AdaptSPEC: Adaptive Spectral Estimation for Nonstationary Time Series

Ori Rosen

Department of Mathematical Sciences, University of Texas at El Paso,

El Paso, Texas 79968, U.S.A.

Sally Wood

Melbourne Business School, University of Melbourne,

Melbourne, Victoria 3053, Australia

and

David S. Stoffer

Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, U.S.A.

We propose a method for analyzing possibly nonstationary time series by adaptively dividing the time series into an unknown but finite number of segments and estimating the corresponding local spectra by smoothing splines. The model is formulated in a Bayesian framework, and the estimation relies on reversible jump Markov chain Monte Carlo (RJMCMC) methods. For a given segmentation of the time series, the likelihood function is approximated via a product of local Whittle likelihoods. Thus, no parametric assumption is made about the process underlying the time series. The number and lengths of the segments are assumed unknown and may change from one MCMC iteration to another. The frequentist properties of the method are investigated by simulation, and applications to EEG and the El Niño Southern Oscillation phenomenon are described in detail.

Key Words: EEG; ENSO; Locally stationary time series; Reversible jump MCMC; Spectral estimation; Whittle likelihood.

1 Introduction

Many time series are realizations of nonstationary random processes, hence estimating their time varying spectra may provide insight into the physical processes that give rise to these

time series. For example, EEG time series are typically nonstationary, and estimating the time varying spectra based on the EEG of epilepsy patients may lead to methods capable of predicting seizure onset. Similarly, analyzing the time varying spectrum of the Southern Oscillation Index (SOI) may further our knowledge of the frequency of the El Niño Southern Oscillation (ENSO) phenomenon and its impact on global climate.

This paper proposes methodology for analyzing possibly nonstationary time series by adaptively dividing the time series into an unknown but finite number of segments and estimating the corresponding local spectra by smoothing splines. The model is formulated in a Bayesian framework, and the estimation relies on reversible jump Markov chain Monte Carlo (RJM-CMC) methods. For a given segmentation of the time series, the likelihood function is approximated via a product of local Whittle likelihoods. Thus, no parametric assumption is made about the process underlying the time series. The number of segments and the length of each segment are assumed unknown and may change from one MCMC iteration to another.

The basic assumptions of the model are that, conditional on the location and number of segments, the time series is piecewise stationary and that the spectrum for each segment is smooth. A detailed description of the model is given in Section 3. In addition to representing time series that have regime changes, the model can be used to approximate slowly varying processes such as locally stationary processes defined in Dahlhaus (1997), or time varying autoregressive processes for which the parameters are allowed to vary slowly with time; see Adak (1998) for details.

Several authors have considered the estimation of locally stationary processes under a variety of assumptions. The estimators that were developed by Dahlhaus (1997) for his evolutionary spectra are consistent, but the method is not computationally efficient and can be problematic when the time series is long. Chiann and Morettin (1999) proposed a wavelet-based version of the estimator proposed by Dahlhaus (1997). Various other approaches have been suggested to overcome the computational difficulty. Ombao et al. (2001) proposed nonparametric estimators based on smooth local exponential functions. Guo et al. (2003) extended the work of Ombao et al. (2001) to allow for simultaneous smoothing in both the time and

frequency domains.

Qin and Wang (2008) use the basic method of Guo et al. (2003) and focus on applying it to the analysis of EEG time series. Our method also estimates a time varying spectral density but differs from Qin and Wang (2008) in three important ways. First, our approach assumes no pre-specified time intervals over which the process is considered locally stationary. Instead, our technique uses the data to determine the size and location of the piecewise stationary time intervals. Second, our estimate of the time varying spectrum is computed by averaging over a range of smoothing parameters, where the averaging is with respect to the posterior density of the smoothing parameters. In contrast, Qin and Wang (2008) estimate a single value of the smoothing parameter which is then plugged in to estimate the time varying spectrum. Using a single value for the smoothing parameter rather than averaging over values of the smoothing parameter results in a wider range of estimates, as demonstrated in Section 4. Third, our approach can capture abrupt changes as well as recover changes that occur more gradually. In Section 4.2 we show that although our approach is piecewise, conditional on a specific partition, it is not necessarily piecewise, unconditional on the partition points, due to the uncertainty surrounding the number and location of the partition points. Indeed, it will only be piecewise stationary if the posterior probability that a partition point occurs at a specific time is equal to 1, for all partition points.

Rosen et al. (2009) estimate the log of the local spectrum using a Bayesian mixture of splines. The basic idea of this approach is to first partition the data into small sections. It is then assumed that the log spectral density of the evolutionary process in any given partition is a mixture of individual log spectra. A mixture of smoothing splines model with time varying mixing weights is used to estimate the evolutionary log spectrum. The mixture model is fit using MCMC techniques that yield estimates of the log spectra of the individual subsections. In contrast to Rosen et al. (2009), the current paper does not use predetermined partitions. Rather, it adaptively divides the time series into segments of variable lengths, rendering the mixture model unnecessary. In addition to more accurate estimation, this also leads to computational saving.

Analyzing locally stationary time series can also be done in the time domain. Kitagawa

and Akaike (1978) suggested fitting piecewise autoregressive (AR) models. Wood et al. (2011) propose a class of models for analyzing possibly nonstationary time series, formed as a mixture of AR models with a common but unknown lag, whose mixing weights are a function of time. The model parameters, including the number of mixture components, are estimated via MCMC methods. Lau and So (2008) use a Dirichlet process mixture of AR processes to flexibly model the predictive density of a time series. Their approach does not handle structural breaks in the time series and their mixture weights are not functions of time. Davis et al. (2006) suggested fitting piecewise AR models using minimum description length and a genetic algorithm for solving the difficult optimization problem. Although Davis et al. (2006) showed that their simulation results for a few locally (and piecewise) stationary AR models perform better than those of Ombao et al. (2001), it is clear that, generally, a parametric technique will outperform a nonparametric technique when the parametric model is correct.

A different time-domain approach is to model the parameter evolution over time. An excellent treatment of the problem based on state-space models with smoothness priors is the influential text by Kitagawa and Gersch (1996). This approach was expanded upon by many authors, for example, West et al. (1999) allowed the parameters of an AR process to change over time by modeling them as a random walk, assuming that the maximum lag in the autoregressive process is fixed. This assumption was relaxed by Prado and Huerta (2002). Gerlach et al. (2000) provide a sampling scheme that allows for smooth parameter evolution, as well as structural breaks in the parameters.

The paper is organized as follows. Sections 2 and 3 present the model and priors for stationary and nonstationary time series respectively. Section 4 outlines the proposed Bayesian inference as well as provides illustrative examples. Section 5 reports results of a simulation study, and Section 6 illustrates the methodology with the analysis of EEG data and indicators for the ENSO phenomenon.

2 Spectral Estimation for Stationary Time Series

2.1 Model

Our approach to the problem of estimating local spectra is best understood by first explaining the technique for estimating the spectral density of a stationary process. Suppose that a stationary time series, $\{X_t\}$, has a bounded positive spectral density, $f(\nu)$, for $-1/2 < \nu \le 1/2$. Given a realization, $x_1, \ldots, x_n$, the periodogram of the data at frequency ν (measured in cycles per unit time) is

$$I_n(\nu) = \frac{1}{n} \left| \sum_{t=1}^{n} x_t \exp(-2\pi i \nu t) \right|^2 .$$

Let $\nu_k = k/n$, for $k = 0, \ldots, n-1$, be the Fourier frequencies. Whittle (1957) showed that, under appropriate conditions, for large n the likelihood of $\boldsymbol{x} = (x_1, \ldots, x_n)'$, given $\boldsymbol{f} = (f(\nu_0), \ldots, f(\nu_{n-1}))'$, can be approximated by

$$p(\boldsymbol{x} \mid f) = (2\pi)^{-n/2} \prod_{k=0}^{n-1} \exp\left\{ -\frac{1}{2}\left[\log f(\nu_k) + I_n(\nu_k)/f(\nu_k) \right] \right\}. \tag{1}$$

Note that in (1), there are only $[n/2]+1$ distinct observations since the spectral density and the periodogram are both even functions of ν. The notation $[n]$ means the largest integer less than or equal n. For ease of notation, in what follows, we assume that n is even. Letting $y_n(\nu_k) = \log I_n(\nu_k)$ and $g(\nu_k) = \log f(\nu_k)$, the representation (1) suggests the log-linear model

$$y_n(\nu_k) = g(\nu_k) + \epsilon_k, \tag{2}$$

where the ϵ_k's are independent, $\epsilon_k \sim \log(\chi_2^2/2)$ for $k = 1, \ldots, n/2 - 1$, and $\epsilon_k \sim \log(\chi_1^2)$ for $k = 0, n/2$. Representation (2) was used by a number of authors for nonparametric estimation of the log spectral density. For example, Wahba (1980) used a frequentist approach for estimating $g(\nu)$ via cubic smoothing splines. Carter and Kohn (1997) achieved the same goal in a Bayesian framework by expressing equation (2) in a state-space form. Carter and Kohn (1997) approximated the error distribution in (2) by a mixture of five normal densities and introduced latent component indicators to facilitate the estimation.

Rather than using (2) for estimating $g(\nu)$, we use (1) directly. The next section provides details on the prior distributions placed on $g(\nu)$.

2.2 Priors

To place a prior on $g(\nu_k)$, we follow Wahba (1990, p.16), and express $g(\nu_k)$ as the sum of its linear and nonlinear components, so that

$$g(\nu_k) = \alpha_0 + \alpha_1 \nu_k + h(\nu_k),$$

where $h(\nu_k)$ is the nonlinear component. We place a linear smoothing spline prior on the vector $\boldsymbol{h} = (h(\nu_0), \ldots, h(\nu_{n/2}))'$, which means that

$$h(\nu) = \tau \int_0^\nu W(v) dv$$

or equivalently $\boldsymbol{h} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{\Omega})$, where τ^2 is a smoothing parameter and $(\boldsymbol{\Omega})_{ij} = \min(\nu_i, \nu_j)$. The parameters α_0 and α_1 are the values of $g(\nu)$ and its first derivative at $\nu = 0$, respectively. The symmetry and periodicity of the spectral density mean that $(\partial g(\nu)/\partial \nu)|_{\nu=0} = 0$. Accordingly, α_1 is set to be identically zero, and the prior on α_0 is $N(0, \sigma_\alpha^2)$, for some large σ_α^2. To complete the prior specification on $g(\nu)$, we follow Gelman (2006) and assume $\tau^2 \sim U(0, c_{\tau^2})$, where c_{τ^2} is a known large value. We express $\boldsymbol{h}$ as a linear combination of basis functions, $\boldsymbol{h} = \boldsymbol{X}\boldsymbol{\beta}$, where the columns of the design matrix $\boldsymbol{X}$ are the Demmler-Reinsch basis functions evaluated at the Fourier frequencies, and $\boldsymbol{\beta}$ is vector of unknown coefficients. We follow Wood et al. (2002) and Rosen et al. (2009) and retain only the basis functions corresponding to the $J = 10$ largest eigenvalues, resulting in significant computational saving. For linear smoothing splines, the jth column of $\boldsymbol{X}$, $j = 1, \ldots, J$, is $\sqrt{2}\cos(j\pi\boldsymbol{\nu})$ (see Eubank (1999)), where $\boldsymbol{\nu} = (\nu_0, \ldots, \nu_{n/2})'$. The prior on $\boldsymbol{\beta}$ is $N(0, \tau^2 I_J)$, where I_J is a $J \times J$ identity matrix.

2.3 Sampling Scheme

The parameters α_0, $\boldsymbol{\beta}$ and τ^2 are drawn from the posterior distribution $p(\alpha_0, \boldsymbol{\beta}, \tau^2 \mid \boldsymbol{y})$, where $\boldsymbol{y} = (y_n(\nu_0), \ldots, y_n(\nu_{n/2}))'$, using MCMC methods, as follows.

