Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Local Spectral Analysis via a Bayesian Mixture of Smoothing Splines

Ori Rosen, David S. Stoffer & Sally Wood

To cite this article: Ori Rosen, David S. Stoffer & Sally Wood (2009) Local Spectral Analysis via a Bayesian Mixture of Smoothing Splines, Journal of the American Statistical Association, 104:485, 249-262, DOI: 10.1198/jasa.2009.0118

To link to this article: https://doi.org/10.1198/jasa.2009.0118

Local Spectral Analysis via a Bayesian Mixture of Smoothing Splines

Ori ROSEN, David S. STOFFER, and Sally WOOD

In many practical problems, time series are realizations of nonstationary random processes. These processes can often be modeled as processes with slowly changing dynamics or as piecewise stationary processes. In these cases, various approaches to estimating the time-varying spectral density have been proposed. Our approach in this article is to estimate the log of the Dahlhaus local spectrum using a Bayesian mixture of splines. The basic idea of our approach is to first partition the data into small sections. We then assume that the log spectral density of the evolutionary process in any given partition is a mixture of individual log spectra. We use a mixture of smoothing splines model with time varying mixing weights to estimate the evolutionary log spectrum. The mixture model is fit using Markov chain Monte Carlo techniques that yield estimates of the log spectra of the individual subsections. In addition to an estimate of the local log spectral density, the method yields pointwise credible intervals. We use a reversible jump step to automatically determine the number of different spectral components.

KEY WORDS: Evolutionary spectra; Locally stationary time series; Mixture of splines; Reversible jump Markov chain Monte Carlo.

1. INTRODUCTION

In many practical problems, the spectral analysis of time series that are realizations of nonstationary random processes is of interest. For example, analyzing the Southern Oscillation Index (SOI) time series is important in monitoring global climate change and predicting rainfall. Priestley (1965) was the first to introduce the concept of evolutionary spectra and a Cramér representation with a time-varying transfer function. This idea was later refined by Dahlhaus (1997) where he established an asymptotic framework for locally stationary processes. The definition of a Dahlhaus locally stationary process will be given in the next section, but throughout the remainder of this section, we use the term locally stationary in a generic sense.

Several authors have considered the estimation of locally stationary processes under a variety of assumptions. The estimators that were developed by Dahlhaus 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. Various other approaches have been suggested to overcome the computational difficulty. Adak (1998) proposed a nonparametric methodology based on segmenting time in such as way as to allow a data dependent choice of window lengths. Ombao, Raz, Von Sachs, and Malow (2001) proposed nonparametric estimators based on smooth local exponential functions. Guo, Dai, Ombao, and von Sachs (2003) extended the work of Ombao et al. (2001) to allow for simultaneous smoothing in both the time and frequency domains. Our primary goal is to establish a Bayesian approach for estimating the time-varying spectrum of a Dahlhaus locally stationary process in a semiautomatic fashion. We will make our goal more precise in the next section.

Another approach to analyzing locally stationary time series is to consider fitting piecewise autoregressive (AR) models; this approach was suggested by Kitagawa and Akaike (1978). Recently Davis, Lee, and Rodriguez-Yam (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. If an investigator is certain the dynamics generating the data are locally stationary AR models, then we would suggest using a parametric approach. If, however, an investigator is uncertain about the dynamics of the process, a nonparametric procedure might be preferred. Indeed, some authors, for example Thomson (1990) in analyzing locally stationary geophysical series, warn strongly that AR spectra can be misleading.

One of the difficulties of the non-Bayesian approaches is that they do not jointly model the uncertainty surrounding the point estimates of the spectra and the point estimates of the number and location of the locally stationary segments. Thus, to make inference regarding the individual locally stationary spectra, plug-in point estimates of the number and location of the locally stationary segments are used. Such an approach underestimates the uncertainty surrounding estimates of the locally stationary spectra.

Bayesian approaches to modeling nonstationary time series include Lavielle (1998) and Punskaya, Andrieu, Doucet, and Fitzgerald (2002). Punskaya et al. (2002) proposed a Bayesian method for fitting piecewise linear regression models like AR models. Their method is based on placing prior distributions on the number of change points, their locations, and the order of the linear regression in each segment. The method is implemented via reversible jump Markov chain Monte Carlo

Ori Rosen is Associate Professor, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX 79968 (E-mail: *ori@math.utep.edu*). David S. Stoffer is Professor, Department of Statistics, University of Pittsburgh, Pittsburgh, PA, 15260 (E-mail: *stoffer@pitt.edu*) and Program Director, Division of Mathematical Sciences, National Science Foundation, 4201 Wilson Blvd., Arlington, VA 22230. Sally Wood is Associate Professor, Melbourne Business School, University of Melbourne, Victoria, 3053, Australia (E-mail: *sallyw@unsw.edu.au*). The work of Ori Rosen and David Stoffer was supported, in part, by the National Science Foundation through grants DMS-0706752, DMS-0804140 (to Rosen), and DMS-0706723, DMS-0805050 (to Stoffer). The authors thank Professor Rob McLean for his suggestion to examine the stationarity of the Southern Oscillation Index and the editors and reviewers for their helpful comments and suggestions.

Journal of the American Statistical Association
March 2009, Vol. 104, No. 485, Theory and Methods
DOI 10.1198/jasa.2009.0118

(MCMC) methods.Unlike the work of Punskaya et al. (2002), our approach is nonparametric.

There are four main contributions of our article. First, we present a novel method of modeling the time varying log spectral density using a mixture of a finite but unknown number of individual log spectra, where the mixture weights of the individual log spectra vary with time. The basic idea of our approach is to first partition the data into small sections. We then assume that in any given partition, the log spectral density of the process is a mixture of individual log spectra. We introduce nonstationarity into our model by allowing the mixing weights of the individual log spectra to change smoothly across partitions.

The second contribution is that we perform model averaging rather than model selection by allowing the number of underlying individual spectra to be unknown and to vary from $j = 1, \ldots, J$. To this end, we construct a MCMC scheme using the reversible jump method of Green (1995) to move between models of varying numbers of components. Thus, our resulting estimate of the time varying log spectra is averaged across models of varying numbers of components.

The third contribution of our article is that we use the output from the MCMC scheme not only to estimate but also to make inference regarding the time varying log spectra, the number of underlying locally stationary spectra, and the location and rate at which the time series changes from one locally stationary process to another.

Finally, segmentation methods, such as those mentioned in our introduction, are entirely local in that estimates are based only on the data in the segment (with perhaps some small overlap from contiguous segments). Our method uses all of the data in the estimation and inference procedures. For example, in Section 7.2, we will analyze seismic data. In those examples, the signals are amplitude modulated wherein the strength of a signal may vary over time but the frequencies remain constant. Any existing segmentation method would ignore this fact and simply base its estimates on the data in each segment. Our method will recognize the fact that there are similarities in the signals over the entire sample and will use the information in all of the data to estimate the evolutionary spectrum.

2. LOCALLY STATIONARY TIME SERIES

The Dahlhaus (1997) approach to modeling evolutionary spectra is based on a time-dependent analog to the classical Cramér spectral representation of stationary processes. The following definition is for the case of a zero-mean process.

Definition 1 (Dahlhaus). A sequence of zero-mean stochastic processes, $\{X_{t,N};\ t = 1, \ldots, N\}$, for $N \geq 1$, is called **locally stationary** with transfer function A^0 if there exists a representation

$$X_{t,N} = \int_{-1/2}^{1/2} A^0_{t,N}(\nu)\, \exp(i2\pi\nu t)\, dZ(\nu), \tag{1}$$

where

(1) $Z(\nu)$ is a zero mean orthogonal increment process on $[-1/2, 1/2]$;

(2) There exists a positive constant K and a smooth function A: $[0, 1] \times [-1/2, 1/2] \to \mathbb{C}$ with $A(u, \nu) = \overline{A(u, -\nu)}$ such that for all N,

$$\sup_{t,\nu} |A^0_{t,N}(\nu) - A(t/N, \nu)| \leq K N^{-1}; \tag{2}$$

(3) For all ν, $A(u, \nu)$ is continuous in u. The Dahlhaus evolutionary spectrum can then be defined as follows.

Definition 2 (Dahlhaus). The **evolutionary spectrum** of a Dahlhaus locally stationary process at time $u \in [0, 1]$ and frequency $\nu \in [-1/2, 1/2]$ is given by $f(u, \nu) = |A(u, \nu)|^2$.

We refer the reader to Dahlhaus (1997) for examples of processes satisfying Definition 1. Note that the first argument of $A(u, \nu)$ is rescaled to live on the unit interval and increasing the number of observations, N, means we allow for more data to be observed locally in the sense of infill asymptotics, rather than in the sense of increasing domain asymptotics (i.e., observing more data in the future). Equation (2) deals with the smoothness of A in u such that it is allowed to change only slowly over time. The idea behind this is essentially that for each fixed N, we implicitly assume some local interval of stationarity about each time point and a smooth change from one interval to the next.

Because a Dahlhaus locally stationary process changes smoothly in time, u, we may approximate the evolutionary spectrum by a piecewise stationary process as follows. For a fixed N, consider a partition of the unit interval into S segments of equal length chosen in such a way that $n = N/S$ is small relative to N. We denote the corresponding intervals by $I_s = (((s-1)/S), (s/S)]$, for $s = 1, \ldots, S$. Next, we define the piecewise stationary process

$$\widetilde{X}_{t,N} = \sum_{s=1}^{S} X_{s,t'}\, \delta(t/N, I_s), \tag{3}$$

where $t' = 1, \ldots, n$ [i.e., $t' = (t-1) \pmod n + 1$], and where $\delta(t/N, I_s) = 1$ if $t/N \in I_s$ and $\delta(t/N, I_s) = 0$ if $t/N \notin I_s$. The stationary process $X_{s,t'}$ is defined to have a spectral density $\widetilde{f}_s(\nu)$, which is given by averaging the evolutionary spectrum in that interval; i.e.,

$$\widetilde{f}_s(\nu) = \frac{1}{S} \int_{u \in I_s} f(u, \nu)\, du \tag{4}$$

for $s = 1, \ldots, S$.

In the asymptotics, we may let $S \to \infty$, but $S/N \to 0$ as $N \to \infty$. Hence, because of the smoothness of $A(u, \nu)$ in u, the local spectral density of $\widetilde{X}_{t,N}$, namely $\widetilde{f}_s(\nu)$ for $t/N \in I_s$, and $s = 1, \ldots, S$, well approximates the evolutionary spectrum, $f(u, \nu)$, of $X_{t,N}$, provided N and S are sufficiently large. In addition, Ombao et al. (2001, Theorem 1) argued that under mild conditions, locally stationary processes can be well approximated by piecewise stationary processes in the sense that the average mean square error between $X_{t,N}$ and $\widetilde{X}_{t,N}$ is $O(n^2/N^2)$.

