideas of this work.

A time-varying autoregression with time-varying order p_t, is described by

$$x_t \quad = \quad \sum_{j=1}^{p_t} \phi_{t,j} x_{t-j} + \epsilon_t, \tag{10}$$

where the autoregressive coefficients change in time according to a random walk, as defined for a TVAR(p). For simplicity, ϵ_t are zero-mean innovations, assumed Gaussian with constant variance σ^2, but extensions to the time-varying case follow easily. Additionally, assume that p_t, the order of the autoregression at time t, is an integer that takes values between a fixed lower bound $p_{\min}$ and a fixed upper bound $p_{\max}$. The TVAR(p_t) model in (10), is a sub-model of a fixed order TVAR($p_{\max}$) described by

$$x_t = \sum_{j=1}^{p_{\max}} \phi_{t,j} x_{t-j} + \epsilon_t, \tag{11}$$

with a $p_{\max}$-dimensional vector of coefficients $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p_t}, 0, \ldots, 0)'$. Model completion requires specification of an initial prior for $(\boldsymbol{\phi}_1, \sigma^2)$ and details concerned with the evolution of model parameters. Relatively diffuse normal/inverse gamma priors are used on $\boldsymbol{\phi}_1$, and vague inverse-gamma priors on σ^2. The evolution of p_t is considered as a first order discrete random walk with known transition probabilities. Posterior inference of the TVAR(p_t) follows a two-stage Gibbs sampling format. Conditional on model orders, the standard sequential updating and retrospective filtering/smoothing algorithms for DLMs apply to update $\boldsymbol{\phi}_t$ and σ^2. The second stage consists of sampling from the conditional posterior distribution of model orders, given the $\boldsymbol{\phi}_t$ for all t and σ^2, via the filtering/smoothing algorithm for discrete random variables of Carter and Kohn (1994). Full description of the simulation algorithm and mathematical details appear in Prado and Huerta (1999).

4.1 Decompositons for time-varying autoregressions

Decomposition of a TVAR(p_t) is obtained via the decomposition theory for a general DLM as earlier described. The representation of the TVAR(p_t) in DLM form involves an evolution matrix $\mathbf{G}_t$ that has p_t distinct non-zero eigenvalues and a zero eigenvalue with multiplicity $p_{\max} - p_t$. The decompositon results of Section 2 apply at each time point, so that

$$x_t = \sum_{j=1}^{c_t} z_{t,j} + \sum_{j=2c_t+1}^{p_t} y_{t,j}, \tag{12}$$

where c_t is the number of complex pairs of non-zero eigenvalues of the system matrix $\mathbf{G}_t$ at time t. Now the number of components depend on time varying c_t and p_t. As in the fixed order TVAR case, $z_{t,j}$ are related to the complex non-zero eigenvalues of $\mathbf{G}_t$ and dominated by a TVARMA(2,1). The $y_{t,j}$ are related to the real non-zero eigenvalues of $\mathbf{G}_t$ and dominated by a TVAR(1).

4.2 Describing changes in the number of latent EEG processes

Consider again the EEG series displayed in Figure 1. The latent components shown in the graph were computed using estimated posterior means for the AR coefficients and the innovations variance of a TVAR(12) model. Here, we model the same series with a TVAR(p_t), where p_t may take values from $p_{\min} = 0$ up to $p_{\max} = 14$. Different values for the lower and upper bands $p_{\min}$ and $p_{\max}$ were considered, leading to similar inferences in terms of the latent structure. Discount factors in the range of $0.99 - 0.999$ were used to control the evolution of the AR coefficients in time. Such values impose smoothness restrictions on the changes of $\boldsymbol{\phi}_t$ in time that are typical in EEG analyses (West *et al.*, 1999). Similarly, the transition probability structure that describes the evolution of p_t in time is specified to impose smoothness conditions, allowing to include or delete only one characteristic root - complex or real - at each time t. Specifically in this example, denoting $q_{ij} = P[p_t = i | p_{t-1} = j]$,

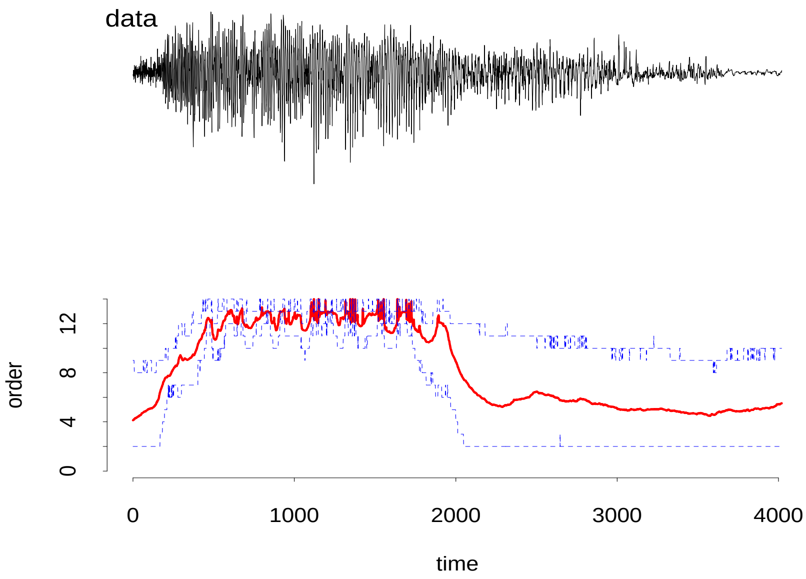


Figure 5: From the top down we have the EEG data and a graph of the estimated posterior mean for model order at each time t with 95% posterior bands.

we take $q_{ii} = 0.99$ for all i, $q_{i,i+1} = q_{i,i-1} = 0.004$, $q_{i,i+2} = q_{i,i-2} = 0.001$ 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 $q_{1,3} = q_{13,11} = 0.002$. In addition, a discrete uniform prior on model order, $P(p_1 = i) = 1/15$ for all i, and relatively diffuse conjugate normal/inverse-gamma priors were used for the AR coefficients and the innovations variance.