1. α_0 and $\boldsymbol{\beta}$ are sampled jointly via a Metropolis-Hastings (M-H) step from

$$p(\alpha_0, \boldsymbol{\beta} \mid \tau^2, \boldsymbol{y}, \boldsymbol{X}) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{n-1}[\alpha_0 + \boldsymbol{x}_k'\boldsymbol{\beta} + \exp(y_n(\nu_k) - \alpha_0 - \boldsymbol{x}_k'\boldsymbol{\beta})] - \frac{\alpha_0^2}{2\sigma_\alpha^2} - \frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right\}, \tag{3}$$

where $\boldsymbol{x}_k'$ is the kth row of $\boldsymbol{X}$.

2. τ^2 is sampled from the truncated inverse gamma distribution, $IG(J/2-1, \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\beta})$, whose density is

$$p(\tau^2 \mid \boldsymbol{\beta}) \propto (\tau^2)^{-J/2} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right), \ \tau^2 \in (0, c_{\tau^2}]. \tag{4}$$

3 Spectral Estimation for Nonstationary Time Series

3.1 Model

To describe our proposed model, let a time series consist of an unknown number of segments, m, and let $\xi_{j,m}$ be the unknown location of the end of the jth segment, $j = 1, \ldots, m$, where $\xi_{0,m}$ and $\xi_{m,m}$ are $t = 0$ and $t = n$, respectively. Then conditional on m and $\boldsymbol{\xi}_m = (\xi_{0,m}, \ldots, \xi_{m,m})'$, we assume that the process $\{X_t\}$ is piecewise stationary. That is,

$$X_t = \sum_{j=1}^{m} X_t^{(j)} \delta_{j,m}(t), \tag{5}$$

where, for $j = 1, \ldots, m$, the processes $X_t^{(j)}$ are independent and stationary with spectral density $f_{j,m}(\nu)$, and $\delta_{j,m}(t) = 1$ if $t \in [\xi_{j-1,m} + 1, \xi_{j,m}]$ and 0 otherwise.

Consider a realization $\boldsymbol{x} = (x_1, \ldots, x_n)'$ from process (5), where the number and locations of the stationary segments are unknown. Let $n_{j,m}$ be the number of observations in the jth segment. We assume that $n_{j,m} \geq t_{\min}$, where $t_{\min}$ is taken to be large enough in order for the local Whittle likelihood to provide a good approximation to the likelihood. Given a partition of the time series $\boldsymbol{x}$, the jth segment consists of the observations $\boldsymbol{x}_{j,m} = \{x_t : \xi_{j-1,m} + 1 \leq t \leq \xi_{j,m}\}$, $j = 1, \ldots, m$, with underlying spectral densities $f_{j,m}$ and periodograms $I_{n_{j,m}}$, evaluated at frequencies $\nu_{k_j} = k_j/n_{j,m}$, $0 \leq k_j \leq n_{j,m} - 1$. For a given partition $\boldsymbol{\xi}_m$, the

7

approximate likelihood of the time series is thus

$$L(f_{1,m}, \ldots, f_{m,m} \mid \boldsymbol{x}, \boldsymbol{\xi}_m) = \prod_{j=1}^{m} (2\pi)^{-n_{j,m}/2} \prod_{k_j=0}^{n_{j,m}-1} \exp\left\{-\frac{1}{2}\left[\log f_{j,m}(\nu_{k_j}) + I_{n_{j,m}}(\nu_{k_j})/f_{j,m}(\nu_{k_j})\right]\right\}.$$

Prior distributions are placed on all the parameters, including the number of segments, m, and the partition, $\boldsymbol{\xi}_m$.

3.2 Priors

For a given number of segments, m, the following priors are used.

1. The priors on the log spectra $g_{j,m}(\nu) = \log f_{j,m}(\nu)$, $j = 1, \ldots, m$, are assumed to be independent and are as given in Section 2.2.

2. The prior on the partition $\boldsymbol{\xi}_m$ is

$$\Pr(\boldsymbol{\xi}_m \mid m) = \prod_{j=1}^{m-1} \Pr(\xi_{j,m} \mid \xi_{j-1,m}, m),$$

where $\Pr(\xi_{j,m} = t \mid m)$ is a discrete uniform such that

$$\Pr(\xi_{j,m} = t \mid m) = 1/p_{jm}$$

for $j = 1, \ldots, m - 1$. The number of available locations for partition point $\xi_{j,m}$ is denoted by p_{jm} and is equal to $n - \xi_{j-1,m} - (m - j + 1)t_{\min} + 1$. This prior states that the first partition point $\xi_{1,m}$ is equally likely to occur at any point in the time series subject to the constraint that there are at least $t_{\min}$ observations in each of the m segments. The prior on subsequent partition points is similar and states that, conditional on the previous partition point, the next partition point is equally likely to occur in any available location, again subject to the constraint that there are at least $t_{\min}$ observations in each segment. We believe that this prior is intuitively appealing, but our framework can accommodate other priors for the partition points.

The prior on the number of segments is a discrete uniform with a maximum number M, so that $\Pr(m = k) = 1/M$ for $k = 1, \ldots, M$. Typically M is chosen to be large enough to

capture all visible locally stationary segments but if, after running the procedure, we find that $\Pr(m = M|\boldsymbol{x}) \not\approx 0$ then we increase M.

4 Bayesian Inference

4.1 Sampling Scheme

Each MCMC iteration consists of two types of moves, within-model moves and between-model moves. An outline of the sampling scheme follows. Further details are given in the Appendix.

Within-model moves

Given a current number of segments m^c, a single partition point ξ_{k^*,m^c} is proposed to be relocated. The corresponding basis function coefficients in the pair of adjacent segments affected by the relocation of the partition point are then updated. These two steps are jointly accepted or rejected in a M-H step. The smoothing parameters are then updated in a Gibbs step as in equation (4). See the Appendix for details.

Between-model moves

The number of segments is either proposed to increase by 1 (birth) or decrease by 1 (death), so that $m^p = m^c + 1$ or $m^p = m^c - 1$, respectively.

- If a birth is proposed ($m^p = m^c + 1$), then an additional partition point is drawn by first selecting a segment to split and then selecting the new partition point from within this segment. Finally, two new smoothing parameters for the new segments are formed from the current single smoothing parameter, and conditional on these new smoothing parameters, two new sets of basis function coefficients are drawn.

- If a death is proposed ($m^p = m^c - 1$), then a partition point is selected to be removed. A single new smoothing parameter is then formed from the adjacent pair of current smoothing parameters, and conditional on the new smoothing parameter, a new set of basis function coefficients is proposed.

4.2   Illustrative Examples

To illustrate the methodology and to demonstrate that our approach can capture changes that occur smoothly as well as abruptly, we present results based on single realizations from a piecewise autoregressive process and a slowly varying autoregressive process. The model is fitted to the data with a total of 10,000 iterations, 2000 of which are used as burn-in. The value of $t_{\min}$ is set to 40, and the number of the spline basis functions is set to 10.

Piecewise autoregressive process

A realization is drawn from the process

$$x_t = \begin{cases} 0.9x_{t-1} + \epsilon_t^{(1)} & \text{for } 1 \le t \le 300 \\ -0.9x_{t-1} + \epsilon_t^{(2)} & \text{for } 301 \le t \le 600 \\ 1.5x_{t-1} - 0.75x_{t-2} + \epsilon_t^{(3)} & \text{for } 601 \le t \le 1000, \end{cases} \tag{6}$$

where $\epsilon_t^{(i)} \overset{\text{iid}}{\sim} N(0,1)$, $i = 1, 2, 3$. Figure 1 presents a realization from model (6). Figure 2
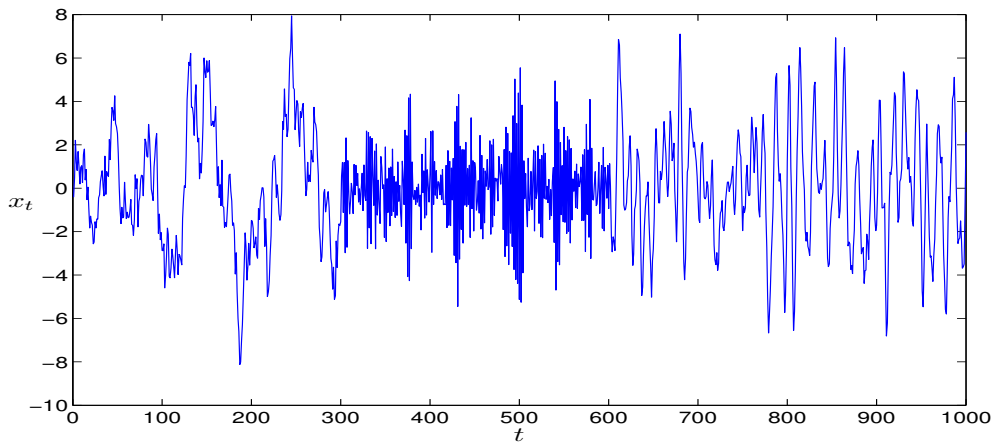

Figure 1: A realization from model (6).

displays the three true log spectral densities (solid line) corresponding to model (6), along with the spline fits (dashed). The posterior probability of three segments is 99.75%. The posterior means of the partition points are $\hat{\xi}_{1,3} = 300.5$ and $\hat{\xi}_{2,3} = 597.9$.
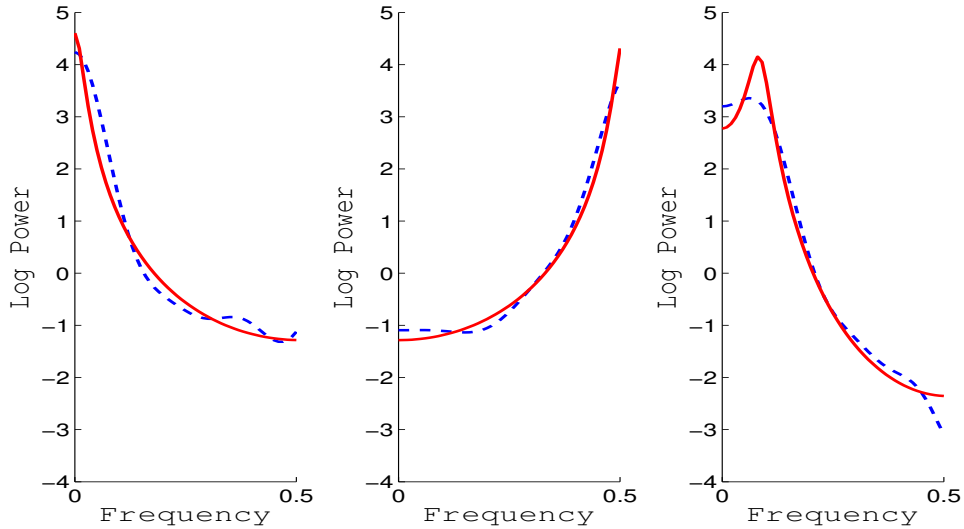
10

Figure 2: Log spectral densities (solid) and their estimates (dashed) for model (6).