Our general technique relies on the approximation given by (3)–(4). To make our technique easier to understand, however, we first describe the case for a stationary process in the next section. After the stationary case has been discussed, we will then describe our technique for the Dahlhaus locally stationary case in the subsequent sections.

3. SPECTRAL ESTIMATION FOR STATIONARY TIME SERIES

As previously indicated, our general 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.

3.1 Model

Suppose that a stationary time series, $\{X_t\}$, has a spectral density given by $f(\nu)$, for $-1/2 < \nu \leq 1/2$. We assume that $f(\nu)$ is bounded and positive. 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 $\boldsymbol{x} = (x_1, \ldots, x_n)$ can be approximated as

$$p(\boldsymbol{x}|f) \propto \prod_{k=0}^{n-1} \exp\left\{-\frac{1}{2}[\log f(\nu_k) + I_n(\nu_k)/f(\nu_k)]\right\} . \qquad (5)$$

Let $y_n(\nu_k)$ be the log of the periodogram evaluated at the Fourier frequencies, then the representation in (5) suggests the log-linear model

$$y_n(\nu_k) = \log f(\nu_k) + \epsilon_k, \qquad (6)$$

where $y_n(\nu_k) = \log I_n(\nu_k)$ for $k = 0, \ldots, [n/2]$, where $[n/2]$ is the largest integer less than or equal $n/2$, 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 $k = 0$, $[n/2]$. Note that in (5), there are only $[n/2] + 1$ distinct observations because the spectral density and the periodogram are both even functions of ν. For ease of notation, in what follows, we assume that n is even.

It is seen from (6) that the log spectral density can be estimated nonparametrically with the log periodogram as the dependent variable. Wahba (1980) used a frequentist approach for estimating the log spectral density via cubic smoothing splines. Carter and Kohn (1997) used a Bayesian approach to modeling the log spectral density as cubic smoothing splines as well by expressing Equation (6) in a state-space form. They approximated the error distribution in (6) by a mixture of five normal distributions and introduced latent component indicators to facilitate the estimation.

3.2 Priors

Let $g(\nu_k) = \log f(\nu_k)$. 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

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

where $h(\nu_k)$ is the nonlinear component. We place a smoothing spline prior on $h(\nu)$, which means that $h(\nu)$ is a zero mean Gaussian process with variance-covariance matrix $\tau^2\Omega$, where the ijth element of Ω is given by

$$\operatorname{cov}(h(\nu_i), h(\nu_j)) = \tau^2 \omega_{ij},$$

where τ^2 is the smoothing parameter and

$$\omega_{ij} = \nu_i^2(\nu_j - \nu_i/3)/2, \ \nu_i \leq \nu_j.$$

The parameters α_0 and α_1 are the value of the log of the spectral density and its first derivative at $\nu = 0$, respectively. The symmetry and periodicity of the spectral density mean that $(\partial \log f(\nu)/\partial\nu)|_{\nu=0} = 0$. Accordingly, α_1 is set to be identically zero and the prior on α_0 is $N(0, c_\alpha)$, for some large c_α. To complete the prior specification on $g(\nu)$, we assume $p(\tau^2) \propto 1/\tau^2$.

To facilitate the computations we follow Wood, Jiang, and Tanner (2002) and write $\boldsymbol{h} = (h(\nu_0), \ldots, h(\nu_{n/2}))'$ as a linear combination of basis functions so that $\boldsymbol{h} = X\boldsymbol{\beta}$, where the columns of the design matrix X are the Demmler-Reinsch basis functions evaluated at the Fourier frequencies and $\boldsymbol{\beta}$ is an $(n/2 + 1) \times 1$ vector of regression coefficients. These basis functions are constructed by decomposing Ω as $\Omega = QDQ'$, where Q is the matrix of eigenvectors of Ω, and D is a diagonal matrix containing the eigenvalues of Ω. Letting $X = QD^{1/2}$ and setting the prior on $\boldsymbol{\beta}$ to be $p(\boldsymbol{\beta}) \sim N(0, \tau^2 I_{(n/2+1)})$, we have $X\boldsymbol{\beta} \sim N(0, \tau^2\Omega)$ as required. In practice, one can retain only the columns of X corresponding to the largest eigenvalues of Ω resulting in computational saving without affecting the fit. The model in (6) can now be expressed as

$$\mathbf{y} = \alpha_0 \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{1}$ is an $(n/2 + 1) \times 1$ vector of ones and $\boldsymbol{\epsilon} = (\epsilon_0, \ldots, \epsilon_{n/2})'$ with the ϵ_k's distributed as in (6). Note that this is the same model as in Carter and Kohn (1997). Both this article and Carter and Kohn (1997) assume a smoothing spline prior on the log spectral density. We chose a different representation of this prior for ease of computation and to avoid the introduction of the latent component indicators used by Carter and Kohn (1997).

4. SPECTRAL ESTIMATION FOR LOCALLY STATIONARY TIME SERIES

In this section, we generalize the spectral analysis of stationary time series described in Section 3 to time series that are Dahlhaus locally stationary as defined in Section 2. Given observations $\{X_{t,N};\ t = 1, \ldots, N\}$ from a Dahlhaus locally stationary process with evolutionary spectrum $f(u,\nu)$, we consider the approximation process $\{\widetilde{X}_{t,N};\ t = 1, \ldots, N\}$ defined in (3). In particular, given a segmentation S, we form the partitioned data $\{X_{s,t'};\ t' = 1, \ldots, n;\ s = 1, \ldots, S\}$ as described in (3). Our goal is to use the partitioned data to estimate the local spectra, $\widetilde{f}_s(\nu)$, as specified in (4), for $s = 1, \ldots, S$ and to use the estimated local spectra as an estimate of the evolutionary spectrum $f(u, \nu)$. The choice of the number of segments, S, will be discussed after we present the model, but the basic idea is that the finer the partition of the unit time interval, the better the estimate of the evolutionary spectrum.

One obvious but primitive approach to this analysis is to estimate the local spectra, $\widetilde{f}_s(\nu)$, individually in each segment $s = 1, \ldots, S$ using the techniques presented in Section 3. Although this approach may be viable if, in fact, the data are Dahlhaus locally stationary and changing slowly over time, we

would like to have an approach that is more robust to departures from this assumption. For example, if the process of interest is piecewise stationary, its time-varying spectra are not smooth and may make jumps at certain time points. If we segment such a process in such a way that a particular partition contains a jump, assuming the data are stationary in that segment will lead to a bad estimate. Another problem with the primitive method, as with all of the segmentation methods previously mentioned, is that they do not use the information from all of the data (or even data in contiguous segments), but rather only from the data in that segment. To overcome these drawbacks, our strategy is to model evolutionary spectra as a mixture of an unknown but finite number of spectra. We make these ideas precise in the following subsection.

4.1 Model

Given a segmentation S, our goal is the estimation of $g_s(\nu) = \log \widetilde{f}_s(\nu)$. To this end, we first calculate the periodogram corresponding to each segment. Let $\mathbf{y}_s = (y_{s,0}, \ldots, y_{s,n/2})'$ be the log-periodogram for segment s, $s = 1, \ldots, S$, evaluated at the Fourier frequencies. We model these observations as

$$y_{s,k} = g_s(\nu_k) + \epsilon_k,$$

where $g_s(\nu)$ is the log of the spectral density for segment s, for $s = 1, \ldots S$, and the ϵ_k's for $k = 0, \ldots, [n/2]$ are distributed as in (6). We model $g_s(\nu)$ as a mixture of an unknown but finite number of spectra so that

$$g_s(\nu) = \sum_{j=1}^{J} g_{js}(\nu) \Pr(j),$$

where J is the maximum number of components, $\Pr(j)$ is the prior probability that the mixture contains j components, and $g_{js}(\nu)$ is the log of the spectral density of a mixture of j components in segment s. For a given number of mixture components j and segment s we model $g_{js}(\nu)$ as

$$g_{js}(\nu) = \sum_{r=1}^{j} \pi_{rjs} \log f_{rj}(\nu) \ ,$$

where $f_{rj}(\nu)$ is the spectral density of the rth component and π_{rjs} is the unknown weight assigned to the rth component in segment s, with $\sum_{r=1}^{j} \pi_{rjs} = 1$. Note that the spectral density $f_{rj}(\nu)$ is common to all segments. The value of π_{rjs} represents the probability that in a mixture of j components, the data in segment s have spectral density $f_{rj}(\nu)$. A key point to note is that these probabilities are parameterized to depend upon the segment s and are modeled using a multinomial logistic regression to be specified in Section 4.5. This means that although the component spectra are common to all segments, a time varying estimate of the spectral density is obtained by allowing the weights of the common spectra to change across segments.

4.2 Segmentation

It is important to note that the choice of the number of segments, S, is not crucial to our estimation process subject to certain constraints. In theory there are potentially as many segments as there are data points. However, practically, we need a minimum number of observations in each segment to estimate the spectral density and for the Whittle approximation to the likelihood to hold. In simulations we found that using a minimum of 64 observations in each segment gave reasonable results (i.e., the estimates are qualitatively close to the actual values, if the true local spectra have well separated peaks). If an investigator is interested in a finer grid of frequencies, however, then he or she she may choose a larger (but relatively small) number of observations in a segment.

Many nonparametric local techniques, including the adaptive techniques that rely on orthogonal libraries and use entropy-based basis algorithms, such as the Best Basis Algorithm of Coifman and Wickerhauser (1992), use an arbitrary maximum level of segmentation of the unit interval to initialize the local analysis. In addition, most of these techniques use dyadic segmentation for ease; by dyadic segmentation we mean intervals of [0, 1] of the form $(k2^{-j}, (k+1)2^{-j})$, $k = 0, \ldots, 2^j - 1$, for levels $j = 1, 2, \ldots J$, where J is the maximum level. Although our method also requires picking a maximal segmentation through the choice of S, it is not crucial that the segmentation be dyadic.

Finally, but perhaps most importantly, is that the parameters of the mixing function π_{rjs} in our model are of more importance than the number of segments because these parameters control the location and rate at which the time series moves from one stationary process to another. The prior on these parameters is discussed in Section 4.5, and the sampling scheme used to explore the posterior distributions of these parameters is described in Section 5.

