AUTOMATIC STATISTICAL ANALYSIS OF BIVARIATE NON-STATIONARY TIME SERIES

May 11, 1999

Hernando Ombao[1], **Jonathan Raz**[1], **Rainer von Sachs**[2], **Beth Malow**[3]

Department of Biostatistics[1], University of Michigan

Institut de Statistique[2], Université Catholique de Louvain

Department of Neurology[3], University of Michigan

Abstract. We propose a new method, the "Auto-SLEX" method, for analyzing bivariate non-stationary processes. The Auto-SLEX method is a procedure that automatically segments the time series into approximately stationary blocks and automatically estimates the time-varying spectra and coherence. It is based on the SLEX (**S**mooth **L**ocalized Complex **EX**ponential) transform which forms a library of orthogonal transforms. The SLEX transform is built on the SLEX basis functions which are complex-valued and are simultaneously localized in time and frequency. We show that the SLEX periodograms are consistent, report simulation results and apply the method to an electroencephalogram (EEG) data set recorded during an epileptic seizure.

Financial support from the contract "Projet d'Actions de Recherche Concertes" no. 98/03-217 of the Belgian Government is gratefully acknowledged.

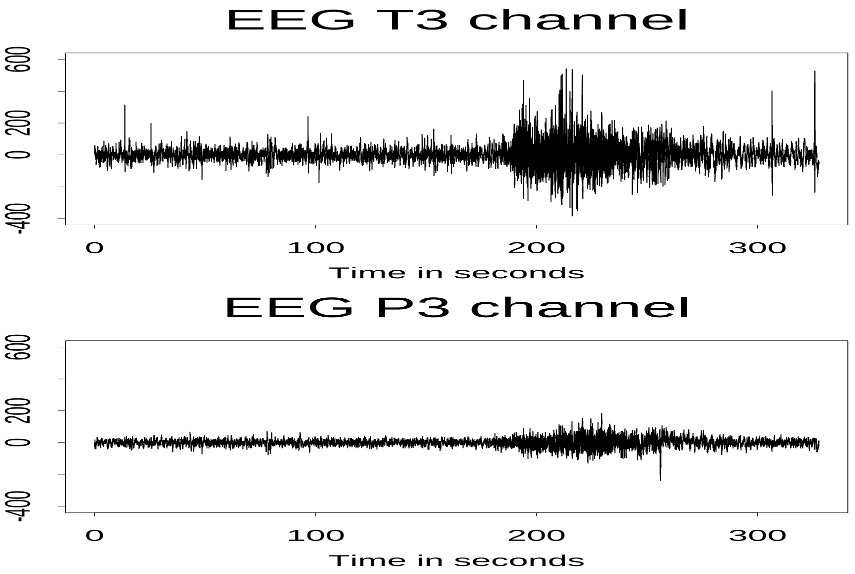

Figure 1: Bivariate EEG time series at channels $T3$ and $P3$. Number of time points $T = 32768$. Sampling rate = 100 Hertz. Subject: Female patient diagnosed with left temporal lobe epilepsy. Data set provided by the Department of Neurology, University of Michigan.

1 Introduction

Electroencephalograms (EEGs) or brain waves are brain electrical potentials recorded by electrodes placed on the scalp of subjects. Examples of EEGs recorded at the left temporal lobe and the left parietal lobe during an epileptic seizure are given in Figure 1. Epileptic seizure EEGs can be modeled as non-stationary processes. As a patient progresses from pre-ictal to post-ictal state, changes in the electrical activity of the brain are manifested in changes in the estimated spectral power. In addition to estimating the power spectra, neurologists are interested in the estimate of coherence between EEGs recorded at different channels. An estimate of coherence is important because it gives an indication on how a seizure spreads from a local area to neighboring regions.

Most methods of analyzing EEGs are built on the premise that EEGs can be modeled approximately by piecewise stationary processes. Hence, the time-varying spectrum can be approximated by the piecewise constant spectra of the stationary processes. The standard approach of estimating the time-varying spectrum is by the windowed Fourier transform (Cohen, 1989; Williams, 1997). This method first divides the time series into blocks of equal sizes and then computes the windowed Fourier transform at each of the blocks. A drawback to this methodology is that the length of the blocks is chosen subjectively. Furthermore, in many applications, it may not be reasonable to require that all blocks or windows should be of equal lengths.

Adak (1998) proposed a methodology that allows a data-dependent choice of the window lengths. The lengths are adapted to the rate of change of the spectrum, i.e., bigger windows are generated for longer periods of stationarity. Adak's methodology divides the time series in a dyadic manner, up to some pre-specified level J, into possibly overlapping blocks and then computes the windowed Fourier transform at each of the blocks. At a particular level $j = 0, 1, \cdots, J$, the time series is divided into 2^j blocks. Adjacent blocks at every level are combined together to form one stationary block if their estimated spectra are close according to some distance measure. Otherwise, they are left as distinct blocks.

Donoho, Mallat and von Sachs [DMvS] (1998) introduced a model of a locally stationary process and then developed a procedure for obtaining an adaptively-smoothed and consistent estimator of the time-varying covariance under this model. Intuitively, a process is said to be locally stationary if one can form an interval around each time point so that the process is approximately stationary at that interval. Dahlhaus (1997) gives a slightly different definition of this local stationarity that we will use in our approach as a model to validate the properties of our estimation procedure from the asymptotic point of view of consistency.