Slowly varying autoregressive process

We now illustrate how AdaptSPEC can model time series that change slowly over time as well as those time series in which there is an abrupt change. We generate one realization from each of the models

$$x_t = a_t x_{t-1} + \epsilon_t \quad \text{where} \quad a_t = -0.5 + t/500 \quad \text{for} \quad t = 1, \ldots, 500 \quad \text{and} \tag{7}$$

$$x_t = a_t x_{t-1} + \epsilon_t \quad \text{where} \quad a_t = \begin{cases} -0.5 & \text{for} \quad t \le 250 \\ 0.5 & \text{for} \quad t > 250, \end{cases} \tag{8}$$

where $\epsilon_t \sim N(0,1)$. In Figure 3, panel (a) shows a realization from model (7), and panel (b) displays the coefficient a_t as a function of t. Panels (c) and (d) display analogous plots corresponding to model (8). Panels (a) and (b) of Figure 4 show the posterior distributions of the number of segments for models (7) and (8), respectively. Figure 5 shows in panel (a) the estimated posterior distribution of the partition point, $\hat{P}(\xi_{1,2} = t \mid \boldsymbol{x})$, and in panel (b), the estimated cumulative distribution function (cdf) of the partition point, $\hat{P}(\xi_{1,2} < t \mid \boldsymbol{x})$, for model (7). Panels (c) and (d) show the analogous plots corresponding to model (8). Figure 5 shows that when the time series changes slowly, so too does the probability that the change occurs before time t, while when the time series changes abruptly, the cdf of the partition point is a step function. The reason for this is that even though our model is a piecewise model, conditional on the partition point, our technique can recover slowly
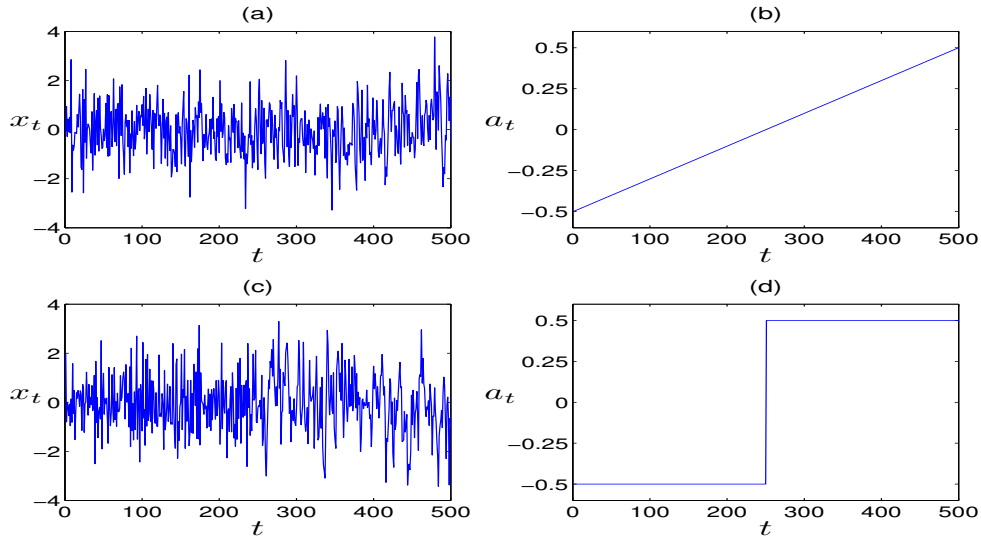
11

Figure 3: Panels (a) and (b): plots of x_t and a_t, respectively, corresponding to model (7); panels (c) and (d): analogous plots corresponding to model (8).
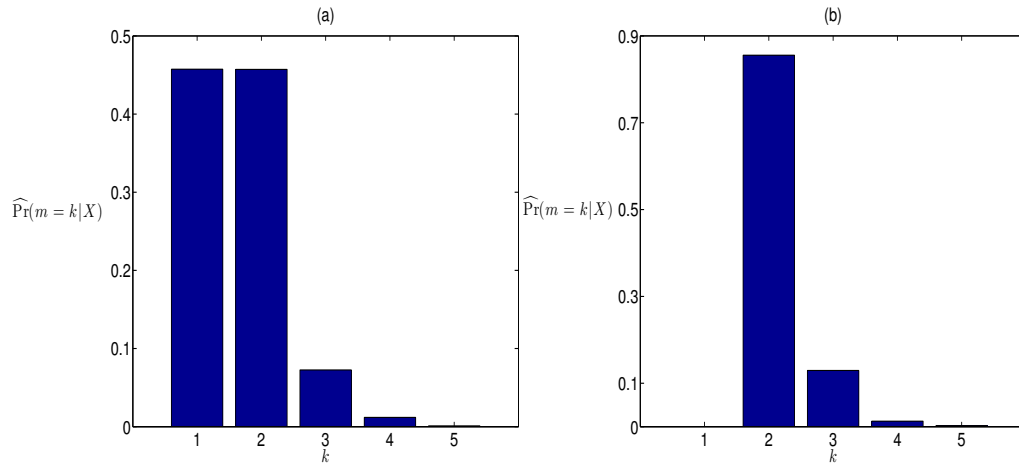

Figure 4: Panels (a) and (b): The estimated posterior distributions of the number of segments for models (7) and (8), respectively.

varying time series by averaging over the possible locations of the partition point. When the posterior probability of the partition point is concentrated on a few values, then the average is taken over only those few values, while if the posterior probability of the partition point has support across the entire time series, then the average is taken across all those values.
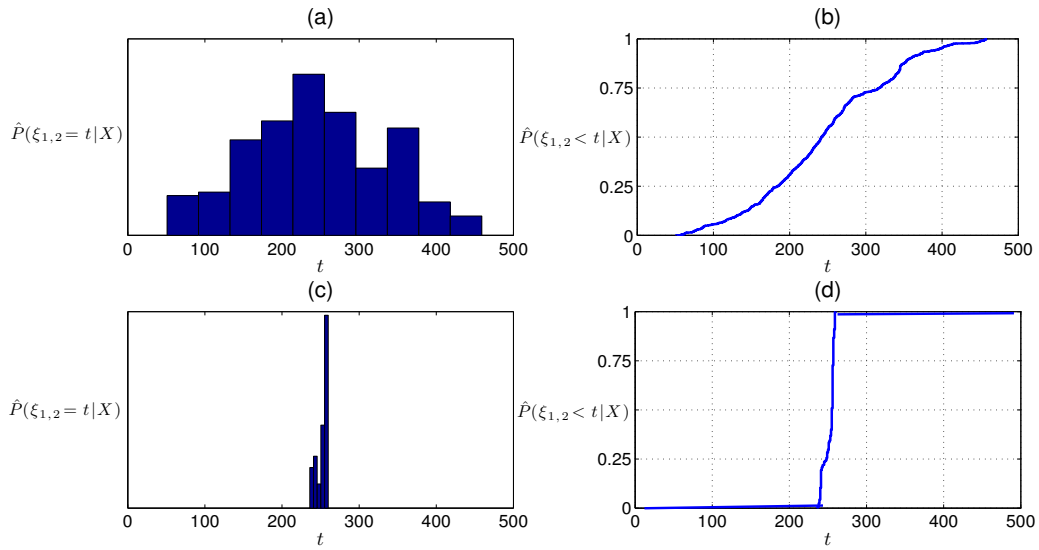
Figure 5: Panels (a) and (b): Plots of $\hat{P}(\xi_{1,2} \mid \boldsymbol{x})$ and $\hat{P}(\xi_{1,2} < t \mid \boldsymbol{x})$, respectively, corresponding to model (7); panels (c) and (d): analogous plots corresponding to model (8).

5 Simulation Study

To evaluate our methodology we consider various models and simulate 50 data sets from each. To fit each data set, a total of 10,000 iterations are used with a burn-in period of 2000 iterations. For each data set, the number of basis functions is set to 10, and the maximum number of segments is set to 4, except in Section 5.3, where it is set to 8. To judge the quality of the estimated spectra, we compute the mean-squared error

$$MSE = \{n(K+1)\}^{-1} \sum_{t=1}^{n} \sum_{k=0}^{K} \{\log \hat{f}(t, \nu_k) - \log f(t, \nu_k)\}^2,$$

where K is set to 50.

5.1 Stationary AR(3) Process

This section demonstrates that AdaptSPEC can be used even if the time series is stationary. The stationary process used for this simulation is taken from Qin and Wang (2008) and is given by

$$x_t = 1.4256x_{t-1} - 0.7344x_{t-2} + 0.1296x_{t-3} + \epsilon_t, \tag{9}$$

where $\epsilon_t \overset{iid}{\sim} N(0, 1)$, $t = 1, \ldots, 256$. Figure 6, panel (a), displays the theoretical time varying
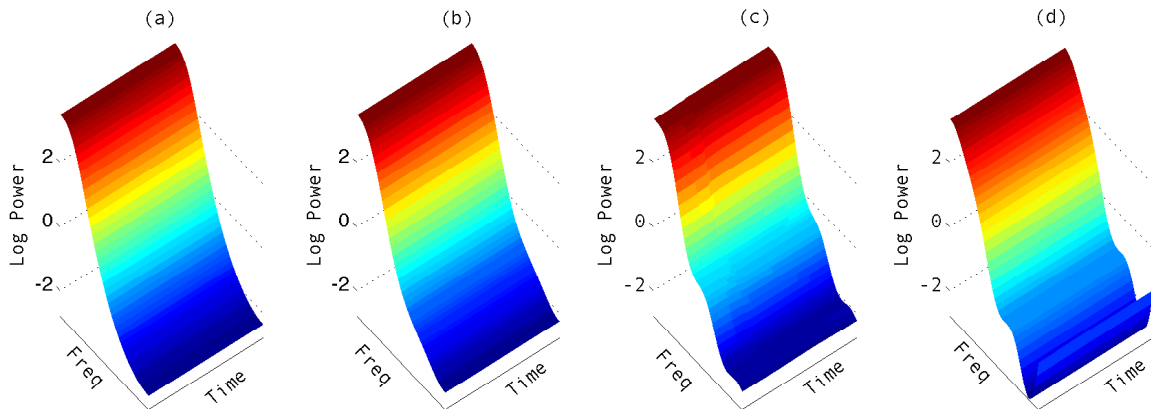

Figure 6: Panel (a): True time varying log spectrum of model (9). Panels (b)–(d): estimated time varying log spectra corresponding to the 10th, 50th and 90th percentiles of MSE, respectively.

log spectrum for model (9). Panels (b)–(d) are the estimated time varying log spectra corresponding to the 10th, 50th and 90th percentiles of the MSE values. The median of the MSE

values and their interquartile range are 0.06 and 0.04, respectively, which is indistinguishable from the results obtained using the DM estimator in Qin and Wang (2008). Their DM estimator was chosen for comparison because it gave the minimum MSE for this simulation setting. Although the median and IQR for the MSE values based on AdaptSPEC are very close to those obtained in Qin and Wang (2008), the DM estimator produces more variable estimates than those presented in this paper. We note that our comparable results were achieved without the assumption of stationarity. In contrast, the results for this setting in Qin and Wang (2008) assumed stationarity. The estimate $\widehat{\Pr}(m = 1|\boldsymbol{x})$ can be considered to be an estimate of the probability that the process is stationary. For this simulation, the median of this estimated probability is equal to 0.99 while the first and third quartiles are 0.93 and 1.00, respectively. These results indicate that if the true process is stationary, AdaptSPEC does not overfit by dividing the time series into more than one segment.