4.3 An Example

To illustrate our proposed method, consider a time series of length 1024 generated from the following piecewise stationary model

$$x_t = \begin{cases} 0.9x_{t-1} + \epsilon_t & \text{if} \quad 1 \le t \le 450 \\ -0.9x_{t-1} + \epsilon_t & \text{if} \quad 451 \le t \le 1024, \end{cases}$$

where $\epsilon_t \sim N(0, 1)$. For illustrative purposes, suppose we know that there are two components (i.e., $\Pr(j = 2) = 1$). We first divide the time series into nonoverlapping segments each containing $n = 64$ data points. This gives a total of $S = 16$ segments. Our estimate of the log spectra in segment s for $s = 1, \ldots 16$ is

$$g_{2s}(\nu) = \pi_{12s} \log f_{12}(\nu) + (1 - \pi_{12s}) \log f_{22}(\nu). \qquad (7)$$

Figure 1 shows the true (solid line) and estimated (dashed line) log spectral density for $x_t = 0.9x_{t-1} + \epsilon_t$ (left panel) and for $x_t = -0.9x_{t-1} + \epsilon_t$ (right panel), $\epsilon_t \sim N(0, 1)$. Figure 2 plots the estimated mixing function π_{12s} as a function of the segment. This figure shows that the probability that the data have spectral density $f_{12}(\nu)$ is close to 0.91 at the beginning of the time series. This probability decreases to 0.5 by the seventh segment and is approximately 0.03 by the end of the time series.

4.4 Prior on the Component Spectra

As in Section 3.2, we express $\log f_{rj}(\nu)$ as

$$\log f_{rj}(\nu_k) = \alpha_{0rj} + h_{rj}(\nu_k)$$

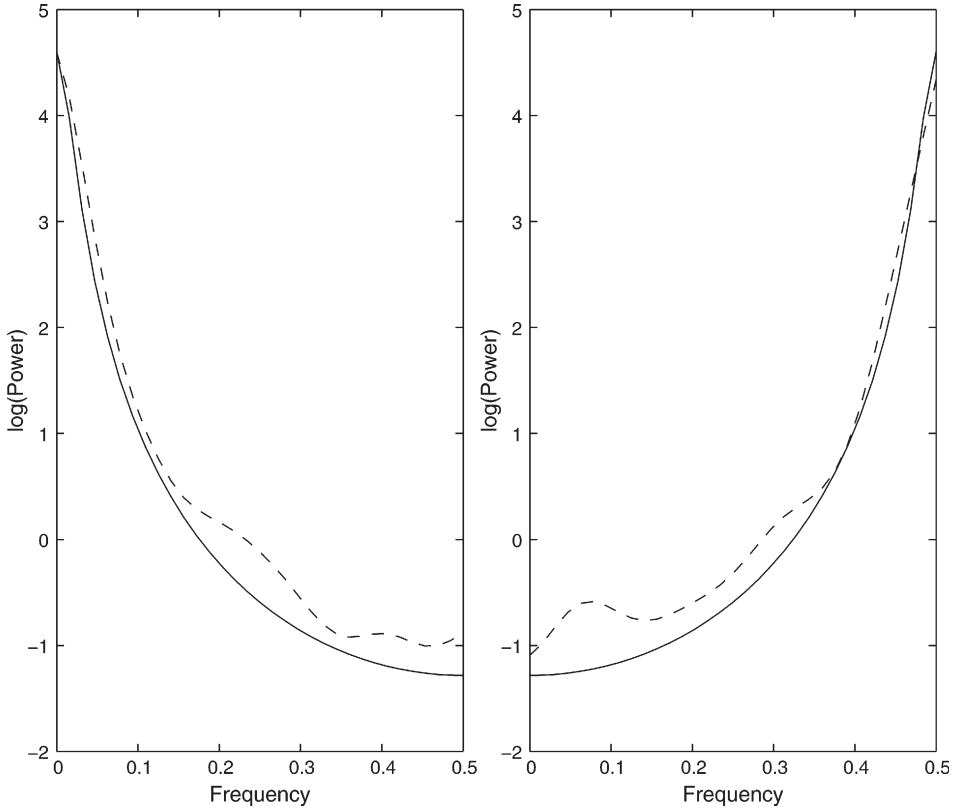

Figure 1. True (*solid lines*) and estimates (*dashed lines*) of the log spectral density. The left panel shows $\log(f_{12})$, which is the log of the spectral density of the time series $x_t = 0.9x_{t-1} + \epsilon_t$. The right panel shows $\log(f_{22})$, the log of the spectral density for $x_t = -0.9x_{t-1} + \epsilon_t$.

and write $\boldsymbol{h}_{rj} = X\boldsymbol{\beta}_{rj}$. The priors on $\boldsymbol{\beta}_{rj}$, and on α_{0rj} for $r = 1, \ldots, j$ and $j = 1, \ldots, J$ are as given in Section 3.2, and are assumed to be independent across all r and j. The prior on τ^2_{rj} is as in Section 3.2, and in addition we impose an ordering on τ^2_{rj} for $r = 1, \ldots, j$ and $j = 1, \ldots, J$, so that for a given j, $\tau^2_{1j} > , \ldots, > \tau^2_{jj}$. This ensures that the likelihood is identified.

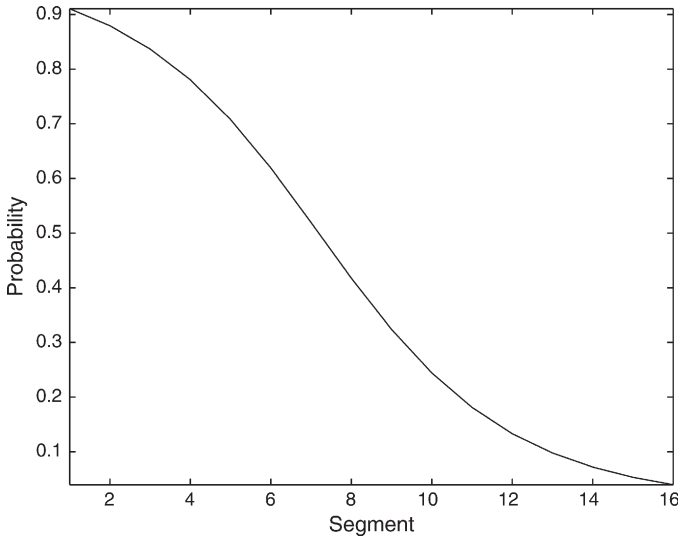

Figure 2. Estimate of the mixing function π_{12s} in (7) based on 16 segments.

4.5 Prior on the Mixing Probabilities

The mixing probabilities are expressed using the multinomial linear logit model so that

$$\pi_{rjs} = \frac{\exp(\boldsymbol{\delta}'_{rj}\mathbf{u}_s)}{\sum_{h=1}^{j}\exp(\boldsymbol{\delta}'_{hj}\mathbf{u}_s)} \tag{8}$$

with parameters $\boldsymbol{\delta}_{rj}$, $r = 1, \ldots, j$ and $j = 1, \ldots, J$. In (8), $\mathbf{u}_s = (1, u_s)'$, where the covariate u_s is taken as $u_s = s/S$, and $\boldsymbol{\delta}_{rj} = (\delta_{0rj}, \delta_{1rj})'$. For identifiability, $\boldsymbol{\delta}_{1j}$ is set to zero. Such logistic weights are also used in the mixtures-of-experts model (Jacobs, Jordan, Nowlan, and Hinton 1991). The priors on $\boldsymbol{\delta}_{rj}$ for $r = 1, \ldots, j$ and $j = 1, \ldots, J$ are bivariate normal with zero mean and variance $\sigma^2_\delta I_2$ and are assumed independent across all r and j. In all of our analyses, σ^2_δ was equal to 4.

4.6 Prior on the number of components, j

Typically, we assume a priori that the maximum number of components is J and that $\Pr(j = k) = 1/J$ for $k = 1, \ldots, J$. In the simulated examples of Section 6, J was set to 10, whereas in the real examples of Section 7, J was set to 20. We do, however, consider the use of Poisson priors in the example of Section 7.1.

5. BAYESIAN INFERENCE

We estimate the log of the spectral density in segment s, for $s = 1, \ldots S$ by its posterior mean $E(g_s|\mathbf{y})$, with all unknown parameters integrated out and we use MCMC to perform the required multidimensional integration.

The expectation $E\{g_s(\nu)|\mathbf{y}\}$ is defined as

$$E\{g_s(\nu)|\mathbf{y}\} = \sum_{j=1}^{J} \int E\{g_s(\nu)|\mathbf{y}, \boldsymbol{\theta}_j, j\} p(\boldsymbol{\theta}_j|\mathbf{y}, j) d\boldsymbol{\theta}_j \Pr(j|\mathbf{y})$$

where $\boldsymbol{\theta}_j = (\boldsymbol{\alpha}_j', \boldsymbol{\beta}'_j, \boldsymbol{\tau}'_j, \boldsymbol{\delta}'_j)'$, $\boldsymbol{\alpha}_j = (\alpha_{01}, \ldots, \alpha_{0j})$, $\boldsymbol{\beta}_j = (\boldsymbol{\beta}'_{1j}, \ldots, \boldsymbol{\beta}'_{jj})'$, $\boldsymbol{\tau}_j = (\tau^2_{1j}, \ldots, \tau^2_{jj})'$, and $\boldsymbol{\delta}_j = (\boldsymbol{\delta}'_{2j}, \ldots, \boldsymbol{\delta}'_{jj})$. This integral cannot be evaluated explicitly and we use MCMC simulation to estimate it. In addition to the point estimates, we construct $(1 - \alpha)$-level pointwise credible intervals for the log spectra by obtaining the $\alpha/2$ and $1 - \alpha/2$ percentiles of the MCMC fitted log spectra based on all the iterates after the burn-in period. Note that these credible intervals reflect the uncertainty surrounding not only our estimate of $\log f_{rj}$ but also our uncertainty surrounding the number of components J and the mixing probabilities π_{rsj}.

To simplify the simulation from the posterior distribution $p(\boldsymbol{\theta}_j, j|\mathbf{y})$, we introduce latent variables that are generated during the simulation. The first of these is the number of components j that are active at any point in the simulation. Then given j define the vector of indicator variables $\boldsymbol{\gamma}_{srj}$ for $s = 1, \ldots, S$ and $r = 1, \ldots, j$, where $\boldsymbol{\gamma}_{srj} = 1$ if $\mathbf{y}_s$ originated from the rth component, and $\boldsymbol{\gamma}_{srj} = 0$, otherwise.