The DMvS method uses the cosine packet transform (CPT) that forms a library (collection) of orthonormal transforms. Cosine packets are like windowed cosine vectors. They are localized in time and frequency. The coefficients are first computed by taking the inner product between the time series and the cosine packets. These coefficients are arranged in a cosine packet table and then squared to form the cosine packet periodograms. The DMvS method automatically segments the time series by applying the Best Basis Algorithm (BBA) of Coifman and Wickerhauser (1991) on the raw periodograms. At each of the blocks chosen by BBA, wavelet-denoised estimators for the log-spectrum are obtained and then exponentiated to form the spectral estimate. The inverse CPT is then applied on the spectral estimate to obtain the estimate of the time-varying covariance.

The primary limitation of the Adak method is that it is based on a non-orthogonal transform. As a consequence, the theory of consistency of the method is not tractable (Adak, *personal communication, 1997*). The Adak method, similar to the DMvS method and our proposed method, employs a dyadic segmentation of the time series data set. Adak (1998), however, failed to demonstrate that the effect of the dyadic segmentation is asymptotically negligible. In addition, the method is for univariate processes only and hence cannot give estimates of the time lag and coherence between between components of a multivariate time series.

DMvS (1998) provide a consistent estimator of the time-varying covariance under the DMvS model of a locally stationary process. A frequency domain approach giving an estimate of the time-dependent spectrum, however, is more directly applicable to the study of epileptic seizure EEGs. The primary limitation of the DMvS method is that it uses the CPT, which is a real-valued transform. Hence, it does not have a phase and is not able to model lag between pairs of components in a multivariate time series. Moreover, it is developed only for univariate processes and hence cannot give estimates of coherence.

In this paper, we propose a methodology, which we call the Auto-SLEX method, that overcomes the limitations of the existing methods. The Auto-SLEX method is a statistical procedure that (i.) automatically segments the bivariate time series into approximately stationary blocks and (ii.) automatically chooses the smoothing parameter for the estimation of time-varying spectra and coherence. The Auto-SLEX method is based on the SLEX (**S**mooth **L**ocalized Complex **EX**ponential) transform. The SLEX basis vectors are related to the Fourier basis vectors. They are obtained by applying two special windows to the Fourier vectors. They are simultaneously localized in time and frequency and hence are able to represent non-stationary time series. The SLEX transform also forms a library of orthonormal transforms. Orthonormal transforms are mathematically elegant and they preserve the energy of the time series. Furthermore, they allow for the theory to be tractable (DMvS, 1998). In addition, the SLEX transform is complex-valued. It has a phase which can model the lag between the components of the bivariate time series. The Auto-SLEX method can give estimates of coherence because it is designed for multivariate non-stationary time series. Although the Auto-SLEX procedure presented in this paper is for bivariate processes only, its extension to the multivariate setting is natural and straightforward. Finally, the Auto-SLEX method yields results that are easy to interpret because they parallel existing spectral methods for analyzing stationary time series.

The rest of the article is organized as follows. In Section 2, we state the Dahlhaus model for locally stationary processes and then discuss apparent intuitive approximations by piecewise stationary and blended stationary processes. The SLEX transform is described in Section 3 and the Auto-SLEX method is discussed in Section 4. In Section 5, theoretical results are presented. In Sections 6 and 7 we show simulation results and apply the method to a bivariate EEG data set before we finally conclude. Proofs are deferred to an appendix.

2 Locally Stationary Processes

In this section, we give a formal definition of the model of locally stationary processes given in Dahlhaus (1997). It is a based on a time-dependent analog to the classical Cramér spectral representation of stationary processes. Our discussion is restricted to univariate zero-mean processes for ease of exposition. The multivariate extension (see also Dahlhaus, 1999) is straightforward and parallels the extension of the classical spectral representation to the multivariate case (Brillinger, 1981, Section 7).

2.1 The Dahlhaus Model

Definition 2.1 (Dahlhaus) *A sequence of zero-mean random variables* $\{X_{t,T}\}_{t=0}^{T-1}$, $T \geq 1$, *is called* **locally stationary** *with transfer function* A^0 *if there exists a representation*

$$X_{t,T} = \int_{-1/2}^{1/2} A^0_{t,T}(\omega)\, e^{i2\pi\omega t}\, dZ(\omega) \ , \quad t = 0, \ldots, T-1, \tag{1}$$

where

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

(ii) *There exists a positive constant* D *and a smooth function* $A : [0,1] \times [-1/2, 1/2] \longrightarrow C$ *with* $A(u,\omega) = \overline{A(u,-\omega)}$ *such that for all* T,

$$sup_{t,\omega} |A^0_{t,T}(\omega) - A(t/T, \omega)| \leq D\ T^{-1}; \tag{2}$$

(iii) *For all* ω, $A(u,\omega)$ *is continuous in* u.

Definition 2.2 (Dahlhaus) *The* **time-varying** *or* **evolutionary spectrum** *at time* $u \in [0,1]$ *and frequency* $\omega \in [-1/2, 1/2]$ *of the Dahlhaus locally stationary process is:* $f_{XX}(u,\omega) = |A(u,\omega)|^2$.