5.2   Comparison with Rosen et al. (2009)

In this section we compare the performance of AdaptSPEC, for $t_{\min}$ values of 20, 40 and 60, with the method of Rosen et al. (2009), which will be referred to as RSW09 in what follows. We simulate data from two processes used in Rosen et al. (2009), given by

$$x_t = a_t x_{t-1} - 0.81 x_{t-2} + \epsilon_t, \quad \text{for} \quad t = 1, \ldots, 1024, \tag{10}$$

where $a_t = 0.8(1 - 0.5\cos(\pi t/1024))$ and $\epsilon_t \overset{iid}{\sim} N(0,1)$, and

$$x_t = \begin{cases} 0.9x_{t-1} + \epsilon_t & \text{if} \quad 1 \leq t \leq 512 \\ 1.69x_{t-1} - 0.81x_{t-2} + \epsilon_t & \text{if} \quad 513 \leq t \leq 768 \\ 1.32x_{t-1} - 0.81x_{t-2} + \epsilon_t & \text{if} \quad 769 \leq t \leq 1024, \end{cases} \tag{11}$$

where $\epsilon_t \overset{iid}{\sim} N(0,1)$. Models (10) and (11) represent a slowly varying and a piecewise autoregressive processes, respectively. Panel (a) of Figure 7 presents boxplots of the MSE values for AdaptSPEC for $t_{\min} = 20, 40, 60$ and for RSW09 corresponding to model (10). Panel (b) of Figure 7 presents analogous boxplots corresponding to model (11). Figure 7 shows that the estimates of the log spectra obtained using AdaptSPEC are superior to those obtained using RSW09 for both the slowly varying AR(2) process as well as for the piecewise autoregressive process. The value of $t_{\min}$ has no significant effect on the MSE values.
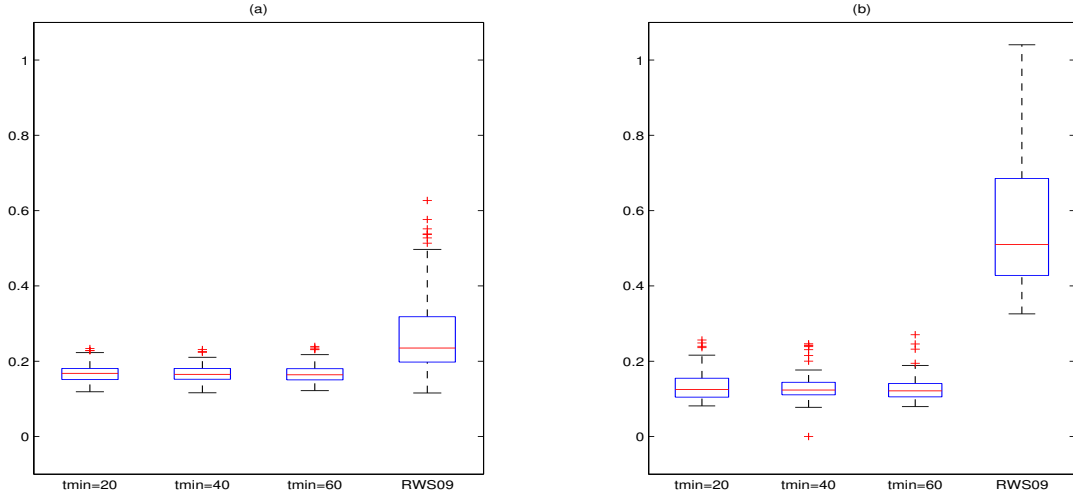
Figure 7: Panel (a): Boxplots of the MSE values corresponding to model (10) based on AdaptSPEC (for $t_{\min} = 20, 40, 60$), and on RSW09. Panel (b): Analogous boxplots corresponding to model (11).

5.3 Frequentist Properties of the Partition Estimate

In this section we focus on the estimation of the partition, that is, the number and location of the partition points. To this end, we use two piecewise autoregressive models, model (11) above and the following model from Wood et al. (2011)

$$
x_t = \begin{cases}
\sum_{k=1}^{6} \phi_{k1} x_{t-k} + \sigma_1 \epsilon_t^{(1)} & \text{for} \quad 1 \leq t \leq 200 \\
\sum_{k=1}^{6} \phi_{k2} x_{t-k} + \sigma_2 \epsilon_t^{(2)} & \text{for} \quad 201 \leq t \leq 1000 \\
\sum_{k=1}^{6} \phi_{k3} x_{t-k} + \sigma_3 \epsilon_t^{(3)} & \text{for} \quad 1001 \leq t \leq 1300 \\
\sum_{k=1}^{6} \phi_{k4} x_{t-k} + \sigma_4 \epsilon_t^{(4)} & \text{for} \quad 1301 \leq t \leq 1600 \\
\sum_{k=1}^{6} \phi_{k5} x_{t-k} + \sigma_5 \epsilon_t^{(5)} & \text{for} \quad 1601 \leq t \leq 2000,
\end{cases}
\tag{12}
$$

with parameter values given in Table 1. A realization from model (12) is displayed in Figure 8.

The simulation setting is as before except that in this section the maximum number of segments is set to 8. Figure 9, panel (a) presents the estimated posterior probability of the number of segments versus the number of segments for each of the 50 simulated samples from model (11). Panel (b) presents the analogous plot for model (12). It is evident that most of the 50 estimated posterior probabilities $\widehat{\Pr}(m = 3|X)$ and $\widehat{\Pr}(m = 5|X)$ under models (11) and (12), respectivley, are greater than 0.9. Figure 10, panel (a) displays density
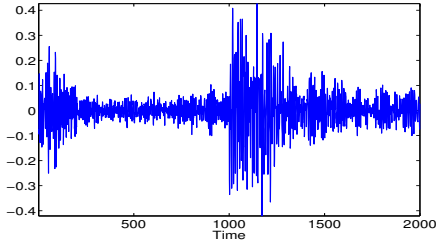
16

Figure 8: A realization from model (12).

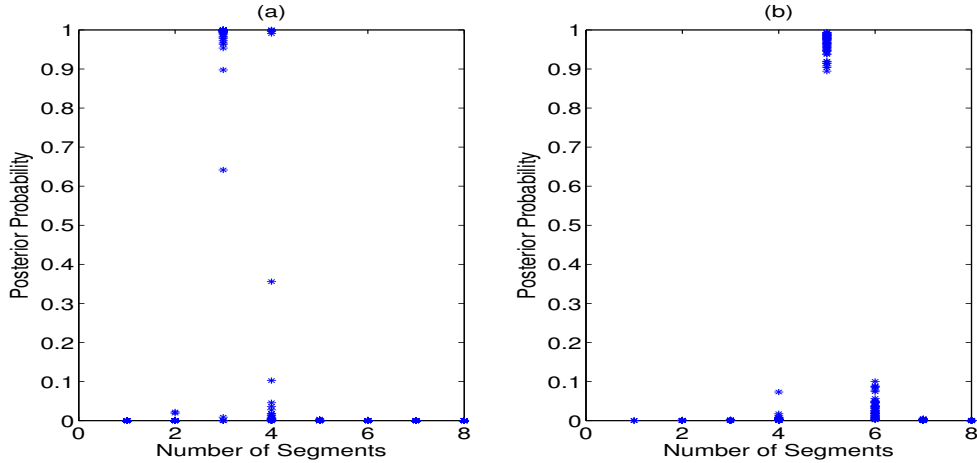| j	ϕ_{1j}	ϕ_{2j}	ϕ_{3j}	ϕ_{4j}	ϕ_{5j}	ϕ_{6j}	σ_j
1	0.89	-0.85	0.25	-0.65	0.32	-0.33	0.04
2	0.70	-0.55	0.315	-0.63	0.11	-0.103	0.02
3	1.34	-1.37	0.895	-0.96	0.58	-0.42	0.07
4	0.98	-0.86	0.43	-0.61	0.20	-0.16	0.03
5	0.80	-0.68	0.25	-0.57	0.17	-0.27	0.02

Table 1: Parameter values for model (12).

Figure 9: Panel (a): Posterior probability of the number of segments vs. the number of segments for each of the 50 samples from model (11). Panel (b): Analogous plot corresponding to model (12).
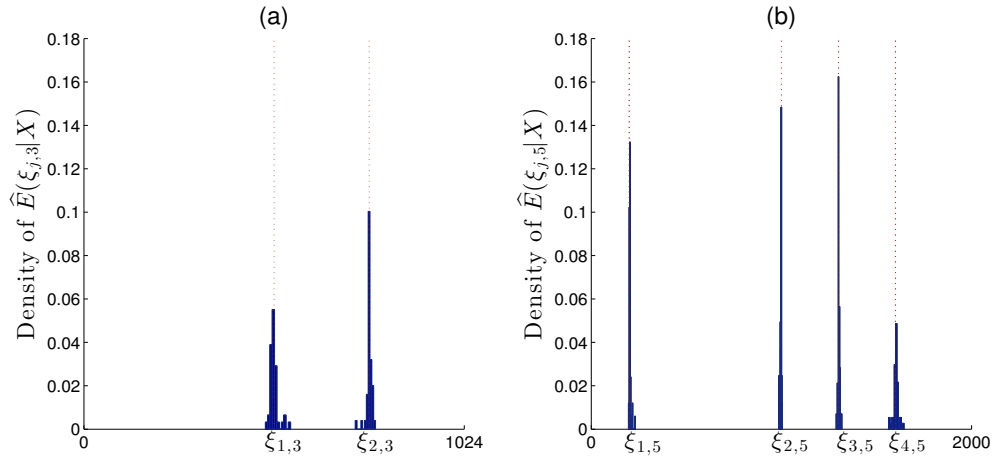

Figure 10: Panel (a): Density histograms of the posterior means of the partition points for a model with 3 segments, based on each of the 50 samples from model (11). The dotted vertical lines denote the true partition points. Panel (b): Analogous plot corresponding to model (12).

17

histograms of the estimated posterior means, $\hat{E}(\xi_{j,3}|X)$ for $j = 1, 2$, corresponding to model (11). The dotted vertical lines are plotted at $\xi_{1,3} = 512$ and $\xi_{2,3} = 768$. Panel (b) displays similar density histograms of $\hat{E}(\xi_{j,5}|X)$, $j = 1, 2, 3, 4$, corresponding to process (12), with $(\xi_{1,5}, \xi_{2,5}, \xi_{3,5}, \xi_{4,5}) = (200, 1000, 1300, 1600)$. Both panels show that our method correctly identifies the locations of the breaks in the time series.

6 Applications

In this section we apply our method of estimating the time varying spectrum to two examples. The first example is the intracranial EEG time series (IEEG) of an epileptic patient with medicine resistant mesial temporal lobe epilepsy and the second is the El Niño Southern Oscillation (ENSO) phenomenon.

6.1 IEEG

The analysis of IEEG time series in the period prior to the onset of an epileptic seizure, known as the preictal period, has been an active area of research over the last decade (Mormann et al., 2005). The motivation for this research has been to predict the onset of an epileptic seizure by developing methods which can distinguish between the preictal period and the interitcal period (the period between seizures) using IEEG time series. If this could be achieved, the benefits to patients who suffer from epilepsy would be enormous. Advance warning of an epileptic seizure would minimize injury and give sufferers a sense of control in their management of the disease.

Most approaches to seizure prediction attempt to identify a preictal state by comparing summary statistics from a univariate or multivariate IEEG time series at different user-specified time intervals and classifying the time interval as either a preictal or interictal state (Andrzejak et al., 2009). For example, Andrzejak et al. (2009) study a collection of bivariate features of an IEEG time series for windows of lengths one and five minutes.

Other approaches by Aksenova et al. (2007) and Temuçin et al. (2005) discriminate between

the preictal and interictal periods by monitoring changes in the spectral density of the IEEG time series rather than specific characteristics of the time series or spectrum because *spectral instabilities might carry more information for seizure anticipation than increases or decreases in a specific physical variable derived from EEG recordings* (Aksenova et al., 2007). Aksenova et al. (2007) assume that the IEEG time series is a locally stationary process with abrupt changes (piecewise stationary) and compute an index which measures the degree of instability of a process. If this index exceeds some individual specific predetermined threshold, they conclude that the preictal period has begun.