5.1 Sampling Scheme

We outline here the MCMC scheme for our model. More details are given in the Appendix. The sampling scheme consists of two parts; a between-model move followed by a within-model move. The number of components j is first initialized, and then conditional on this value, the other model parameters $\boldsymbol{\alpha}_j$, $\boldsymbol{\beta}_j$, $\boldsymbol{\tau}_j$, and $\boldsymbol{\delta}_j$ are initialized.

1. **Between Model Move**
A new value of j is proposed, and conditional on this value, parameter values for $\boldsymbol{\alpha}_j$, $\boldsymbol{\beta}_j$, $\boldsymbol{\tau}_j$, and $\boldsymbol{\delta}_j$ are proposed. These proposed values are then accepted or rejected using a Metropolis-Hastings step.
2. **Within Model Move**
Given the value of j, the parameters specific to a model of j components are then updated as follows.
(1) Let $\boldsymbol{\beta}^*_{rj} = (\alpha_{rj}, \boldsymbol{\beta}'_{rj})'$, $r = 1, \ldots, j$, $\boldsymbol{\beta}^*_j = (\boldsymbol{\beta}^{*'}_{1j}, \ldots, \boldsymbol{\beta}^{*'}_{jj})'$, and $X^* = (\boldsymbol{1}, X)$. Generate $\boldsymbol{\beta}^*_j$ from $p(\boldsymbol{\beta}^*_j|\boldsymbol{\tau}_j, \boldsymbol{\gamma}_j, X^*, \tilde{\mathbf{y}})$ via Metropolis-Hastings steps where $\boldsymbol{\gamma}_j = \{\gamma_{srj}\}$, for $r = 1, \ldots, j$ and $s = 1, \ldots, S$, are the component indicators, and $\tilde{\mathbf{y}} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_S)'$.
(2) Generate $\boldsymbol{\tau}_j$ from $p(\boldsymbol{\tau}_j \mid \beta_j)$.
(3) Generate $\boldsymbol{\delta}_j$ from $p(\boldsymbol{\delta}_j \mid \gamma_j, U)$ via a Metropolis-Hastings step, where U is the matrix whose sth row is $\mathbf{u}'_s$, $s = 1, \ldots, S$.
(4) Let $\gamma_{sj} = r$ if $\gamma_{srj} = 1$. Generate the component indicators from $p(\gamma_{sj} = r|\boldsymbol{\beta}^*_j, \boldsymbol{\delta}_j, X^*, \mathbf{y}_s)$.The proposal densities for generating $\boldsymbol{\beta}^*_j$ and $\boldsymbol{\delta}_j$ are multivariate normal. The resulting acceptance rates are around 30% and 80%, respectively.

6. SIMULATIONS

We conducted a simulation study using three settings; the first two are taken from Ombao et al. (2001). The first is a slowly varying AR(2) process, the second is a piecewise stationary model, and the third is a time varying AR(6) process. We generated 200 time series from the slowly varying AR(2) process and 50 time series from each of the other two processes. For all three settings, each time series is of length $N = 1024$, divided into $S = 16$ segments of size $n = 64$. The Gibbs sampler for our model was then run for each time series with 100,000 iterations, including a burn-in period of 20,000 iterations. A run of 100,000 iterations took about an hour on a 3.4 GHz Linux PC, using Fortran 90 code.

6.1 Slowly Varying AR(2) Process

In this simulation setting, 200 realizations are generated from the model

$$x_t = a_t x_{t-1} - 0.81 x_{t-2} + \epsilon_t, t = 1, \ldots, 1024, \qquad (9)$$

where $a_t = 0.8(1 - 0.5\cos(\pi t/1024))$ and $\epsilon_t \overset{iid}{\sim} N(0, 1)$. In this case, the process is neither piecewise stationary, nor does it have a dyadic structure. Figure 3 shows a realization of the process (9) and a plot of a_t. Note that a_t changes slowly as a function of t, which in turn results in a slowly varying spectrum as a function of t.

Figure 4 presents the true time-varying log spectrum of process (9) on the left, as well as the average of the 200 estimated log spectra, on the right. In this figure, darker shades correspond to higher power, and time has been rescaled to the unit interval. Each of these estimated log spectra is obtained as described in Section 5. We see that our method does a good job, on average, in estimating the true time-varying spectrum of the process. Following Ombao et al. (2001), we use the averaged squared error (ASE) to assess the distance between the true log spectrum and its estimate. In our notation, ASE is given by

$$\text{ASE} = \{N(n/2 + 1)\}^{-1} \times \sum_{t=1}^{N} \sum_{k=0}^{n/2} \left\{\log \widehat{f}(t/N, \nu_k) - \log f(t/N, \nu_k)\right\}^2, \qquad (10)$$

where $f(\cdot, \cdot)$ is the true time-varying spectrum, and

$$\log \widehat{f}(t/N, \nu_k) = \sum_{s=1}^{S} \widehat{g}_s(\nu_k)\, \delta(t/N, I_s),$$

where $\widehat{g}_s(\nu_k)$ the estimate of the log of time-varying spectrum $g_s(\nu_k)$ as discussed in Section 5 and $\delta(t/N, I_s)$ is defined below (3). The average of the 200 ASE values corresponding to the 200 realizations was 0.261 with a standard deviation of 0.093. The 10th, 25th, 50th, 75th, and 90th percentiles of the 200 ASE values were 0.17, 0.20, 0.23, 0.31, and 0.39, respectively. Figure 5 displays, from left to right, the individual log-spectrum estimates corresponding to the 10th, 50th, and 90th percentiles of the 200 ASE values.

To compare our method with the AutoSLEX method presented in Ombao et al. (2001), we repeated the study using AutoSLEX. One of the inputs to the AutoSLEX method is the maximum level of the dyadic segmentation, which determines the maximum number of segments. In this simulation, we set the maximum level to 4, which corresponds to segments of size $n = 64$. The average of the 200 ASE values in this case was 0.759 with a standard deviation of 0.070. The 10th, 25th, 50th, 75th, and 90th percentiles of the 200 ASE values were 0.68, 0.72, 0.75, 0.80, and 0.84, respectively. (We note that these

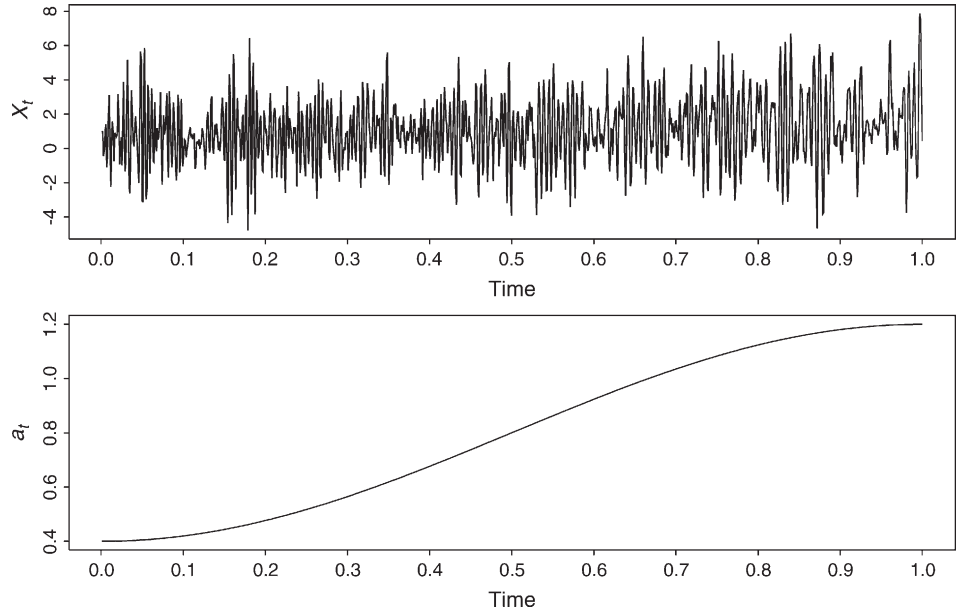

Figure 3. *Top*: A realization of the slowly varying AR(2) process given in (9). *Bottom*: The coefficient a_t as a function of t for process (9). In both figures time is rescaled to the unit interval, N = 1,024.

results are in conflict with those reported in Ombao et al. 2001, however, our results are based on the AutoSLEX Matlab programs obtained from one of the authors. The calculations of ASE for our method and for AutoSLEX were performed in an identical manner.) It is evident that our method shows a substantial improvement over AutoSLEX.

6.2 Piecewise Stationary Process

Although we assume local stationarity, our technique will also work in the case that the process is piecewise stationary. The following is an example of the viability of our method in this case. In this simulation setting, the time series are generated from the following model

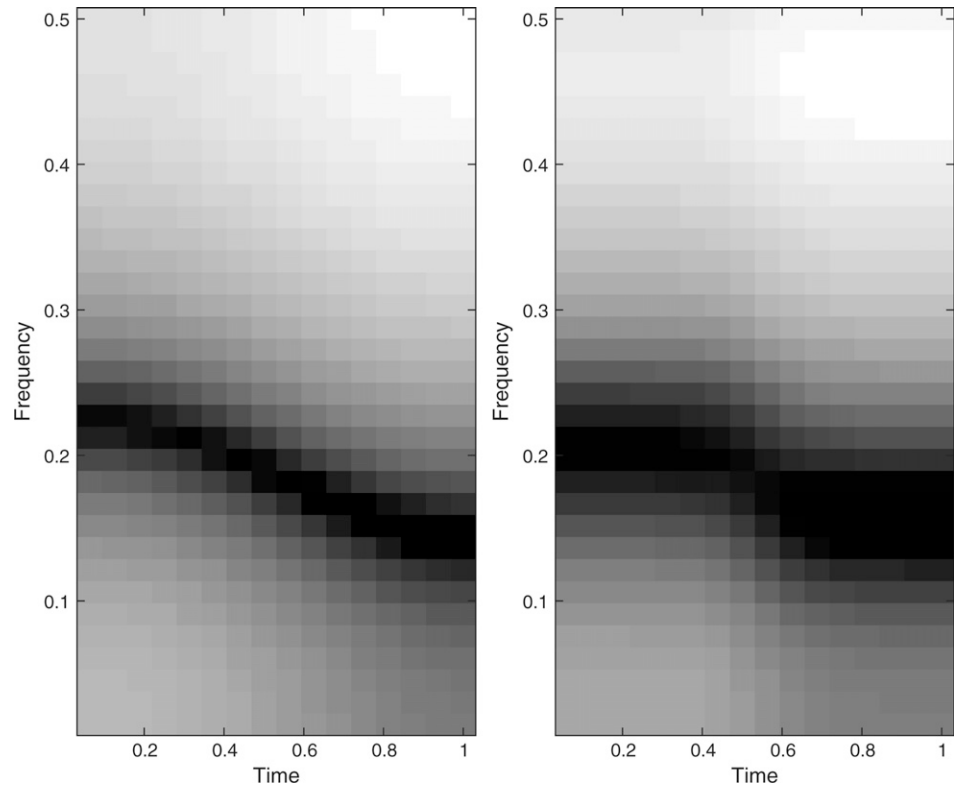

Figure 4. *Left*: True time-varying log-spectrum of process (9). *Right*: Estimated log-spectrum based on averaging the 200 fitted log-spectra.