Remarks:
(1.) Note that the first argument of $A(t/T,\omega)$ is rescaled to live on the unit interval. Increasing the number of observations, T, does not mean looking into the future. Rather, this asymptotic framework allows for more data to be observed at a local structure and to do asymptotic inference starting from a single realization of $\{X_{t,T}\}$. This is possible by equation (2) which says that $A^0_{t,T}(\omega) \approx A(t/T,\omega)$, i.e, the smoothness of A in u controls the change of $A^0_{t,T}$ as a sequence in t such that it is allowed to change only slowly over time. The idea behind this is essentially that, for each fixed T, one implicitly assumes some local interval of stationarity about each time point and a smooth change from one interval to the next.
(2.) In the literature (Dahlhaus, 1996, 1997, Neumann and von Sachs, 1997) there exists a variety of possible smoothness conditions on $A(u,\omega)$. We will state ours in the various context of the theorems we intend to show. Compare also the slightly different model approach of Adak (1998) which includes a local smoothness condition directly in her formulation of equation (2).
(3.) Having defined the time-varying spectrum $f(u,\omega)$ of the model allows us to develop asymptotic estimation theory for this object as a function in the (re-scaled) time–frequency plane $[0,1] \times [-1/2, 1/2]$. This parallels non-parametric estimation of curves where questions of consistency, asymptotic normality and related measures (such as L_2-risks) of the quality of an estimator of $f(u,\omega)$, e.g. a smoothed time-varying periodogram, are addressed.

The next two subsections discuss two classes of processes which can serve as intuitive simple approximations to locally stationary processes, the class of **piecewise stationary processes** and the class of **blended stationary processes**. Both, but in particular the second one, are also useful to motivate the construction of estimators of the time-varying spectrum of a locally stationary process.

2.2 Piecewise Stationary Processes

A piecewise stationary processes can be defined (see, e.g., Adak, 1998) to be

$$\tilde{X}_{t,T} = \sum_{b=1}^{B} X_t^b \ \chi(u, I_b), \tag{3}$$

where X_t^b are stationary processes with spectra $f^b(\lambda)$; $I_b = [u_{b-1}, u_b)$; $\chi(u, I_b) = 1$ if $u \in I_b$ and $\chi(u, I_b) = 0$ if $u \notin I_b$. Time is rescaled to live on the interval $[0,1]$ so that $0 = u_0 \leq \cdots \leq u_B = 1$ forms a partition of $[0,1]$. It is not difficult to show that locally stationary processes can be well approximated by piecewise stationary processes.

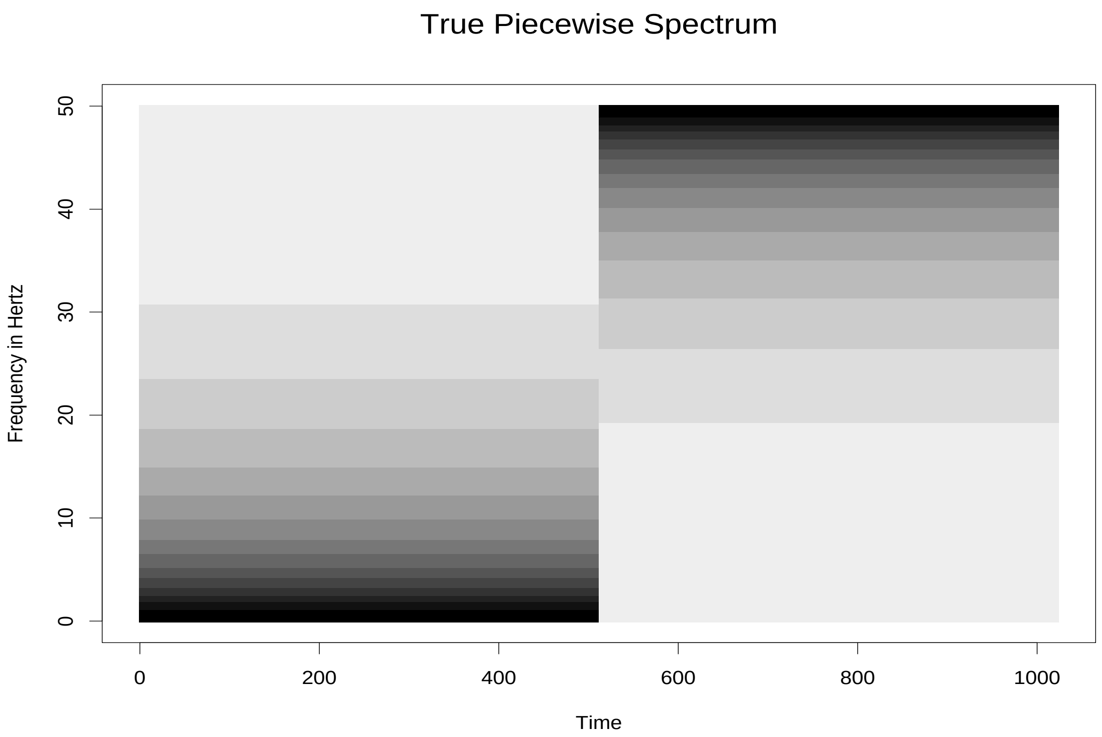

Figure 2: Time-frequency plot or the time-varying spectrum of a piecewise stationary process: At time $0 \le t \le 511$, the process is AR(1)(0.9). At time $512 \le t \le 1023$, the process is AR(1)(-0.9)