Modeling the time varying spectrum of IEEG time series captures both the instability of the spectrum across time as well as the difference in features of the spectrum at a given time. Qin and Wang (2008) is an example of such an approach. Qin and Wang (2008) assume that the IEEG time series is locally stationary and estimate the time varying spectrum non-parametrically using smoothing spline ANOVA, as in Guo et al. (2003). To estimate the smoothing parameters, Qin and Wang (2008) use generalized maximum likelihood and generalized approximate cross validation. They partition the data into 64 time blocks and select 32 equally spaced frequency points to compute the locally stationary periodograms. Guo et al. (2003) note that *one potential limitation of our approach is its smoothness assumption, which cannot handle abrupt jumps in the time varying spectrum.*

The data were collected by the EEG Lab of the University of Pennsylvania (D'Alessandro et al., 2001) and are shown in Figure 11. The data consist of five-minute interval IEEG time series from two channels taken at different times and were analyzed by Qin and Wang (2008). The signal was sampled at the rate of 200 signals per second so that each time series consists of 60,000 observations. Panels (a) and (c) of Figure 11 are the IEEG time series taken from channels 1 and 2, respectively, during the preictal period which in this example was taken to be five minutes prior to the onset of a known seizure. Panels (b) and (d) are similar plots of five-minute intervals of the IEEG time series extracted at least four hours prior to seizure onset and are therefore representative of the interictal period. Figure 12 shows the time varying spectra for the four time series. From figures 11 and 12 it can be seen that the spectra of the IEEG series for the two channels during the interictal periods are markedly different from the spectra during the preictal periods and different from each
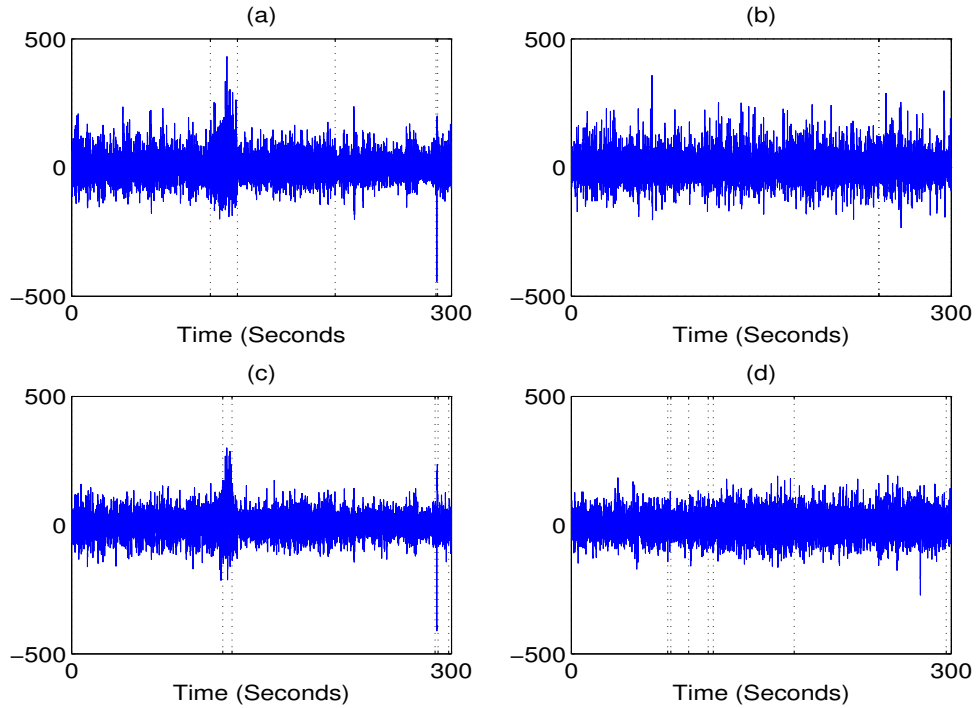
19

other.

Figure 11: Plots of the IEEG time series. Panel (a) is the IEEG time series for a five-minute time interval immediately prior to seizure onset for channel 1. Panel (c) is an analogous plot for channel 2. Panel (b) is a plot of a five-minute interval of the IEEG time series extracted at least four hours prior to seizure onset for channel 1, and panel (d) is an analogous plot for channel 2. The dotted lines represent the partition points corresponding to the modal number of segments for each time series.

Number of Segments	Preictal		Interictal	
	Channel 1	Channel 2	Channel 1	Channel 2
3	0.00	0.00	0.99	0.00
4	0.00	0.00	0.01	0.00
5	0.00	0.07	0.00	0.00
6	0.98	0.59	0.00	0.00
7	0.02	0.34	0.00	0.00
8	0.00	0.00	0.00	0.06
9	0.00	0.00	0.00	0.85
10	0.00	0.00	0.00	0.08
11	0.00	0.00	0.00	0.01

Table 2: Posterior probability of number of segments for two channels during the preictal period and the interictal period

Given the large number of observations in each time series, the maxiumum number of segments, M, was set at 12. Table 2 shows the posterior probability of the number of segments for each of the time series, while Figure 11 also shows the partition points (dotted lines) corresponding to the modal number of segments for each series. For both channels, the modal number of segments for the preictal period is 6.

The spectra for the initial locally stationary segment, ($t < 100$ secs), of the preictal period for both channels display characteristics typical of pre-seizure spectra documented in the literature. There is an increase in power at a frequency of 60 Hz and again at 85Hz. This is consistent with the findings of Niederhauser et al. (2003) who found that the frequency components in the range 25-60Hz appear before a seizure, and with Rampp and Stefan (2006) who found that there is a *specific association of high-frequency oscillations (in the range 80 - 500 Hz) with an epileptic network function*. The estimated spectra during this first segment are different from those estimated by Qin and Wang (2008) during a similar period. Our estimate exhibits very pronounced fluctuations in power at high frequency for both channels, while the estimate of Qin and Wang (2008) shows only slight fluctuations in power for channel 1 and no fluctuations in power for channel 2. This may be because the method of Qin and Wang (2008) oversmooths and therefore cannot capture these fluctuations.

The spectra for the segment $100 < t < 150$ secs, of the preictal period in both channels, (Figure 12, panels (a) and (c)), are distinguished from the spectra for the first segment by an increase in power at low frequencies (in the range 10-50 Hz). This is consistent with the findings of Qin and Wang (2008), although Qin and Wang (2008) refer to this increase in power at low frequencies as a *power build-up* and the plots of their time varying spectra indicate that this *power build-up* is gradual. In contrast, our method suggests that this increase occurs abruptly. From visual inspection of the time series, it appears that the change in the time series is abrupt rather than gradual, so that the gradual increase estimated in Qin and Wang (2008) may be due to oversmoothing.

The duration of this power surge at low frequencies during the preictal period is approximately twenty seconds for channel 1 and seven seconds for channel 2. Following this surge, the spectra for $t > 150$ secs, of the preictal period for both channels revert back to the

spectra of the initial locally stationary segment. Afterwards, there is a brief power surge at very low frequencies (< 10Hz) and high frequencies (>70 Hz) which occurs approximately 12 seconds before seizure onset. However, this second surge lasts for only one second.

The spectrum for the interictal period in channel 1, (Figure 12, panel (b)), does not have the fluctuations in power at high frequencies which characterize the spectra of the preictal period for both channels, and while the spectrum of channel 2, (Figure 12, panel (d)), shows slight fluctuations, these are not as pronounced as those for the preictal period. The spectrum for channel 1 only has one power surge which lasts for less than half a second before returning to its previous state. In contrast, the spectrum for channel 2 experiences a few small power surges between 4 to 2 minutes prior to the end of the time series. However, the power surges for the interictal periods occur at high frequencies, while those for preictal periods occur at low frequencies.

In summary, the spectra of the preictal period differ from those of the interictal period in three ways. First, spectra of the preictal period exhibit pronounced fluctuations in power at high frequencies, while spectra of the interictal period do not. Second, the spectra of the preictal period have large power surges at low frequencies while the spectra of the interictal period have smaller power surges at high frequency. Third, the duration of the power surges during the preictal period are longer than those during the interical period. The large number and short duration of locally stationary segments found in the interictal period is consistent with the findings in Kaplan et al. (2005).

6.2   ENSO

The National Oceanic and Atmospheric Administration (NAOO) defines the El Niño Southern Oscillation (ENSO) as a disruption of the ocean-atmosphere system in the Tropical Pacific having important consequences for weather and climate around the globe. In recent years there has been much research and debate on changes in the structure of ENSO. Many researchers have reported changes in the frequency of ENSO (Trenberth and Hoar, 1996, 1997; An and Wang, 2000) and the intensity of ENSO (Timmermann et al., 2004). The NOAA states on its website (http://www.ncdc.noaa.gov/oa/climate/globalwarming.html)
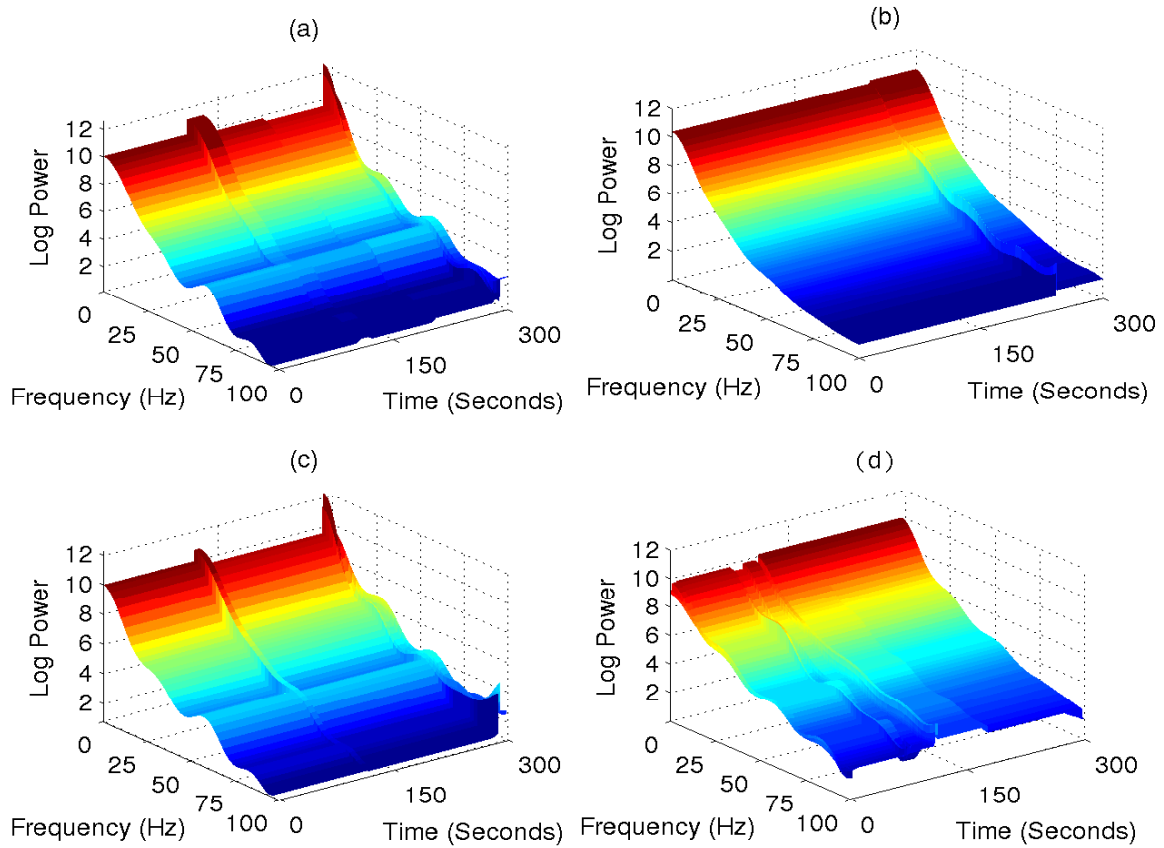
Figure 12: Plots of the time varying spectrum of the IEEG time series. Panel (a) is the time varying spectrum for the IEEG time series for a 5-minute time interval immediately prior to seizure onset for channel 1. Panel (c) is an analogous plot for channel 2. Panel (b) is a plot of the time varying spectrum for the IEEG time series for a 5-minute interval extracted at least 4 hours prior to seizure onset for channel 1 and panel (d) is an analogous plot for channel 2.