Figure 5 displays, from the top down, the data and the trajectory in time of the estimated posterior mean for model order (solid line) with 95% posterior probability bands (dotted lines). The instantaneous posterior means and probability bands for model order are based on 4,000 samples taken from 17,000 iterations of the Gibbs sampler after a burn-in of 3,000 iterations for MCMC convergence. The graph shows that the model order is higher roughly between $t = 400$ and $t = 2000$, indicating that the latent structure is more complex during this period than at the beginning of the seizure and after $t = 2000$. The posterior mean oscillates around 12 between $t = 400$ and $t = 2000$, with 95% bounds in the range from 10 to 14. Approximately at $t = 1800$ the uncertainty on model order starts to increase, with 95% posterior bands in the 2 to 10 range. This is consistent with the relatively broad posterior bands observed in the graphs of the trajectories in time of the characteristic frequency and modulus of component (1) in the decomposition obtained with a TVAR(12) (see Figure 1).

5 Other recent, current and potential future directions

Time-varying autoregressive models constitute a suitable class of models to study the behaviour of non-stationary time series. The related decomposition theory summarised here has proven useful in a variety of applications where the interest lies in discovering and interpreting latent structure in the series. Via efficient MCMC simulation, the model may be extended to have time varying order which permits to describe the changes in the number of latent components.

In connection with the use of simulation methods for model implementation more broadly, pressing research issues arise in contexts where analysis is necessarily sequential, such as in fast processing of speech and in other communications signal processing applications. The recent upsurge in research and algorithmic development involving particle filtering methods (Pitt and Shephard 1999; Liu and West 2000; and Doucet *et al.* 2000) has led to a range of new and efficient sequential filtering algorithms that may be applied to sequential analysis in many different time series frameworks. Use of these methods in applications of TVAR models in speech signal processing has been a highlight of the work of Godsill *et al.* (2000b); see also the related discussions in Godsill *et al.* (2000a) and Doucet *et al.* (2000). In addition to developing filtering and smoothing methodology, these authors explore TVAR models in which the autoregressive coeeficients evolve in time according to models that are alternatives to the usual random walks. These include, in particular, novel models that are random walk on the time-varying partial autocorrelations rather than raw AR coefficients. In addition to providing a statistically intuitive parametrisation, these models have physical interpretation in a speech processing context. This is a currently active research domain, and one that we should see growing in the near future.

Additional research frontiers consider the multivariate decomposition results of Section 2, and aim to extend the prior specifications of Huerta and West (1999) for univariate AR processes to multivariate vector autoregressive models. In fact, a first extension proposes a diagonal VAR(p) with a prior that allows for possible zero characteristic roots, i.e. takes into account model order uncentainty, but also allows for potential ties of characteristic roots across series. Extending such thinking to multivariate time-varying vector autoregressions is a further research challenge.

Software

Some readers may be interested in software that implements the TVAR analysis and decomposition methodology, and the component-based model uncertainty analyses of AR models, as described and exemplified here. See `www.stat.duke.edu/isds-info/software.html` for free software and further information.

Acknowledgements

Research reviewed here was partially supported by the National Science Foundation of the USA under grants DMS-9704432 and DMS-9707914.

Bibliography

Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics*, **18**.

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.

Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*, 2nd edn. San Francisco: Holden-Day.

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

Doucet, A., Freitas, J. F. G. De and Gordon, N. J. (2000) *Sequential Monte Carlo Methods in Practice.* New York: Springer-Verlag.

Doucet, A., Godsill, S. J. and West, M. (2000) Monte Carlo filtering and smoothing with application to time-varying spectral estimation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. II, pp. 701–704. IEEE.

Frühwirth-Schnatter, S. (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.

Godsill, S. J., Doucet, A. and West, M. (2000a) Maximum a *posteriori* sequence estimation using Monte Carlo particle filters. *Annals of the Institute of Statistical Mathematics.*

Godsill, S. J., Doucet, A. and West, M. (2000b) Methodology for Monte Carlo smoothing with application to time-varying autoregressions. Technical Report 00-01. ISDS, Duke University.

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.

de Jong, P. and Shephard, N. (1995) The simulation smoother for time series models. *Biometrika*, **82,** 339–350.

Kitagawa, G. (1983) Changing spectrum estimation. *J. Sound and Vibration*, **89,** 443–445.

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

Krystal, A. D., Prado, R. and West, M. (1999) New methods of time series analyis of non-stationary EEG data: Eigenstructure decompositions of time varying autoregressions. *Clinical Neurophysiology*, **110,** 2197–2206.

Liu, J. and West, M. (2000) Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (eds J. F. G. De Freitas A. Doucet and N.J. Gordon). New York: Springer-Verlag.

Pitt, M. K. 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 Huerta, G. (1999) Time varying autoregressions with model order uncertainty. Technical Report. Centro de Estadística y Software Matemático, Universidad Simón Bolívar.

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.

Prado, R., West, M. and Krystal, A.D. (2000). *Journal of the Royal Statistical Society, Series* C.

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.

Weiner, R.D. and Krystal, A.D. (1993) EEG monitoring of ECT seizures. In *The Clinical Science of Electroconvulsive Therapy, Washington, DC* (ed. C.E. Coffey), pp. 93–109. American Psychiatry Press.

West, M. (1997) Time series decomposition. *Biometrika*, **84,** 489–494.

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.