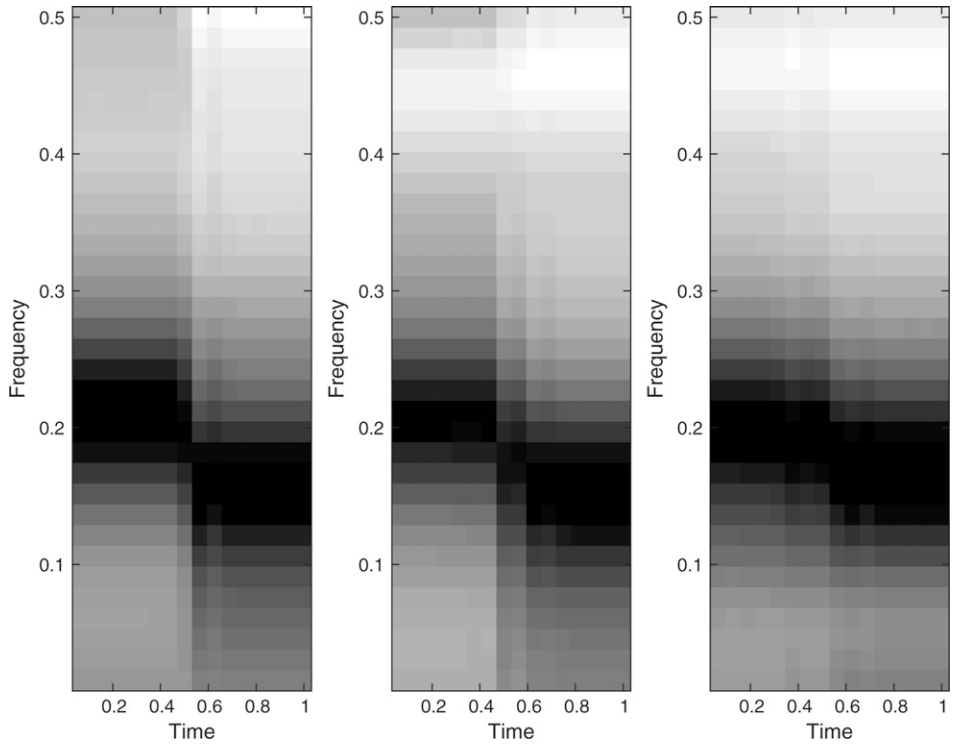

Figure 5. From left to right, individual log-spectrum estimates corresponding to the 10th, 50th, and 90th percentiles of the 200 ASE values for process (9).

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

where $\epsilon_t \overset{\text{iid}}{\sim} N(0, 1)$. In model (11), the lengths of the stationary time series are powers of 2.

Figure 6 shows the true time-varying log spectrum of process (11) on the left, as well as the average of the 50 estimated log spectra, on the right. It is clear from Figure 6 that, on average, our technique provides a good estimate of the true piecewise spectra in this example.

6.3 Time Varying AR(6) Process

In this example, we generate 50 realizations, each of length $N = 1024$, from a time varying autoregressive process of lag 6. In particular, this process can be expressed as $\phi_t(B)x_t = \epsilon_t$, $\varepsilon_t \overset{\text{iid}}{\sim} N(0, 1)$, where $B^m x_t = x_{t-m}$, and

$$\phi_t(B) = 1 - \phi_{t1}B - \ldots - \phi_{t6}B^6. \tag{12}$$

Assume the characteristic polynomial in (12) has three pairs of conjugate complex roots $(a_{tj}^{-1}, a_{tj}^{*-1})$, $j = 1, 2, 3$, where the superscript * denotes the complex conjugate, i.e.,

$$\phi_t(B) = (1 - a_{t1}B)(1 - a_{t1}^*B)(1 - a_{t2}B)(1 - a_{t2}^*B)(1 - a_{t3}B) \times (1 - a_{t3}^*B).$$

To make these roots vary with time, let $a_{tj}^{-1} = A_j \exp(2\pi i\theta_{tj})$, $j = 1, 2, 3$, where $i = \sqrt{-1}$, and let the θ_{tj}s satisfy

$$\begin{aligned} \theta_{t1} &= 0.05 + (0.1/(N-1))t \\ \theta_{t2} &= 0.25 \\ \theta_{t3} &= 0.45 - (0.1/(N-1))t, \end{aligned}$$

for $t = 1, \ldots, N$. The values of A_1, A_2, and A_3 are 1.1, 1.12, and 1.1, respectively. The corresponding coefficients, ϕ_{tj}, $t = 1, \ldots, N$, $j = 1, \ldots, 6$ are obtained as the negative coefficients of the polynomial in (12).

We fitted our method to each of the 50 realizations. Figure 7 presents the true time varying log-spectrum on the left, the average of the 50 estimated log-spectra in the middle, and an estimated log-spectrum based on a single realization. From these plots, it is evident that our method provides good estimates of the true time varying spectrum.

7. APPLICATIONS

In this section we give two examples to illustrate our methodology. The total number of iterations is 300,000 in the first example and 400,000 in the second.

7.1 Southern Oscillation Index

In recent years there has been much research and debate on climate change. One area of research is the El Niño/Southern Oscillation (ENSO) phenomenon. ENSO is an irregular low-frequency oscillation between a warm El Niño state and a cold La Niña state. The SOI is an indicator of the ENSO phenomenon and is calculated to be the standardized anomaly of the mean sea-level pressure difference between Tahiti and Darwin. The strong El Niños of 1982/83 together with the more

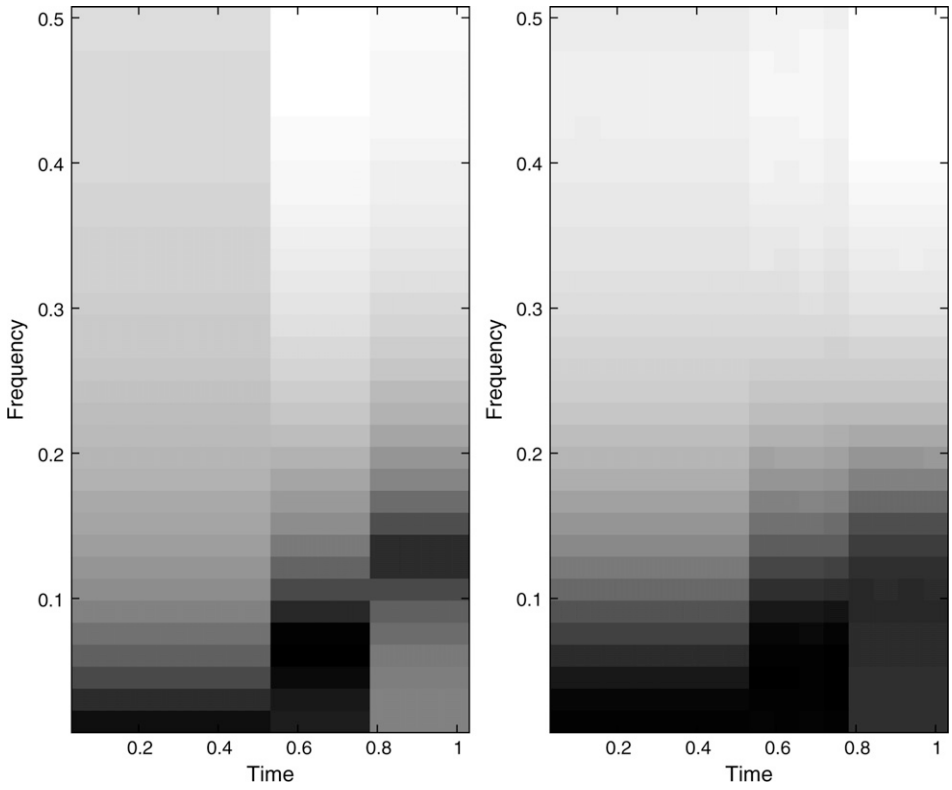

Figure 6. *Left*: True time-varying log-spectrum of process (11). *Right*: Estimated log-spectrum based on averaging the 50 fitted log-spectra.

frequent occurrences of El Niño in recent decades have raised the question of whether human-induced global warming has changed the structure of the ENSO time series. (Timmermann et al. 1999). Trenberth and Hoar (1996) used quarterly standardized seasonal Sea Level Pressure Anomalies at Darwin (DSLPA) as an indicator of ENSO and assumed an ARMA(3,1) model as the data-generating process. They tested explicitly for a change in the time series from 1981 onwards and found that given the values of the index from 1882 to 1981, the likelihood of observing the 1990–1995 ENSO is about 1:3000 years. The SOI is considered to be a better indicator of ENSO than the DSLPA (Chen, 1982) and therefore we use this index as our data.

The data, shown in Figure 8, are monthly values of the SOI from January 1876 to April 2008 and are available at *http://www.bom.gov.au/climate/current/soihtm1.shtml.*

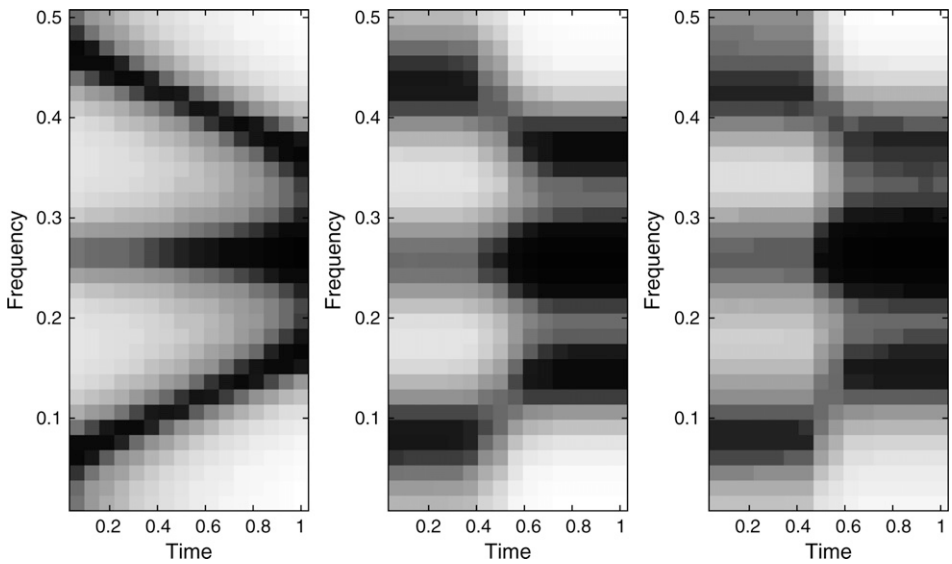

Figure 7. *Left*: True time-varying log-spectrum of the time-varying AR(6) process described in Section 6.3. *Middle*: Estimated log-spectrum based on averaging the 50 fitted log-spectra. *Right*: Estimated log-spectrum based on a single realization.