Number of Segments	SOI	Niño3.4	DSLPA
1	0.95	0.93	0.99
2	0.05	0.07	0.01
3	0.00	0.00	0.00

Table 3: Posterior probability of number of segments for the three indices, the SOI, Niño3.4 and DSLPA.

that *it is also true that El Niños have been more frequent and intense in recent decades.*

However, recent work by Solow (2006), Nicholls (2008) and Rosen et al. (2009) suggests that the frequency and intensity of ENSO have not changed over the last century. In this section we analyze the structure of ENSO by modeling the time varying spectrum of three indicators of ENSO. The first indicator is the Southern Oscillation Index (SOI). The SOI is the monthly standardized anomaly of the mean sea-level pressure difference between Tahiti and Darwin and is available from the Australian Bureau of Meteorology at http://www.bom.gov.au/climate/current/soihtm1.shtml. Reliable measurements of this index are available from 1876 to the present, and so we use the entire dataset in this example. The second indicator of ENSO is the Niño3.4 index which is the sea surface temperature (SST) averaged across the region 5S-5N, 120W-170W. From 1880–1950 the SST used to construct the Niño3.4 index was measured by buoys or passing ships but from 1950 onwards satellite measurements of the SST were used instead. There are many different versions of the Niño3.4 index, however, the differences among versions are very small after 1950. For this reason, we confine our analysis to this time period and use the Hadley Center SST dataset HadSST1. The third index is the mean sea level pressure at Darwin anomalies (DSLPA) from 1951 to 2010, available from the NOAA at http://www.cpc.ncep.noaa.gov/data/indices/darwin. This indicator was chosen solely because it is the indicator used by Trenberth and Hoar (1996), and we note that the SOI and Niño3.4 have been considered to be better indicators of ENSO than the DSLPA for some time (Chen, 1982). Plots of the three indices appear in Figure 13.

We set the maximum number of segments to be 4 for the three series and ran the sampling
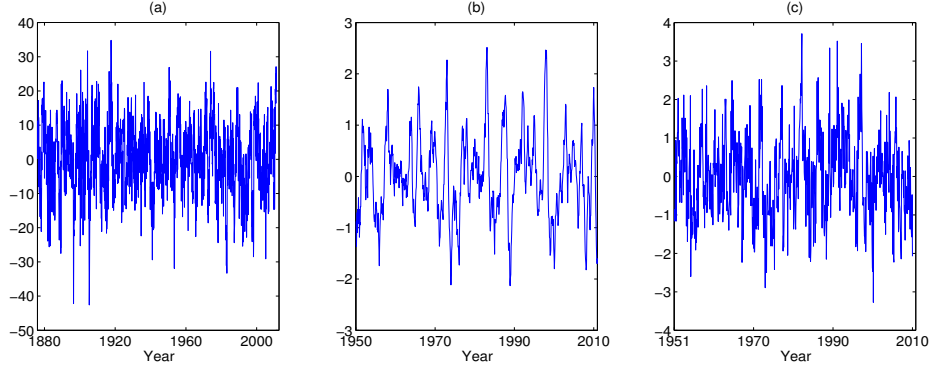
Figure 13: Plots of the SOI index from 1876–2011, panel (a), the Niño3.4 index from 1950-2011, panel (b) and the DSLPA from 1951–2010, panel (c).

scheme 6000 iterations with a burn-in of 2000 iterations. The posterior probability of the number of segments appears in Table 3. The time varying spectra of all three indices appear in Figure 14. Our results indicate that the spectra of the three indices are very similar and that they do not vary over time. Therefore it is very unlikely that there has been a change in the frequency or intensity of ENSO. The estimated posterior probabilities that the time series are stationary, i.e., $\widehat{\Pr}(m = 1|\boldsymbol{x})$ are 0.95, 0.93 and 0.99 for the SOI, Niño3.4 and DSLPA indices, respectively. These results confirm the findings of Rosen et al. (2009), Solow (2006) and Nicholls (2008).

One explanation for the difference between these findings and the earlier study of Trenberth and Hoar (1996) is that Trenberth and Hoar (1996) tested explicitly if there had been a change in frequency from 1981 onwards. In our model, explicitly testing for a change in frequency in the SOI or the Niño3.4 index from a specific time, t^*, is equivalent to assuming *a priori* that $\Pr(\xi_{12} = t^* \mid m = 2, \boldsymbol{x}) = 1$. This is a very strong prior belief, in effect the uncertainty surrounding the number of segments and the partition was ignored by Trenberth and Hoar (1996). In contrast, our methodology makes no assumptions regarding the number of segments or the partition. For a full discussion of why our results differ from those of Trenberth and Hoar (1996), see the discussion in Rosen et al. (2009). Given the results in this paper, in Rosen et al. (2009), Solow (2006) and Nicholls (2008), we suggest the NOAA update their website to reflect recent research findings.
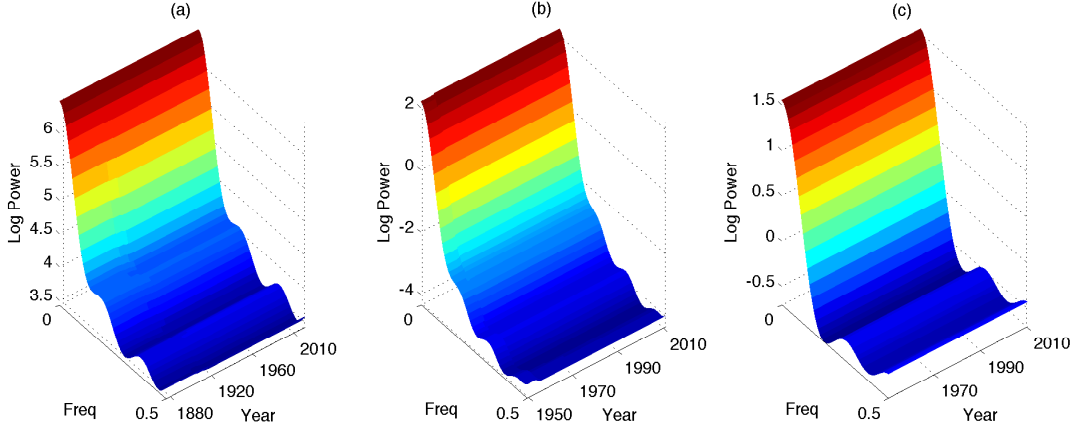
Figure 14: Time varying log spectrum of the SOI index from 1876–2011, panel (a), the Niño3.4 index from 1950-2011, panel (b) and the DSLPA from 1951–2010, panel (c).

Acknowledgements

Appendix: Details of the Sampling Scheme

As described in the paper, a partition of the time series into m segments is denoted by $\boldsymbol{\xi}_m = (\xi_{0,m}, \ldots, \xi_{m,m})$. Also, let $\boldsymbol{\tau}_m^2 = (\tau_{1,m}^2, \ldots, \tau_{m,m}^2)'$ and $\boldsymbol{\beta}_m = (\boldsymbol{\beta}'_{1,m}, \ldots, \boldsymbol{\beta}'_{m,m})'$ where $\boldsymbol{\beta}_{j,m}$ is a vector of unknown coefficients for the jth segment in a partition of m segments, for $j = 1, \ldots, m$. To simplify the notation, we assume that $\boldsymbol{\beta}_{j,m}$, $j = 1, \ldots, m$, includes $\alpha_{0j,m}$ as its first entry (see equation (3)). In what follows, current and proposed values are denoted by the superscripts c and p, respectively.

26

1. **Between-Model Moves:**

Let $\boldsymbol{\theta}_m = (\boldsymbol{\xi}'_m, \boldsymbol{\tau}^{2\prime}_m, \boldsymbol{\beta}'_m)'$ and suppose the chain is currently at $(m^c, \boldsymbol{\theta}^c_{m^c})$. We propose to move to $(m^p, \boldsymbol{\theta}^p_{m^p})$ by drawing $(m^p, \boldsymbol{\theta}^p_{m^p})$ from a proposal density $q(m^p, \boldsymbol{\theta}^p_{m^p} \mid m^c, \boldsymbol{\theta}^c_{m^c})$ and accepting this draw with probability

$$\alpha = \min\left\{1, \frac{p(m^p, \boldsymbol{\theta}^p_{m^p}|\boldsymbol{x}) \times q(m^c, \boldsymbol{\theta}^c_{m^c} \mid m^p, \boldsymbol{\theta}^p_{m^p})}{p(m^c, \boldsymbol{\theta}^c_{m^c}|\boldsymbol{x}) \times q(m^p, \boldsymbol{\theta}^p_{m^p} \mid m^c, \boldsymbol{\theta}^c_{m^c})}\right\},$$

where $p(\cdot)$ denotes a target density, which is the product of an approximate likelihood times prior densities. The specific forms of the target and proposal densities depend on the move type and are detailed below. We first outline the proposal density $q(m^p, \boldsymbol{\theta}^p_{m^p} \mid m^c, \boldsymbol{\theta}^c_{m^c})$.

$$
\begin{aligned}
q(m^p, \boldsymbol{\theta}^p_{m^p} \mid m^c, \boldsymbol{\theta}^c_{m^c}) &= q(m^p \mid m^c) \times q(\boldsymbol{\theta}^p_{m^p} \mid m^p, m^c, \boldsymbol{\theta}^c_{m^c}) \\
&= q(m^p \mid m^c) \times q(\boldsymbol{\xi}^p_{m^p}, \boldsymbol{\tau}^{2p}_{m^p}, \boldsymbol{\beta}^p_{m^p} \mid m^p, m^c, \boldsymbol{\theta}^c_{m^c}) \\
&= q(m^p \mid m^c) \times q(\boldsymbol{\xi}^p_{m^p} \mid m^p, m^c, \boldsymbol{\theta}^c_{m^c}) \times q(\boldsymbol{\tau}^{2p}_{m^p} \mid \boldsymbol{\xi}^p_{m^p}, m^p, m^c, \boldsymbol{\theta}^c_{m^c}) \\
&\quad \times q(\boldsymbol{\beta}^p_{m^p} \mid \boldsymbol{\tau}^{2p}_{m^p}, \boldsymbol{\xi}^p_{m^p}, m^p, m^c, \boldsymbol{\theta}^c_{m^c}).
\end{aligned}
$$

Thus, $(m^p, \boldsymbol{\theta}^p_{m^p})$ is drawn by first drawing m^p, followed by $\boldsymbol{\xi}^p_{m^p}$, $\boldsymbol{\tau}^{2p}_{m^p}$ and finally $\boldsymbol{\beta}^p_{m^p}$. Details on how each of these quantities is sampled are provided next.