In fact this is just a variant of (the proof of) our more general Theorem 2.1 below on the approximation of locally stationary processes by blended stationary processes. Hence, the time-varying spectrum $f(u, \omega)$ of a locally stationary process can be approximated by a piecewise constant spectrum of each stationary process :

$$f(u, \omega) \approx \sum_{b=1}^{B} f^b(\omega) \chi(u_{b-1} \le u < u_b) \tag{4}$$

In general, the number of blocks, B, needs to increase with the length, T, to make the approximation error tend to zero (see, again, Theorem 2.1 below). In order to assure that more observations fall within each stationary segment, we need to satisfy: $\frac{B}{T} \to 0$ as $B \to \infty$ and $T \to \infty$. This equivalent to saying that the upper bound for the segment lengths M fulfills: $\frac{M}{T} \to 0$ as $T \to \infty$.

Time-frequency plot

A useful tool in analyzing non-stationary time series is the **time-frequency plot**. A time-frequency plot is a representation of the time-varying spectrum. It gives the concentration of power at each frequency at each time point. An example of a time-frequency plot of a process defined below is given in Figure 2. Define a random process Y_t to be an autoregressive process of order p, denoted AR(p), with parameters $\theta_1, \ldots, \theta_p$ if it is generated by the equation

$$Y_t \quad = \quad \theta_1 Y_{t-1} + \ldots + \theta_p Y_{t-p} + \xi_t \tag{5}$$

where ξ_t are independent $N(0,1)$ random variables. The time-frequency plot in Figure 2 is a representation of the time-varying spectrum of a piecewise stationary process defined as

$$X_t = \begin{cases} X_{1,t} & \text{if } 0 \leq t < 511 \\ X_{2,t} & \text{if } 512 \leq t < 1023 \end{cases}$$

where $X_{1,t}$ and $X_{2,t}$ are first order auto-regressive processes AR(1) with parameters $\theta_1 = 0.9$ and $\theta_1 = -0.9$ respectively. In all time-frequency plots in this paper, darker shades represent higher power (or stronger coherence). The power of the spectrum at the first half of the process is concentrated at lower frequencies. At the second half, power is shifted to higher frequencies.

Obviously a limitation of piecewise stationary processes is that the spectral properties change abruptly at the change points and hence may not be reasonable models for processes that evolve slowly over time. Blended stationary processes, which will be discussed in the next subsection, overcome these difficulties because they allow some overlap between two adjoining stationary processes.

2.3 Blended Stationary Processes

A good model for most biological processes is one that allows for spectra that are not "blocky" but have smooth transitions between adjoining stationary blocks. In this section, we define *blended* stationary processes to be

$$\widehat{X}_{t,T}^B = \sum_{b=1}^{B} W^b \left(\, [t/T - u_{b-1}]/[u_b - u_{b-1}] \, \right) X_t^b \ , \tag{6}$$

where X_t^b are stationary processes with spectra $f^b(\lambda)$. The weights W overlap and must satisfy: $\sum_{b=1}^{B} W^b (\, [t/T - u_{b-1}]/[u_b - u_{b-1}] \,) = 1$ for all t. Moreover, the number B of stationary regimes must fulfill $B \to \infty$, but $B/T \to 0$ as $T \to \infty$. This equivalent to saying that the upper bound for the segment lengths M fulfills: $\frac{M}{T} \to 0$ as $T \to \infty$.

We now state the result that the class of blended stationary processes (including piecewise stationary processes as a special case) can approximate locally stationary processes.

Theorem 2.1 *Let $\{X_{t,T}\}$ be a locally stationary process with $A(u,\omega)$ being Lipschitz in u. Then there exists a sequence of blended stationary process $\{\widehat{X}_{t,T}^B\}$ and a positive constant Q such that for all $T \geq 1$,*

$$\frac{1}{T} \sum_{t=0}^{T-1} E \left(X_{t,T} - \widehat{X}_{t,T}^B \right)^2 \ \leq \ Q \ M^2/T^2 \ . \tag{7}$$

Proof. Given in the appendix.

Consequently, the time-varying spectrum $f(u,\omega)$ of a locally stationary process can be approximated by:

$$f(u,\omega) \approx \sum_{b=1}^{B} W^b\,(\,[u-u_{b-1}]/[u_b-u_{b-1}]\,)\,f^b(\omega). \tag{8}$$

3 The SLEX Transform

3.1 The SLEX basis vectors

Fourier vectors are ideal at representing stationary time series because they are perfectly localized in frequency and the spectral properties of stationary processes do not change with time. Fourier vectors, however, do not adequately represent processes whose spectral properties evolve with time. In order to alleviate the time localization problem, smooth compactly supported windows have been applied to the Fourier vectors (Daubechies, 1992). Windowed Fourier vectors, however, are generally no longer orthogonal. In fact, the Balian-Low theorem says that there does not exist a smooth window such that the windowed Fourier basis vectors are both (i.) orthogonal and (ii.) localized in time and frequency (Wickerhauser, 1994). The SLEX basis vectors, on the other hand, are simultaneously orthogonal and localized in time and frequency. They evade the Balian-Low obstruction because they are constructed by applying two, rather than just one, special smooth windows to the complex exponentials.