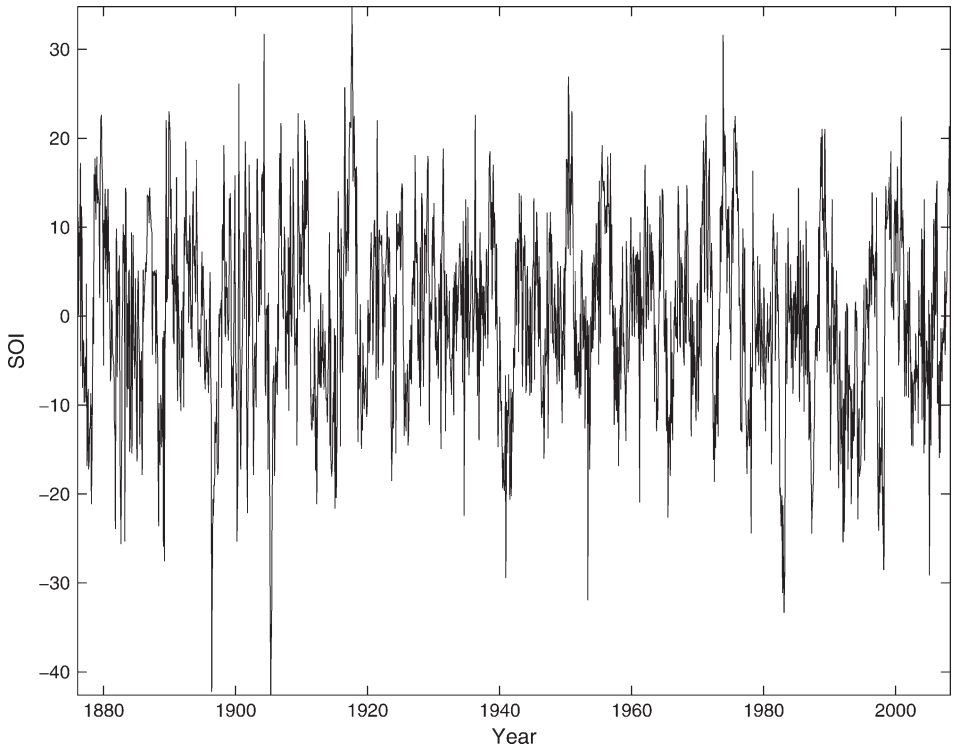

Figure 8. Monthly values of the SOI from January 1876 to April 2008.

We divided the data into 24 segments, each containing 64 observations (leaving out the first 52 observations), and fitted our model. We used a uniform prior for the number of components in the mixture, with an upper limit of 20 components [i.e., $\Pr(j = k) = 0.05$ for $k = 1, \ldots, 20$]. Using this prior, our method estimates $\Pr(j = 1|\boldsymbol{y}) = 0.75$(i.e., the posterior probability of the time series being stationary is 0.75). To study the effect of the prior on the posterior distribution of the number of components, we redid the analysis using a Poisson distribution, with mean λ [i.e., $\Pr(j = k) = \lambda^k e^{-\lambda}/k!$] for three different values of λ. The results of this analysis appear in Table 1. This table shows that a single component has the highest posterior probability for all priors on the number of components. Even when $\lambda = 5$ and hence $\Pr(j = 1) = 0.03$, the posterior probability is $Pr(j = 1|\mathbf{y}) = 0.66$. This finding is in contrast to Trenberth and Hoar (1996) who found that the time series is nonstationary. There are a number of explanations for this disparity.

The first is that we use the SOI rather than the DSLPA as an indicator of ENSO. To test if the cause of the difference in findings was due to using different indicators of ENSO, we repeated the analysis of Trenberth and Hoar using the SOI data. As in Trenberth and Hoar (1996), we grouped the monthly data for the period 1876–1979 into seasons and fitted a parametric model for the time series using Akaike information criterion (AIC) as the basis for model selection. Similar to Trenberth and Hoar, an ARMA(3,1) of the form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \theta_1\, e_{t-1}, e_t \sim N(0, \sigma^2) \quad (13)$$

was the model selected, with parameter estimates $\hat{\phi}_1 = 1.303$, $\hat{\phi}_2 = -0.2707$, $\hat{\phi}_3 = -.1876$, $\theta_1 = -0.7524$, and $\hat{\sigma}^2 = 49$. In (13) y_t is the average of the SOI for the three months in season t. This model was then used to generate 1,000,000 years of data. Trenberth and Hoar focused on the period December 1989–May 1995, where 22 successive positive anomalies with a median of 0.94 mb were recorded, and asked the question "If the ARMA(3,1) model is the true model, how likely is it that we observe 22 successive negative DSLPA, and how likely is it that we observe a median value of 0.94 mb for any 22 season period?" The answers to these questions were about 2 in 8,000 years and 1 in 3,000 years, respectively. From this they concluded that the time series structure of the DSLPA had changed. Repeating the same analysis for the SOI data, we obtained the distribution of run lengths of either sign, see Figure 9. Our results are of a similar order of magnitude to those computed by Trenberth and Hoar, but comparison is difficult because the maximum run length for the SOI data were only 11, not 22. However, except for one slightly positive SOI (0.20 for December, January, and February in 92/93), the maximum run length would have been 17. The frequencies of observing run lengths of 11, 17, 22, or greater appear in

Table 1. Posterior probabilities of the number of components in the model fitted to the SOI data as a function of the prior distribution

	Prior			
	Uniform	Poisson		
Number of components	$P(j = k) = 0.05$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$
1	0.755	0.94	0.75	0.66
2	0.241	0.06	0.24	0.33
3	0.004	0.00	0.01	0.01
4	0.000	0.00	0.00	0.00

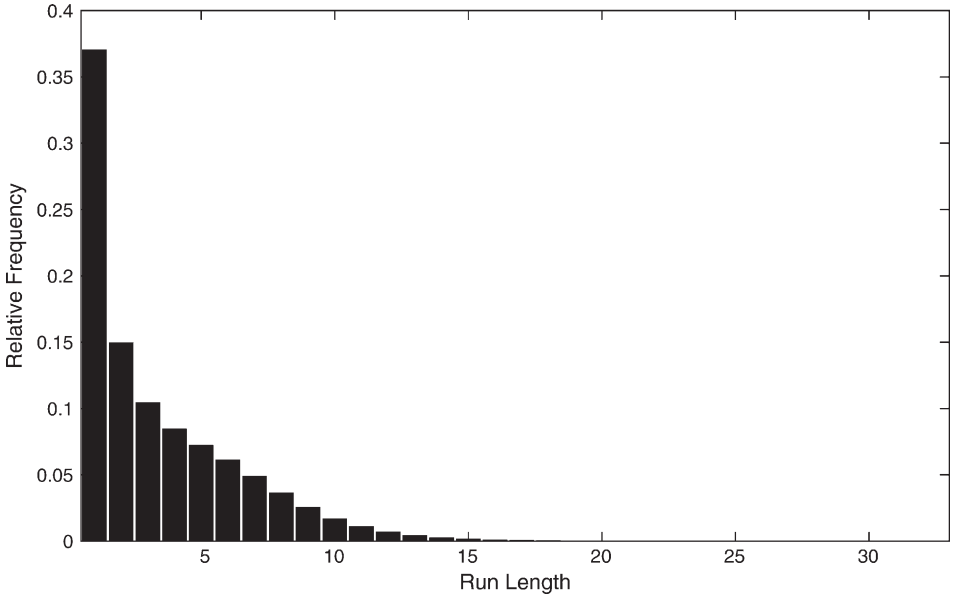

Figure 9. Distribution of the number of runs for simulated values of the SOI assuming an ARMA(3,1) process is the correct model.

Table 2. This table shows that the estimated frequencies of events are very similar for both the SOI and the DSLPA, which is not surprising given that the ARMA(3,1) estimated for the SOI is very similar to the ARMA(3,1) process estimated for the DSLPA (Trenberth and Hoar, 1996). However, different conclusions would be reached because the behavior of the SOI for the period in question (i.e., December 1989–May 1995), is different from the behavior of the DSLPA in the same period. A similar picture emerges for the test of the 22 season median value. The median value of the SOI from December 1989 to May 1995 was -7.5. Using the data-generating model given in (13), we computed that a 22 season median of -7.5 could be expected to happen about once in 130 years, a very different value from once in 3,000 years.

There are a number of other differences in our analysis of the ENSO effect. Trenberth and Hoar (1996) assumed an ARMA(3,1) model for the data-generating process, whereas the method presented in this article models changes in the frequency domain to avoid assuming a particular parametric model.

Another explanation is that Trenberth and Hoar (1996) tested explicitly if there had been a change from 1981 onward. In our model, explicitly testing for a difference in the SOI from 1981 onward is equivalent to assuming a priori that in a mixture of two components, $\pi_{12s} = 1$ for segments s corresponding to time periods before 1981 and $\pi_{12s} = 0$, otherwise. In effect, the uncertainty surrounding the number and location of change points has been ignored by Trenberth and Hoar (1996). In contrast, our methodology makes no assumptions regarding the location and number of any possible nonstationarity, and our estimate of $\Pr(j = 1|\boldsymbol{y})$ integrates out the uncertainty regarding both the mixing function $\boldsymbol{\pi}$ and the number of possible change points.

A plot of the estimate of the log power together with 95% posterior intervals appears in Figure 10. Figure 10 shows a peak at about $\nu = 0.02$, which corresponds to a 48 month (4 year) cycle. We note that in our analysis, by modeling the evolution of the log spectra, we are modeling the evolution of the second moment. It may be that the SOI is a nonstationary process but the nonstationarity is apparent only in higher-order moments.