(a) The number of segments, m^p, is proposed from $q(m^p|m^c)$. Let M be the maximum number of segments allowed, and $m^c_{2\min}$ be the current number of segments which contain at least $2\,t_{\min}$ observations, then

$$
q(m^p = k \mid m^c) = \begin{cases} 1/2 & \text{if } k = m^c - 1, m^c + 1 \text{ and } m^c \neq 1, M,\ m^c_{2\min} \neq 0 \\ 1 & \text{if } k = m^c - 1 \text{ and } m^c = M \text{ or } m^c_{2\min} = 0 \\ 1 & \text{if } k = m^c + 1 \text{ and } m^c = 1 \end{cases}
$$

(b) Conditional on m^p, a new partition, $\boldsymbol{\xi}^p_{m^p}$, a new vector of smoothing parameters, $\boldsymbol{\tau}^{2p}_{m^p}$, and a new vector of coefficients, $\boldsymbol{\beta}^p_{m^p}$, are then proposed as follows.

    i. **Birth:**

    Suppose $m^p = m^c + 1$, then

    A. A partition,

$$\boldsymbol{\xi}^p_{m^p} = (\xi^c_{0,m^c}, \ldots, \xi^c_{k^*-1,m^c}, \xi^p_{k^*,m^p}, \xi^c_{k^*,m^c}, \ldots, \xi^c_{m^c,m^c})$$

is proposed from $q(\boldsymbol{\xi}_{m^p}^p \mid m^p, m^c, \boldsymbol{\theta}_{m^c}^c)$. This partition is proposed by first selecting at random a segment $j = k^*$ to split. A point t^* within segment $j = k^*$ is then selected to be the proposed partition point, subject to the constraint that $\xi_{k^*-1,m^c}^c + t_{\min} \leq t^* \leq \xi_{k^*,m^c}^c - t_{\min}$. The proposal probability function is

$$
\begin{aligned}
q(\xi_{j,m^p}^p = t^* \mid m^p, m^c, \boldsymbol{\xi}_{m^c}^c) &= p(j = k^* \mid m^p, m^c, \boldsymbol{\xi}_{m^c}^c) \\
&\times p(\xi_{k^*,m^p}^p = t^* \mid j = k^*, m^p, m^c, \boldsymbol{\xi}_{m^c}^c) \\
&= \frac{1}{m_{2\min}^c} \times \frac{1}{n_{k^*,m^c} - 2t_{\min} + 1}.
\end{aligned}
$$

B. A vector of smoothing parameters

$$
\boldsymbol{\tau}_{m^p}^{2p} = (\tau_{1,m^c}^{2c}, \ldots, \tau_{k^*-1,m^c}^{2c}, \tau_{k^*,m^p}^{2p}, \tau_{k^*+1,m^p}^{2p}, \tau_{k^*+1,m^c}^{2c}, \ldots, \tau_{m^c,m^c}^{2c})
$$

is proposed from $q(\boldsymbol{\tau}_{m^p}^{2p} \mid m^p, \boldsymbol{\xi}_{m^p}^p, m^c, \boldsymbol{\theta}_{m^c}^c) = q(\boldsymbol{\tau}_{m^p}^{2p} \mid m^p, \boldsymbol{\tau}_{m^c}^{2c})$. We follow Green (1995) and propose the additional smoothing parameters for the newly split segment, τ_{k^*,m^p}^{2p} and τ_{k^*+1,m^p}^{2p}, by drawing $u \sim U[0,1]$ and letting τ_{k^*,m^p}^{2p} and τ_{k^*+1,m^p}^{2p} be deterministic functions of u and τ_{k^*,m^c}^{2c}. Specifically,

$$
\begin{aligned}
\tau_{k^*,m^p}^{2p} &= \tau_{k^*,m^c}^{2c} \times \frac{u}{1-u} \\
\tau_{k^*+1,m^p}^{2p} &= \tau_{k^*,m^c}^{2c} \times \frac{1-u}{u}.
\end{aligned}
$$

C. A vector of coefficients

$$
\boldsymbol{\beta}_{m^p}^p = (\boldsymbol{\beta}_{1,m^c}^{c\prime}, \ldots, \boldsymbol{\beta}_{k^*-1,m^c}^{c\prime}, \boldsymbol{\beta}_{k^*,m^p}^{p\prime}, \boldsymbol{\beta}_{k^*+1,m^p}^{p\prime}, \boldsymbol{\beta}_{k^*+1,m^c}^{c\prime} \ldots, \boldsymbol{\beta}_{m^c,m^c}^{c\prime})'
$$

is proposed from $q(\boldsymbol{\beta}_{m^p}^p \mid \boldsymbol{\tau}_{m^p}^{2p}, \boldsymbol{\xi}_{m^p}^p, m^p, m^c, \boldsymbol{\theta}_{m^c}^c) = q(\boldsymbol{\beta}_{m^p}^p \mid \boldsymbol{\tau}_{m^p}^{2p}, \boldsymbol{\xi}_{m^p}^p, m^p)$. The pair of vectors $\boldsymbol{\beta}_{k^*,m^p}^p$ and $\boldsymbol{\beta}_{k^*+1,m^p}^p$ are drawn from normal approximations to their posterior conditional distributions $p(\boldsymbol{\beta}_{k^*,m^p}^p \mid \boldsymbol{x}_{k^*}^p, \tau_{k^*,m^p}^{2p}, m^p)$ and $p(\boldsymbol{\beta}_{k^*+1,m^p}^p \mid \boldsymbol{x}_{k^*+1}^p, \tau_{k^*+1,m^p}^{2p}, m^p)$, where $\boldsymbol{x}_{k^*}^p$ and $\boldsymbol{x}_{k^*+1}^p$ denote the subsets of the time series belonging to segment k^* and k^*+1 respectively, see equation (3). Note that $\boldsymbol{\xi}_{m^p}^p$ determines $\boldsymbol{x}_*^p = (\boldsymbol{x}_{k^*}^{p\prime}, \boldsymbol{x}_{k^*+1}^{p\prime})'$, and so $\boldsymbol{\xi}_{m^p}^p$ has

been suppressed here. For example, $\boldsymbol{\beta}^p_{k^*,m^p}$ is drawn from $N(\boldsymbol{\beta}^{\max}_{k^*}, \Sigma^{\max}_{k^*})$, where $\boldsymbol{\beta}^{\max}_{k^*} = \arg\max\limits_{\boldsymbol{\beta}^p_{k^*,m^p}} p(\boldsymbol{\beta}^p_{k^*,m^p} \mid \boldsymbol{x}^p_{k^*}, \tau^{2p}_{k^*,m^p}, m^p)$ and

$$\Sigma^{\max}_{k^*} = \left\{ -\frac{\partial^2 \log p(\boldsymbol{\beta}^p_{k^*,m^p} \mid \boldsymbol{x}^p_{k^*}, \tau^{2p}_{k^*,m^p}, m^p)}{\partial \boldsymbol{\beta}^p_{k^*,m^p} \partial \boldsymbol{\beta}^{p\prime}_{k^*,m^p}} \Big|_{\boldsymbol{\beta}^p_{k^*,m^p}=\boldsymbol{\beta}^{\max}_{k^*}} \right\}^{-1}.$$

The acceptance probability for the birth move is $\alpha = \min\{1, A\}$, where

$$
\begin{aligned}
A &= \frac{p(\boldsymbol{\theta}^p_{m^p} \mid \boldsymbol{x}, m^p)p(\boldsymbol{\theta}^p_{m^p} \mid m^p)p(m^p)}{p(\boldsymbol{\theta}^c_{m^c} \mid \boldsymbol{x}, m^c)p(\boldsymbol{\theta}^c_{m^c} \mid m^c)p(m^c)} \\
&\quad \times \frac{p(m^c \mid m^p)p(\boldsymbol{\beta}^c_{k^*,m^c})}{p(m^p \mid m^c)p(\xi^{m^p}_{k^*,m^p} \mid m^p, m^c)p(u)p(\boldsymbol{\beta}^p_{k^*,m^p})p(\boldsymbol{\beta}^p_{k^*+1,m^p})} \times \left| \frac{\partial(\tau^{2p}_{k^*,m^p}, \tau^{2p}_{k^*+1,m^p})}{\partial(\tau^{2c}_{k^*,m^c}, u)} \right|,
\end{aligned}
$$

where $p(u) = 1$, $0 \leq u \leq 1$, $p(\boldsymbol{\beta}^p_{k^*,m^p})$ and $p(\boldsymbol{\beta}^p_{k^*+1,m^p})$ are the Gaussian proposal densities $N(\boldsymbol{\beta}^{\max}_{k^*}, \Sigma^{\max}_{k^*})$ and $N(\boldsymbol{\beta}^{\max}_{k^*+1}, \Sigma^{\max}_{k^*+1})$, respectively, and the Jacobian is

$$\left| \frac{\partial(\tau^{2p}_{k^*,m^p}, \tau^{2p}_{k^*+1,m^p})}{\partial(\tau^{2c}_{k^*,m^c}, u)} \right| = \frac{2\tau^{2c}_{k^*,m^c}}{u(1-u)} = 2(\tau^p_{k^*,m^p} + \tau^p_{k^*+1,m^p})^2.$$

ii. **Death:**

If $m^p = m^c - 1$, then the reverse of a birth move is performed.

A. A partition

$$\boldsymbol{\xi}^p_{m^p} = (\xi^c_{0,m^c}, \dots, \xi^c_{k^*-1,m^c}, \xi^c_{k^*+1,m^c}, \dots, \xi^c_{m^c,m^c})$$

is proposed by selecting one of $m^c - 1$ partition points to remove. Let $j = k^*$ be the partition point selected for removal. Among m^c segments there are $m^c - 1$ partition points available for removal. Our proposal is to make each partition point equally likely so that

$$q(\xi^p_{j,m^p} \mid m^p, m^c, \boldsymbol{\xi}^c_{m^c}) = \frac{1}{m^c - 1}.$$

B. A vector of smoothing parameters,

$$\boldsymbol{\tau}^{2p}_{m^p} = (\tau^{2c}_{1,m^c}, \dots, \tau^{2c}_{k^*-1,m^c}, \tau^{2p}_{k^*,m^p}, \tau^{2c}_{k^*+2,m^c}, \dots, \tau^{2c}_{m^c,m^c})$$

is proposed from $q(\boldsymbol{\tau}_{m^p}^{2p} \mid m^p, \boldsymbol{\xi}_{m^p}^{p}, m^c, \boldsymbol{\theta}_{m^c}^{c}) = q(\boldsymbol{\tau}_{m^p}^{2p} \mid m^p, \boldsymbol{\tau}_{m^c}^{2c})$. A single smoothing parameter, τ_{k^*,m^p}^{2p}, is formed from τ_{k^*,m^c}^{2c} and τ_{k^*+1,m^c}^{2c} by reversing the process described in step (b) i B, i.e.,

$$\tau_{k^*,m^p}^{2p} = \sqrt{\tau_{k^*,m^c}^{2c}\tau_{k^*+1,m^c}^{2c}}.$$

C. A vector of coefficients

$$\boldsymbol{\beta}_{m^p}^{p} = (\boldsymbol{\beta}_{1,m^c}^{c\prime}, \ldots, \boldsymbol{\beta}_{k^*-1,m^c}^{c\prime}, \boldsymbol{\beta}_{k^*,m^p}^{p\prime}, \boldsymbol{\beta}_{k^*+2,m^c}^{c\prime}, \ldots, \boldsymbol{\beta}_{m^c,m^c}^{c\prime})'$$

is proposed from $q(\boldsymbol{\beta}_{m^p}^{p} \mid \boldsymbol{\tau}_{m^p}^{2p}, \boldsymbol{\xi}_{m^p}^{p}, m^p, m^c, \theta_{m^c}^{c}) = q(\boldsymbol{\beta}_{m^p}^{p} \mid \boldsymbol{\tau}_{m^p}^{2p}, \boldsymbol{\xi}_{m^p}^{p}, m^p)$. A single vector of coefficients, $\boldsymbol{\beta}_{k^*,m^p}^{p}$, is drawn from a normal approximation to its posterior distribution, $p(\boldsymbol{\beta}_{k^*,m^p}^{p} \mid \boldsymbol{x}, \tau_{k^*,m^p}^{2p}, \boldsymbol{\xi}_{m^p}^{p}, m^p)$, as in step (b) i C.

The acceptance probability is the inverse of that of the birth move. If the move is accepted then $m^c = m^p$ and $\boldsymbol{\theta}_{m^c}^{c} = \boldsymbol{\theta}_{m^p}^{p}$.

2. **Within-Model Moves:**

For this type of move, m is fixed, and so the notation indicating the dependence on the number of segments is dropped. Within-model moves consist of two parts; first, a segment relocation move is performed, and then, conditional on the relocation, all the basis function coefficients are updated. The two steps, jointly, are either accepted or rejected via a M-H step. The smoothing parameters are then updated via a Gibbs step.