A SLEX basis function $\phi_k(u)$ has the form

$$\phi_k(u) = \Psi_+(u)\exp(i\ 2\pi\lambda_k u) + \Psi_-(u)\exp(-i\ 2\pi\lambda_k u) \tag{9}$$

where $\lambda_k \in (-1/2, 1/2]$. The SLEX basis function has support on $[-\eta, 1+\eta]$, where $0 < \eta < 0.5$. Thus, SLEX functions at different dyadic blocks can overlap. Though they overlap, the SLEX basis functions remain orthogonal.

SLEX basis functions generalize directly to orthogonal SLEX basis vectors for representing time series. Define $\alpha_1 > \alpha_0$ to be two time points, and let $S = \{\alpha_0, \alpha_0+1, \ldots, \alpha_1 - 1\}$ be a block of time points. Define $|S| = \alpha_1 - \alpha_0$ and $\epsilon = [\eta\ |S|]$, where $[\ \cdot\ ]$ denotes the greatest integer less than or equal to its argument. A SLEX basis vector has elements $\{\phi_{S,k,t}\ ;\ t = \alpha_0 - \epsilon, \ldots, \alpha_0, \ldots, \alpha_1 - 1, \ldots, \alpha_1 - 1 + \epsilon\}$ where

$$\begin{aligned}
\phi_{S,k,t} &= \phi_k\left((t-\alpha_0)/|S|\right) \\
&= \Psi_+\left((t-\alpha_0)/|S|\right)\ \exp\{i\omega_k(t-\alpha_0)\}\ + \\
&\ \Psi_-\left((t-\alpha_0)/|S|\right)\ \exp\{-i\omega_k(t-\alpha_0)\}
\end{aligned}$$

where $\omega_k = 2\pi k/|S|$. Different blocks are defined by different values of α_0 and α_1.

SLEX basis functions and vectors overlap, and thus they allow for smooth transitions between adjacent blocks. They are orthogonal, despite the overlap, because of the inclusion of the smooth window Ψ_-. They are localized in time because the windows Ψ_+ and Ψ_- have compact support. Finally, they are localized in frequency because the windows are smooth.

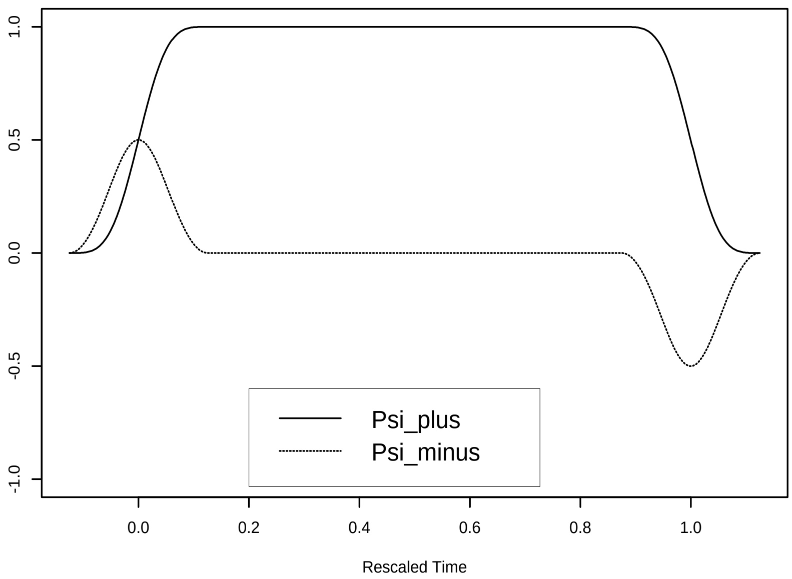

Figure 3: The tapers $\Psi_+(u)$ and $\Psi_-(u)$ used in the construction of the SLEX function that has support on $[-\eta, 1+\eta]$ where $\eta = 0.125$.

3.2 Computing the SLEX transform

The first property of the SLEX transform is that it smoothly partitions the time axis by allowing the SLEX vectors to overlap. The SLEX transform divides the time series in a dyadic fashion. Suppose that each component of the time series $\{X_t, Y_t\}_{t=0}^{T-1}$ has length $T = 2048 = 2_0^J$ where $J_0 = 11$. (To simplify the discussion, we only consider the SLEX transform of one component of the bivariate time series.) At resolution level $j = 0$, the time series is kept undivided as one single block or segment. At resolution level j, the time series is divided into 2^j overlapping segments. We define $S(j,b)$ to be the b^{th} block on level j. The SLEX coefficients on block $S(j,b)$ are defined as:

$$d_{j,b}^{(X)}(\omega_k) = \frac{1}{\sqrt{M_j}} \sum_t X_t \overline{\phi_{j,b,k,t}} \tag{10}$$

where $M_j = T/2^j$ and $\phi_{j,b,k,t} = \phi_{S(j,b),k,t}$ is the SLEX basis vector on block $S(j,b)$ oscillating at frequency $\omega_k = 2\pi k/M_j$ where $k = -(M_j/2-1), \ldots, M_j/2$.