7.2 Earthquakes and Explosions

In this section we illustrate our methodology with seismic traces of an earthquake and a mining explosion taken from Shumway and Stoffer (2006). These data were measured at a recording station in Scandinavia, and are each of length 2048. Figures 11 and 12 show the data along with their estimated log spectrum. These particular time series both consist of two waves, the compression wave, also known as primary or P wave, which is the start of the series and the shear, or S wave, which arrives at the midpoint of the series. The delay between the arrival of the P wave and the arrival of the S wave is used to get a quick and reasonably accurate estimate of the location of the event. The analysis of such seismic data is one of critical importance for monitoring a comprehensive test-ban treaty. As argued by many authors [e.g., Shumway and Stoffer (2006, Chap. 7)], distinguishing between the seismic traces of earthquakes and explosions is best accomplished in the frequency domain. Hence, two problems are of interest in this analysis. The first one is to identify the arrival of the S wave, and the second problem is to estimate the time-varying spectrum of the

Table 2. Frequency of occurrence of length of a run in SOI

	Data	
Run length	SOI	DSLPA
>11	1 in 47 years	1 in 55 years
>17	1 in 800 years	1 in 910 years
>22	1 in 9,200 years	1 in 8,600 years

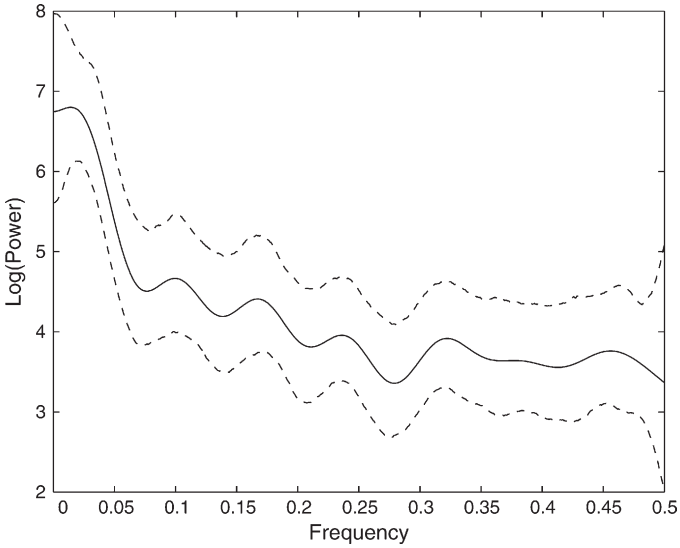

Figure 10. Estimated log spectrum (*solid line*) and 95% posterior intervals (*dashed line*) against frequency for monthly values of the SOI from June 1878 to May 2006.

process to distinguish between the seismic traces of explosions and earthquakes.

To illustrate our method, we fitted the mixture model to both time series after dividing them into 32 segments of 64 observations each and computing the corresponding log periodograms. The Gibbs sampler was run for 400,000 iterations with a burn-in period of 200,000 iterations. Table 3 presents the posterior probabilities of the number of components for the two time series. The S component for the earthquake shows power at the low frequencies only, and the power remains strong for a long time. In contrast, the explosion shows power at higher frequencies than the earthquake, and the power of the signals (P and S waves) does not last as long as in the case of the earthquake. Although the estimated time-varying spectra are qualitatively similar to those considered in Shumway and Stoffer (2006, chap. 4 and 7), the additional information contained in the number of components needed to estimate the time-varying spectrum of each seismic trace, as expressed in Table 3, may add a new dimension with which to discriminate between earthquakes and explosions.

APPENDIX: DETAILS OF THE SAMPLING SCHEME

First initialize the number of components, j. Conditional on this value, initialize the indicators, γ_{srj}, for $s = 1, \ldots, S$ and $r = 1, \ldots, j$ and $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j^{*\prime}, \boldsymbol{\tau}'_j, \boldsymbol{\delta}_j')'$, where $\boldsymbol{\beta}_j^* = (\alpha_{01}, \boldsymbol{\beta}'_{1j}, \ldots, \alpha_{0j}, \boldsymbol{\beta}'_{jj})'$, $\boldsymbol{\tau}_j = (\tau_{1j}^2, \ldots, \tau_{jj}^2)'$, and $\boldsymbol{\delta}_j = (\boldsymbol{\delta}'_{2j}, \ldots, \boldsymbol{\delta}'_{jj})'$. Note that $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_{11}^{*\prime}, \tau_{11}^2)'$.

1. **Moving Between Models**.

Let j^c and j^p be the current and proposed number of components in the model, respectively. Let $\boldsymbol{\theta}^c$ and $\boldsymbol{\theta}^p$ be the corresponding parameter values (i.e., $\boldsymbol{\theta}^c = \boldsymbol{\theta}_{j^c}$ and $\boldsymbol{\theta}^p = \boldsymbol{\theta}_{j^p}$). Let $V^c = (j^c, \boldsymbol{\theta}^c)$ and $V^p = (j^p, \boldsymbol{\theta}^p)$.

Using a proposal density $q(V^c, V^p)$, we draw V^p from $P(V^p|\tilde{\mathbf{y}})$ as follows.

(1) If $1 < j^c < J$, propose a value for j^p with transition probability $q(j^c \to j^p = j^c \pm 1) = 0.5$. If $j^c = 1$, then $q(j^c \to j^p = 2) = 1$, and if $j^c = J$, then $q(j^c \to j^p = j - 1) = 1$.

(2) If $j^p = j^c + 1$, then $\boldsymbol{\theta}^p = (\boldsymbol{\beta}_{j^c}^{*\prime}, \boldsymbol{\beta}_{j^p j^p}^{*\prime}, \boldsymbol{\tau}'_{j^c}, \tau_{j^p j^p}^2, \boldsymbol{\delta}'_{j^c}, \boldsymbol{\delta}'_{j^p j^p})'$, where $\boldsymbol{\delta}_{j^p j^p}$, $\tau_{j^p j^p}^2$, and $\boldsymbol{\beta}_{j^p j^p}^*$ are generated as follows.

(a) Draw $\boldsymbol{\delta}_{j^p j^p}$ from $N(0, \sigma_\delta^2 I_2)$.

(b) Draw $\tau_{j^p j^p}^2$ from $U(0, \tau_{j^c j^c}^2)$.

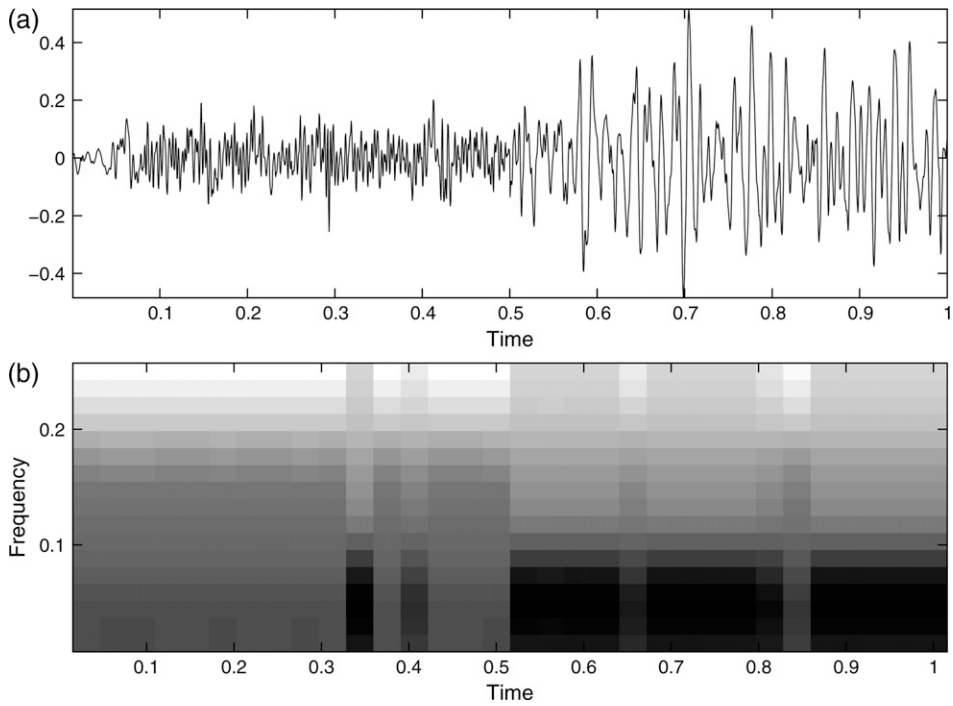

Figure 11. (*a*) The earthquake time series and (*b*) its estimated time-varying log spectrum.

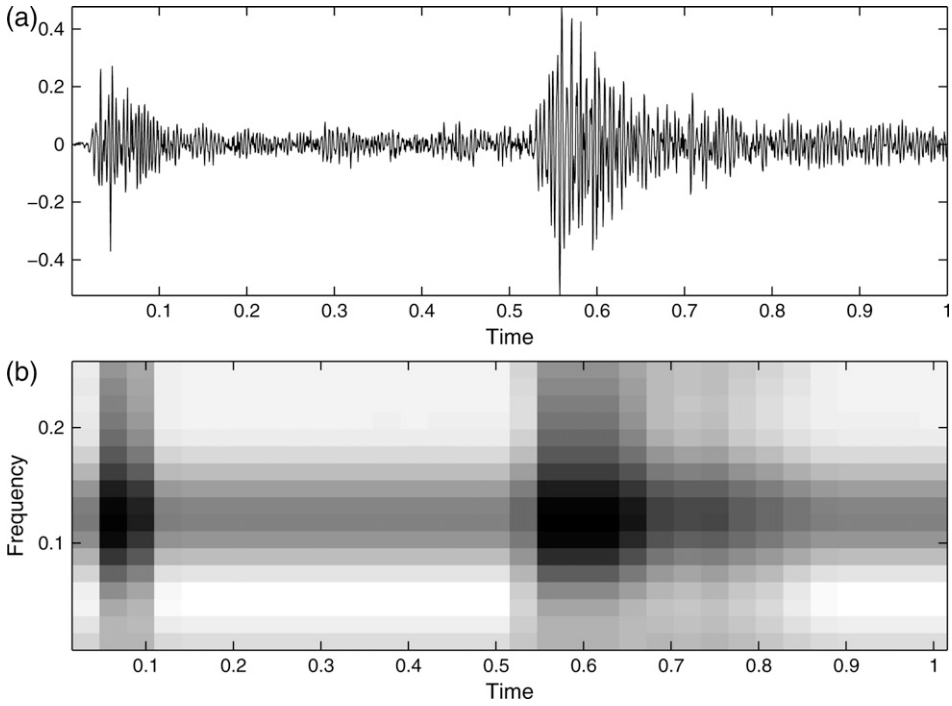

Figure 12. (*a*) The explosion time series and (*b*) its estimated time-varying log spectrum.