(a) Suppose the chain is at $\boldsymbol{\theta}^{c} = (\boldsymbol{\xi}^{c}, \boldsymbol{\beta}^{c})$; we propose to move to $\boldsymbol{\theta}^{p} = (\boldsymbol{\xi}^{p}, \boldsymbol{\beta}^{p})$ as follows.

i. Select a partition point, ξ_{k^*}, to relocate from $m-1$ possible partition points. Then select a position in the interval $[\xi_{k^*-1}, \xi_{k^*+1}]$, subject to the constraint that the new location is at least $t_{\min}$ from each of ξ_{k^*-1} and ξ_{k^*+1}, so that

$$\Pr(\xi_{k^*}^{p} = t) = \Pr(j = k^*) \times \Pr(\xi_{k^*}^{p} = t \mid j = k^*), \qquad (A.1)$$

where $\Pr(j = k^*) = (m-1)^{-1}$.

In order to explore the parameter space efficiently, we construct a mixture distribution for $\Pr(\xi^p_{k^*} = t \mid j = k^*)$, so that

$$\Pr(\xi^p_{k^*} = t \mid j = k^*) = \pi \, q_1(\xi^p_{k^*} = t \mid \xi^c_{k^*}) + (1 - \pi) \, q_2(\xi^p_{k^*} = t \mid \xi^c_{k^*}), \quad \text{(A.2)}$$

where $q_1(\xi^p_{k^*} = t \mid \xi^c_{k^*}) = (n_{k^*} + n_{k^*+1} - 2t_{\min} + 1)^{-1}$, $\xi_{k^*-1} + t_{\min} \le t \le \xi_{k^*+1} - t_{\min}$, and

$$q_2(\xi^p_{k^*} = t \mid \xi^c_{k^*}) = \begin{cases} 0 & \text{if } |t - \xi^c_{k^*}| > 1 \\ 1/3 & \text{if } |t - \xi^c_{k^*}| \le 1, \ n_{k^*} \ne t_{\min} \text{ and } n_{k^*+1} \ne t_{\min} \\ 1/2 & \text{if } t - \xi^c_{k^*} \le 1, \ n_{k^*} = t_{\min} \text{ and } n_{k^*+1} \ne t_{\min} \\ 1/2 & \text{if } \xi^c_{k^*} - t \le 1, \ n_{k^*} \ne t_{\min} \text{ and } n_{k^*+1} = t_{\min} \\ 1 & \text{if } t = \xi^c_{k^*}, \ n_{k^*} = t_{\min} \text{ and } n_{k^*+1} = t_{\min}. \end{cases}$$

As can be seen, the support of q_1 has $n_{k^*} + n_{k^*+1} - 2t_{\min} + 1$ time points, while that of q_2 has at most three. Using q_2 alone results in a relatively high acceptance rate of the M-H step but explores the parameter space too slowly. Adding the q_1 component allows bigger jumps which in turn leads to faster exploration of the parameter space. Fixing π at a relatively small value, say 0.2, combines a relatively high acceptance rate with fast exploration of the parameter space.

ii. Draw $\boldsymbol{\beta}^p_j$, $j = k^*, k^* + 1$, from an approximation to $\prod_{j=k^*}^{k^*+1} p(\boldsymbol{\beta}_j \mid \boldsymbol{x}^p_j, \tau^2_j)$, as in step 1 (b) (i) C.

The proposal density, evaluated at $\boldsymbol{\beta}^p_j$, $j = k^*, k^* + 1$, is

$$q(\boldsymbol{\beta}^p_* \mid \boldsymbol{x}^p_*, \boldsymbol{\tau}^2_*) = \prod_{j=k^*}^{k^*+1} q(\boldsymbol{\beta}^p_j \mid \boldsymbol{x}^p_j, \tau^2_j),$$

where $\boldsymbol{\beta}^p_* = (\boldsymbol{\beta}^{p\prime}_{k^*}, \boldsymbol{\beta}^{p\prime}_{k^*+1})'$ and $\boldsymbol{\tau}^2_* = (\tau^2_{k^*}, \tau^2_{k^*+1})'$. In a similar fashion, the proposal density is evaluated at the current values of $\boldsymbol{\beta}^c_* = (\boldsymbol{\beta}^{c\prime}_{k^*}, \boldsymbol{\beta}^{c\prime}_{k^*+1})'$. The draw $\boldsymbol{\beta}^p_*$ is accepted with probability

$$\alpha = \min \left\{ 1, \frac{p(\boldsymbol{x}^p_* \mid \boldsymbol{\beta}^p_*) p(\boldsymbol{\beta}^p_* \mid \boldsymbol{\tau}^2_*) q(\boldsymbol{\beta}^c_* \mid \boldsymbol{x}^c_*, \boldsymbol{\tau}^2_*)}{p(\boldsymbol{x}^c_* \mid \boldsymbol{\beta}^c_*) p(\boldsymbol{\beta}^c_* \mid \boldsymbol{\tau}^2_*) q(\boldsymbol{\beta}^p_* \mid \boldsymbol{x}^p_*, \boldsymbol{\tau}^2_*)} \right\},$$

where $\boldsymbol{x}^c_* = (\boldsymbol{x}^{c\prime}_{k^*}, \boldsymbol{x}^{c\prime}_{k^*+1})'$. If the draw is accepted, then $(\xi^c_{k^*}, \boldsymbol{\beta}^c_*) = (\xi^p_{k^*}, \boldsymbol{\beta}^p_*)$.

(b) Draw $\boldsymbol{\tau}^{2p}$ from

$$p(\boldsymbol{\tau}^2_* \mid \boldsymbol{\beta}_*) = \prod_{j=k^*}^{k^*+1} p(\tau^2_j \mid \boldsymbol{\beta}_j)$$

31

and accept with probability 1. Equation (4) gives the density of $p(\tau^2 \mid \boldsymbol{\beta})$ without the proportionality constant.

References

Adak, S. (1998), "Time-dependent Spectral Analysis of Nonstationary Time Series," *Journal of the American Statistical Association*, 93, 1488–1501.

Aksenova, T., Volkovych, V., and Villa, A. (2007), "Detection of spectral instability in EEG recordings during the preictal period," *Journal of Neural Engineering*, 4, 173–178.

An, S.-I. and Wang, B. (2000), "Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency," *Journal of Climate*, 13, 2044–2055.

Andrzejak, R., Chicharro, D., Elger, C., and Mormann, F. (2009), "Seizure prediction: any better than chance?" *Clinical Neurophysiology*, 120, 1465–1478.

Carter, C. and Kohn, R. (1997), "Semiparametric Bayesian inference for time series with mixed spectra," *Journal of the Royal Staistical Society B*, 59, 255–268.

Chen, W. (1982), "Assessment of southern oscillation of sea-level pressure indices," *Monthly Weather Review*, 110, 800–807.

Chiann, C. and Morettin, P. (1999), "Estimation of time-varying linear systems," *Statististical Inference for Stochastic Processes*, 2, 253–285.

Dahlhaus, R. (1997), "Fitting time series models to nonstationary processes," *Annals of Statistics*, 25, 1–37.

D'Alessandro, M., Vachtsevanos, G., Esteller, R., Echauz, J., and Litt, B. (2001), "A generic approach to selecting the optimal feature for epileptic seizure prediction," *IEEE International Meeting of the Engineering in Medicine and Biology Society*.

Davis, R., Lee, T., and Rodriguez-Yam, G. (2006), "Structural breaks estimation for nonstationary time series models," *Journal of the American Statistical Association*, 101, 223–239.

Eubank, R. (1999), *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker, 2nd ed.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 1–19.

Gerlach, R., Carter, C., and Kohn, R. (2000), "Efficient Bayesian inference for dynamic mixture models," *Journal of the American Statistical Association*, 95, 819–828.

Green, P. (1995), "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Guo, W., Dai, M., Ombao, H., and von Sachs, R. (2003), "Smoothing spline ANOVA for time-dependent spectral analysis," *Journal of the American Statistical Association*, 98, 643–652.

Kaplan, A. Y., Fingelkurts, A. A., Borisov, S. V., and Darkhovsky, B. S. (2005), "Nonstationary nature of the brain activity as revealed by EEG/MEG: methodological, practical and conceptual challenges," *Signal Process.*, 85, 2190–2212.

Kitagawa, G. and Akaike, H. (1978), "A procedure for the modeling of nonstationary time series," *Annals of the Institute of Statistical Mathematics*, 30, 351–363.

Kitagawa, G. and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, Springer-Verlag Inc.

Lau, J. and So, M. (2008), "Bayesian Mixture of autoregressive models," *Computational Statistics and Data Analysis*, 53, 38–60.

Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R., Kraskov, A., David, P., Elger, C., and Lehnertz, K. (2005), "On the predictability of epileptic seizures," *Clinical Neurophysiology*, 116, 569–587.

Nicholls, N. (2008), "Recent trends in the seasonal and temporal behaviour of the El Niño–Southern Oscillation," *Geophysical Research Letters*, 35, L19703.

Niederhauser, J., Esteller, R., Echauz, J., Vachtsevanos, G., and Litt, B. (2003), "Detection of seizure precursors from depth-EEG using a sign periodogram transform," *IEEE Transcactions of Biomedical Engineering*, 50, 449–458.

Ombao, H., Raz, J., Von Sachs, R., and Malow, B. (2001), "Automatic statistical analysis of bivariate nonstationary time series," *Journal of the American Statistical Association*, 96, 543–560.

Prado, R. and Huerta, G. (2002), "Time-varying autoregressions with model order uncertainty," *Journal of Time Series Analysis*, 23, 599–618.

Qin, L. and Wang, Y. (2008), "Nonparametric spectral analysis with applications to seizure characterization using EEG time series," *The Annals of Applied Statistics*, 2, 1432–1451.

Rampp, S. and Stefan, H. (2006), "Fast activity as a surrogate marker of epileptic network function?" *Clinical Neurophysiology*, 117, 2111–2117.

Rosen, O., Wood, S., and Stoffer, D. (2009), "Local spectral analysis via a Bayesian mixture of smoothing splines," *Journal of the American Statistical Association*, 104, 249–262.

Solow, A. (2006), "An ENSO shift revisited," *Geophysical Research Letters*, 33, L22602.

Temuçin, C., Tokçaer, A., and Bilir, E. (2005), "Detection of EEG background abnormalities in epilepsy by a new spectral index," *Clinical Neurophysiology*, 116, 933–947.

Timmermann, A., Jin, F.-F., and Collins, M. (2004), "Intensification of the annual cycle in the tropical pacific due to greenhouse warming," *Geophysical Research Letters*, 31, L12208.

Trenberth, K. and Hoar, T. (1996), "The 1990-1995 El Niño– Southern Oscillation event: longest on record," *Geophysical Research Letters*, 23, 57–60.

— (1997), "El Niño and climate change," *Geophysical Research Letters*, 24, 3057–3060.

Wahba, G. (1980), "Automatic smoothing of the log periodogram," *Journal of the American Statistical Association*, 75, 122–132.

— (1990), *Spline Models for Observational Data*, vol. 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, Philadelphia: SIAM.

West, M., Prado, R., and Krystal, A. (1999), "Evaluation and comparison of EEG traces: latent structure in non-stationary time series," *Journal of the American Statistical Association*, 94, 1083–1095.

Whittle, P. (1957), "Curve and periodogram smoothing," *Journal of the Royal Statistical Society B*, 19, 38–47.

Wood, S., Rosen, O., and Kohn, R. (2011), "Bayesian mixtures of autoregressive models," *Journal of Computational and Graphical Statistics*, 20, 174–195.

Wood, S. A., Jiang, W., and Tanner, M. (2002), "Bayesian mixture of splines for spatially adaptive nonparametric regression." *Biometrika*, 89, 513 – 528.