3.3 Best Segmentation

The SLEX transform forms a library of orthonormal transforms. After computing the SLEX transform, a well-defined cost (Equation 14) is computed at each of the blocks. The cost for a particular segmentation is the *sum* of the costs at all the blocks that define that segmentation. The SLEX table given in Figure 5 is an example of a SLEX library and the

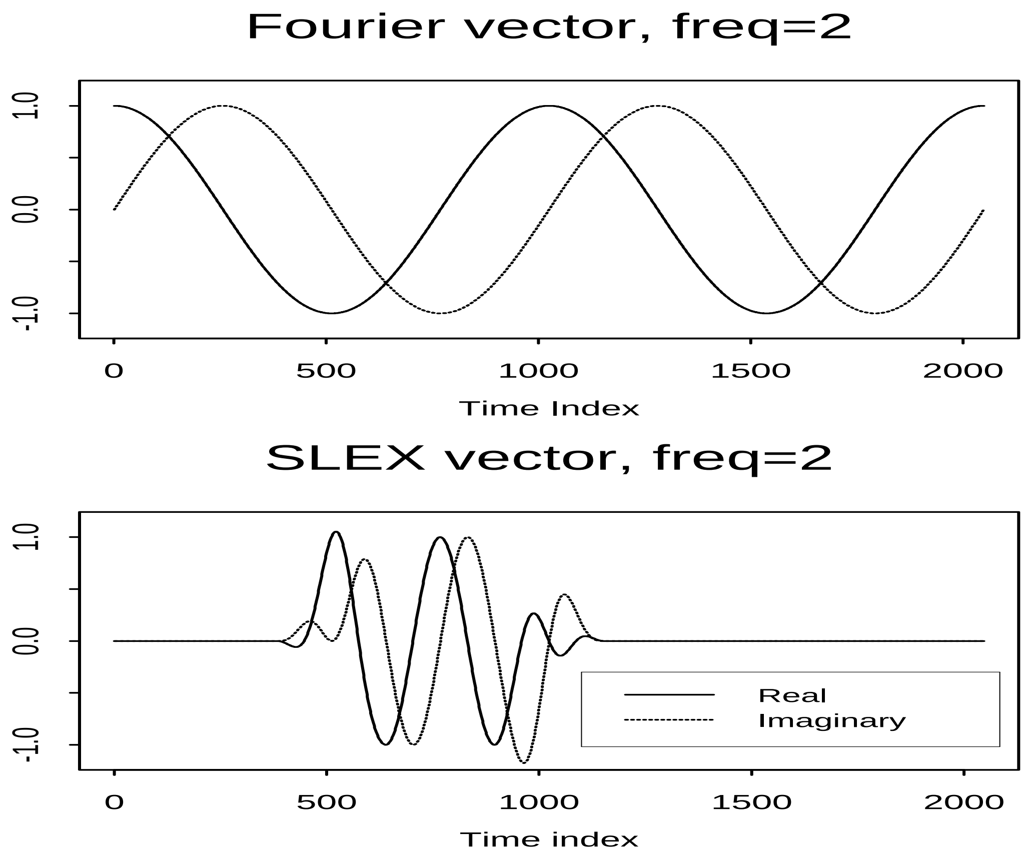

Figure 4: Top: Real and imaginary parts of a Fourier vector defined at time $t \in \{0, \ldots, 2043\}$. Bottom: Real and imaginary parts of a SLEX vector defined at time $t \in \{\alpha_0 - \epsilon, \ldots, \alpha_1 + \epsilon\}$ where $\alpha_0 = 512$, $\alpha_1 = 1023$ and $\epsilon = 64$.