(c) Draw $\alpha_{j^p j^p}$ from $N(0, \sigma_\alpha^2)$.
(d) Draw $\boldsymbol{\beta}_{j^p j^p}$ from $N(0, \tau^2_{j^p j^p} \cdot I_k)$, where k is the number of basis functions.
(e) Accept the proposed parameter values with probability

$$\epsilon = \min\left\{1, \frac{P(V^p|\tilde{\mathbf{y}})q(V^p, V^c)}{P(V^c|\tilde{\mathbf{y}})q(V^c, V^p)}\right\} \tag{A.1}$$

(3) If $j^p = j^c - 1$, then

(a) $\boldsymbol{\theta}^p = \boldsymbol{\theta}_{j^c-1}$
(b) Accept the proposed parameter values with probability (A.1).

2. **Updating Within a Model**

Given the new values of j and $\boldsymbol{\theta}_j$, the model parameters are updated as follows.

(1) If $j > 1$, then

(a) Let $\gamma_{sj} = r$ if $\gamma_{srj} = 1$. Generate component indicators γ_{sj}, $s = 1, \ldots, S$, from

$$p(\gamma_{sj} = r|\boldsymbol{\beta}_j^*, \boldsymbol{\delta}_j, X^*, \mathbf{y}_s) = \pi_{rjs} p(\mathbf{y}_s|X^*, \boldsymbol{\beta}_{rj}^*) / \sum_{h=1}^{j} \pi_{hjs} p(\mathbf{y}_s|X^*, \boldsymbol{\beta}_{hj}^*)\,, \tag{A.2}$$

where

$$\begin{aligned} P(\mathbf{y}_s|X^*, \boldsymbol{\beta}^*) \;\propto\; & \exp\left(\tfrac{y_{s,0}-\mu_0}{2} - \tfrac{1}{2}\exp(y_{s,0}-\mu_0)\right) \\ & \times \exp\left(\tfrac{y_{s,n/2}-\mu_{n/2}}{2} - \tfrac{1}{2}\exp(y_{s,n/2}-\mu_{n/2})\right) \\ & \times \textstyle\prod_{k=1}^{n/2-1} \exp\left(y_{s,k} - \mu_k - \exp(y_{s,k}-\mu_k)\right) \end{aligned} \tag{A.3}$$

and μ_k is the kth element of $X^*\boldsymbol{\beta}^*$.

(b) Generate $\boldsymbol{\delta}_j$ from $p(\delta_j \mid \gamma_j, U)$ via a Metropolis-Hastings step. The conditional posterior distribution of $\boldsymbol{\delta}_j$ is

$$\begin{aligned} p(\boldsymbol{\delta}_j|\boldsymbol{\gamma}_j, U) \;&\propto\; p(\boldsymbol{\delta}_j) \prod_{s=1}^{S}\prod_{r=1}^{j} \pi_{rjs}^{\gamma_{srj}} \\ &=\; p(\boldsymbol{\delta}_j) \prod_{s=1}^{S}\prod_{r=1}^{j} \left(\frac{\exp(\boldsymbol{\delta}_{rj}'\mathbf{u}_s)}{\sum_{h=1}^{j}\exp(\boldsymbol{\delta}_{hj}'\mathbf{u}_s)}\right)^{\gamma_{srj}}. \end{aligned} \tag{A.4}$$

The proposal density is $N(\boldsymbol{\delta}_j^{\max}, \Sigma_\delta)$, where $\boldsymbol{\delta}_j^{\max}$ is the value of $\boldsymbol{\delta}_j$ maximizing $p(\boldsymbol{\delta}_j|\boldsymbol{\gamma}_j, U)$ in (A. 4), and $\Sigma_\delta = \left\{-\frac{\partial^2 \log p(\boldsymbol{\delta}_j|\boldsymbol{\gamma}_j, U)}{\partial \boldsymbol{\delta}_j \partial \boldsymbol{\delta}_j'} \,|\, _{\boldsymbol{\delta}_j = \boldsymbol{\delta}_j^{\max}}\right\}^{-1}$.

(2) For any j:

(a) Generate $\boldsymbol{\beta}_j^*$ from its conditional posterior distribution

$$\begin{aligned} p(\boldsymbol{\beta}_j^*|\boldsymbol{\tau}_j, \boldsymbol{\gamma}_j, X^*, \tilde{\mathbf{y}}) \;&=\; p(\boldsymbol{\beta}_j^*) \textstyle\prod_{s=1}^{S}\prod_{r=1}^{j} p(\mathbf{y}_s|X^*, \boldsymbol{\beta}_{rj}^*)^{\gamma_{srj}} \\ &=\; \textstyle\prod_{s=1}^{S}\prod_{r=1}^{j} p(\boldsymbol{\beta}_{rj}^*) p(\mathbf{y}_s|X^*, \boldsymbol{\beta}_{rj}^*)^{\gamma_{srj}} \end{aligned} \tag{A.5}$$

as follows. For $r = 1, \ldots, j$:
Generate the vector $\boldsymbol{\beta}_{rj}^*$ via a Metropolis-Hastings step. The conditional posterior distribution of $\boldsymbol{\beta}_{rj}^*$ is

Table 3. Posterior probabilities of the number of components

No. of components	Earthquake	Explosion
1	0.000	0.000
2	0.960	0.098
3	0.040	0.700
4	0.000	0.193
>5	0.000	0.009

$$p(\boldsymbol{\beta}^*_{rj}|\tau^2_{rj}, \boldsymbol{\gamma}_j, X^*, \tilde{\mathbf{y}}) = p(\boldsymbol{\beta}^*_{rj}) \prod_{s=1}^{S} p(\mathbf{y}_s|X^*, \boldsymbol{\beta}^*_{rj})^{\gamma_{srj}} .$$

The proposal density is $N(\boldsymbol{\beta}^{*\max}_{rj}, \Sigma_{\beta^*_{rj}})$, where $\boldsymbol{\beta}^{*\max}_{rj} = \arg\max_{\boldsymbol{\beta}^*_{rj}} p(\boldsymbol{\beta}^*_{rj}|\tau^2_{rj}, \boldsymbol{\gamma}_j, X^*, \tilde{\mathbf{y}})$ and $\Sigma_{\beta^*_{rj}} = \{-(\partial^2 \log p(\boldsymbol{\beta}^*_{rj}|\tau^2_{rj}, \boldsymbol{\gamma}_j, X^*, \tilde{y}))/\partial\boldsymbol{\beta}^*_{rj}\partial\boldsymbol{\beta}^*_{rj}|_{\boldsymbol{\beta}^*_{rj}=\boldsymbol{\beta}^{*\max}_{rj}}\}^{-1}$.

(b) Generate τ_j from $p(\boldsymbol{\tau}_j|\boldsymbol{\beta}_j, \tilde{\mathbf{y}}) = p(\boldsymbol{\tau}_j|\boldsymbol{\beta}_j)$, such that $\tau^2_{1j} \geq \tau^2_{2j} \geq \cdots \geq \tau^2_{jj}$. Note that $p(\tau^2_{rj}|\boldsymbol{\beta}_{rj}) = \text{IG}((k/2), (1/2)\boldsymbol{\beta}'_{rj}\boldsymbol{\beta}_{rj})$, where k is the number of basis functions.

[Received July 2007. Revised July 2008.]

REFERENCES

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

Carter, C. K. and Kohn, R. (1997), "Semiparametric Bayesian Inference for Time Series with Mixed Spectra," *Journal of the Royal Statistical Society, B,* 59, 255–268.

Chen, W. Y. (1982), "Assessment Of Southern Oscillation Sea-Level Pressure Indices," *Monthly Weather Review,* 110, 800–807.

Chiann, C., and Morettin, P. A. (1999), "Estimation of Time-varying Linear Systems," *Statistical Inference for Stochastic Processes,* 2, 253–285.

Coifman, R., and Wickerhauser, M. (1992), "Entropy Based Algorithms for Best Basis Selection," *IEEE Transactions on Information Theory,* 38, 712–718.

Dahlhaus, R. (1997), "Fitting Time Series Models to Nonstationary Processes," *Annals of Statistics,* 25, 1–37.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006), "Structural Breaks Estimation for Nonstationary Time Series Models," *Journal of the American Statistical Association,* 101, 223–239.

Green, P. J. (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.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991), "Adaptive Mixtures of Local Experts," *Neural Computation,* 3, 79–87.

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.

Lavielle, M. (1998), "Optimal Segmentation of Random Processes," *IEEE Transactions on Signal Processing,* 46, 1365–1373.

Ombao, H. C., Raz, J. A., Von Sachs, R., and Malow, B. A. (2001), "Automatic Statistical Analysis of Bivariate Nonstationary Time Series," *Journal of the American Statistical Association,* 96, 543–560.

Priestley, M. (1965), "Evolutionary Spectra and Nonstationary Processes," *Journal of the Royal Statistical Society:* Series B, 27, 204–237.

Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. J. (2002), "Bayesian Curve Fitting using MCMC with Applications to Signal Segmentation," *IEEE Transactions on Signal Processing,* 50, 747–758.

Shumway, R. H., and Stoffer, D. S. (2006), *Time Series Analysis and Its Applications: With R Examples* (2nd ed.), New York: Springer.

Thomson, D. J. (1990), "Time Series Analysis of Holocene Climate Data," *Philosophical Transactions of the Royal Society, A,* 330, 601–616.

Timmermann, A., Oberhuber, J., Bacher, A., Esch, M., Latif, M., and Roeckner, E. (1999), "Increased El Niño Frequency in a Climate Model Forced by Future Greenhouse Warming," *Nature,* 398, 694–697.

Trenberth, K. E., and Hoar, T. J. (1996), "The 1990–1995 El Niño-Southern Oscillation Event: Longest on Record," *Geophysical Research Letters,* 23, 57–60.

Wahba, G. (1980), "Automatic Smoothing of the Log Periodogram," *Journal of the American Statistical Association,* 75, 122–132.

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

Whittle, P (1957), "Curve and Periodogram Smoothing," *Journal of the Royal Statistical Society:* Series B, 19, 38–47.

Wood, S., Jiang, W., and Tanner, M. A. (2002), "Bayesian Mixture of Splines for Spatially Adaptive Nonparametric Regression," *Biometrika,* 89, 513–528.