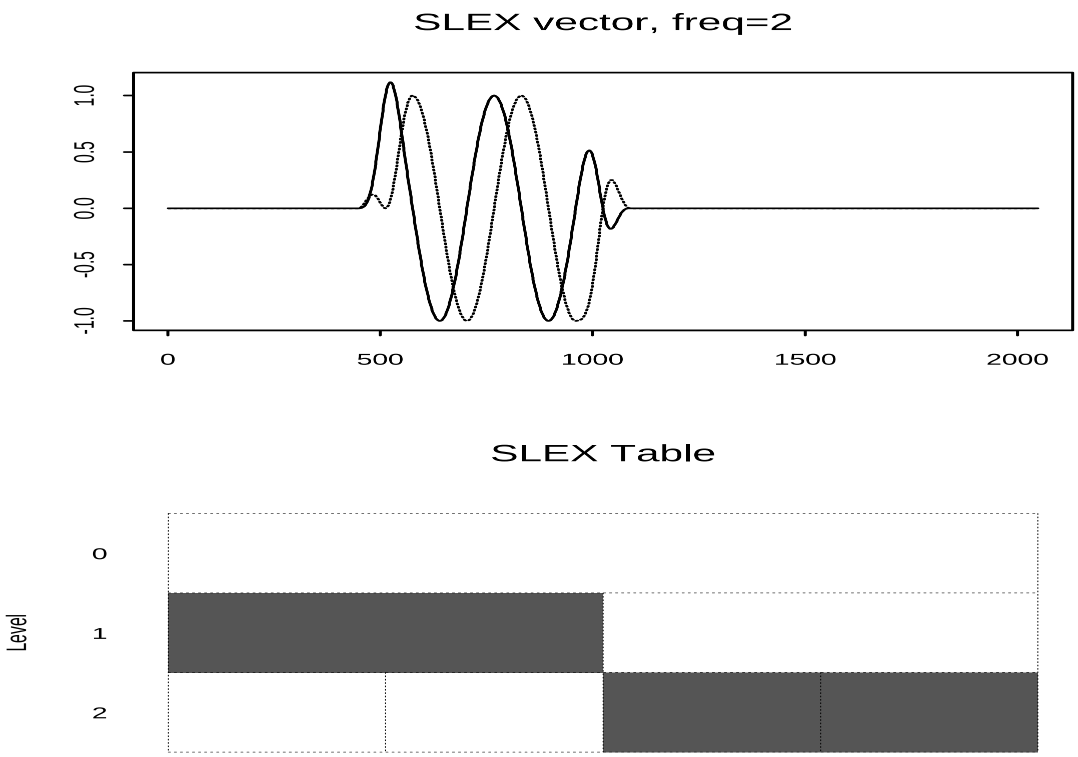

Figure 5: Top: Real and imaginary parts of a SLEX basis vector on block indexed by time $t \in \{512, \ldots, 1023\}$. Bottom: The SLEX table where the coefficients are arranged according to increasing frequency. The coefficient corresponding to the SLEX vector above is put on the second block of level 2. The shaded blocks represent one orthonormal transform.

shaded blocks represent one orthonormal transform (or one segmentation). The cost for this particular transform is the sum of the costs at each of the shaded blocks. The Best Basis Algorithm (BBA) of Coifman and Wickerhauser (1991) is an efficient algorithm that searches for the segmentation, among the orthonormal transforms in the library, that has the smallest cost. Wickerhauser (1994) discusses BBA and the cost functions in detail.

Remarks:

(1.) The choice of the deepest level of the transform, J, should be governed by the following principles. The length of each of the blocks at level $j, j = 0, \ldots, J$, is $M_j = T/2^j$. Hence, as level J increases, the number of observations on the finest level decreases. While this is positive from the point of view that bias due to non-stationarity can be controlled, fewer observations in a block means reduced frequency resolution and increased variance of the spectral estimates.

(2.) The amount of overlap, ϵ, at each level need not be the same. In order to ensure orthogonality of the SLEX vectors, it is sufficient that the maximum amount of overlap ϵ_j at any level j must satisfy $\epsilon_j \leq \epsilon_J \leq M_J/2$.

(3.) DMvS motivates the use of BBA as follows: Desirable theoretical results, such as consistency of the spectral estimators, are obtained in part due to the fact that the Fourier transform approximately diagonalizes the covariance matrix of a stationary process. This idea can be transfered to locally stationary processes. It turns out that *the* transform, among the orthonormal transforms in a library, that minimizes the cost is the "best" diagonalizer of the covariance of the locally stationary process. Please refer to DMvS (Sections 2, 10 to 12 and 14) for details.

4 The Auto-SLEX Method

The Auto-SLEX Method is an automatic statistical procedure that simultaneously segments the time series and gives the kernel-smoothed time-varying estimate of the spectra and coherence. Given a non-stationary bivariate time series (X_t, Y_t), we aim to estimate the time-varying spectra f_{XX} and f_{YY} of the univariate components, the cross-spectrum $f_{XY} = \overline{f_{YX}}$, and the coherence $R_{XY} = |f_{XY}|/\sqrt{f_{XX} f_{YY}}$.

The method consists of two general steps: (1.) automatic smoothing and (2.) automatic segmentation and estimation. The periodograms are first smoothed over frequency using a kernel smoother with bandwidth automatically selected by Generalized Cross Validation (GCV). Then the BBA is applied to the smooth periodograms to obtain the segmentation and the estimates of the time-varying spectra and coherence.

Automatic smoothing

The main output of this step are the kernel-smoothed periodograms and cross-periodograms using the bandwidth selected by Generalized Cross Validation (GCV). We note that the SLEX periodograms in each block are approximately independent and have a gamma distribution whose parameters depend on the unknown spectrum. Thus, the problem of spectral estimation can be considered in terms of a Generalized Additive Model (GAM) with a gamma distribution. Hastie and Tibshirani (1990) suggested sf choosing the smoothing parameter in GAMs by minimizing a form of the GCV function. The details of the GCV method is given in Ombao, Raz and von Sachs (1999).

We now set the notation. Let $d_{j,b}^{(X)}(\omega_k)$ and $d_{j,b}^{(Y)}(\omega_k)$ $(k=-(M_j/2+1),\ldots,M_j/2)$ be the SLEX coefficients on block $S(j,b)$. The raw periodograms and cross-periodograms are defined as $I_{j,b}^{(XX)}(\omega_k)=|d_{j,b}^{(X)}(\omega_k)|^2$, $I_{j,b}^{(YY)}(\omega_k)=|d_{j,b}^{(Y)}(\omega_k)|^2$ and $I_{j,b}^{(XY)}(\omega_k)=d_{j,b}^{(X)}(\omega_k)\overline{d_{j,b}^{(Y)}(\omega_k)}$ respectively. Let $\widehat{I}_{j,b,\nu}^{(XX)}$ and $\widehat{I}_{j,b,\nu}^{(YY)}$ be the periodograms smoothed using the bandwidth ν. We denote the "error degrees of freedom" by $df_{j,b,\nu}=1-\text{trace}(\mathbf{H}_\nu)/(M_j/2+1)$ where $\mathbf{H}_\nu$ is the smoother matrix.

The GCV functions for the components $\{X_t\}$ and $\{Y_t\}$ are, respectively,

$$GCV_{j,b}^{X}(\nu) = \frac{2}{(df_{j,b,\nu})^2}\sum_{k=0}^{M_j/2}\left\{\frac{I_{j,b}^{XX}(\omega_k)}{\widehat{I}_{j,b,\nu}^{XX}(\omega_k)} - \log\frac{I_{j,b}^{XX}(\omega_k)}{\widehat{I}_{j,b,\nu}^{XX}(\omega_k)} - 1\right\} \tag{11}$$

$$GCV_{j,b}^{Y}(\nu) = \frac{2}{(df_{j,b,\nu})^2}\sum_{k=0}^{M_j/2}\left\{\frac{I_{j,b}^{YY}(\omega_k)}{\widehat{I}_{j,b,\nu}^{YY}(\omega_k)} - \log\frac{I_{j,b}^{YY}(\omega_k)}{\widehat{I}_{j,b,\nu}^{YY}(\omega_k)} - 1\right\}. \tag{12}$$

The best bandwidth for the block $S(j,b)$ minimizes the function

$$GCV_{j,b}^{XY}(\nu) = GCV_{j,b}^{X}(\nu) + GCV_{j,b}^{Y}(\nu). \tag{13}$$

Thus, the "automatic-smoothing" step gives kernel-smoothed periodograms and cross-periodograms that use the optimal bandwidth. We denote these by $\widetilde{I}_{j,b}^{(XX)}$, $\widetilde{I}_{j,b}^{(YY)}$ and $\widetilde{I}_{j,b}^{(XY)}$ respectively.

Remark: In order to ensure that the estimate of the spectral density matrix is non-negative, the kernel smoothers for the periodograms and cross-periodograms within each block must use the *same* bandwidth. Hence, we used a single minimizing function in Equation (13) that combines the GCV for the two components.

Automatic Segmentation and Estimation

The next step is to apply the BBA to the smoothed periodograms. In our simulations and applications, we used the log-energy cost plus a complexity penalty term as motivated in DMvS (1998). The cost at each block $S(j,b)$ is defined as:

$$\text{Cost}(j,b) = \left\{\sum_{k=-(M_j/2\,-\,1)}^{M_j/2} \log\widetilde{I}_{j,b}^{(XX)}(\omega_k) + \log\widetilde{I}_{j,b}^{(YY)}(\omega_k)\right\} + \beta\sqrt{M_j} \tag{14}$$

where β is the complexity penalty parameter. The cost for one segmentation is the *sum* of the costs for all blocks that define that segmentation. The BBA then searches for the segmentation with the smallest cost. The estimates of the time-varying spectra $\widetilde{I}_{j,b}^{(XX)}$ and $\widetilde{I}_{j,b}^{(YY)}$ and of the cross-spectra $\widetilde{I}_{j,b}^{(XY)}$ are extracted from blocks $S(j,b)$ that define the best segmentation selected by BBA. The estimate of the time-varying coherence is given by

$$\widetilde{R}_{j,b}^{(XY)}(\omega_k) = |\widetilde{I}_{j,b}^{(XY)}(\omega_k)|/\sqrt{\widetilde{I}_{j,b}^{(XX)}(\omega_k)\widetilde{I}_{j,b}^{(YY)}(\omega_k)}\ . \tag{15}$$

Remarks:
(1.) In order to make meaningful coherence estimates, it is necessary that only one segmentation should be made for the multivariate time series. In the procedure, we combined the cost each of the components of the bivariate time series to obtain a single segmentation.
(2.) The cost function (Equation 14) includes the term $\beta\sqrt{M_j}$. This penalty term safeguards the procedure from obtaining the best segmentation with too many or too few blocks. A small value of β leads to a procedure that tends to select a segmentation with too many blocks. Although this is positive from the point of view that there will be less bias due to non-stationarity, this implies having less observations within each block and hence variances of the estimates will be inflated. A large value of β leads to a procedure that will tend to select a segmentaion with very few blocks. Although variance of the estimates is reduced, having too few blocks may lead to bias due to non-stationarity (i.e, error due to not splitting a non-stationary block). The penalty parameter β can either be approximated or computed via a data-driven procedure. In this paper, we set $\beta = 1$, which is motivated by the thresholds set by Gao (1997). Excellent empirical results are obtained using $\beta = 1$.

Relationship of the Auto-SLEX method to the Adak and DMvS methods

A common thread that ties the three methods is the use of transforms that divide the time series in a dyadic manner. The transforms used, however, are different. Adak uses the non-orthogonal windowed Fourier transform and DMvS uses the real-valued cosine packet transform. The Auto-SLEX method uses the SLEX transform that is complex-valued and forms a library of orthonormal transforms.

Adak (1998), suggested using two types of cost functions: (1.) a distance measure between spectral estimates; (2.) a distance measure plus a complexity penalty term. Adak's method (2) is supposed to safeguard against choosing a segmentation that divides the time series into many blocks. It incorporates a data-driven procedure for selecting the best penalty parameter. The Auto-SLEX method also uses a cost function that includes a complexity penalty for too many blocks. Simulation results in Section 6 showed that the present implementation of the Auto-SLEX method, where the complexity penalty parameter is set to $\beta = 1$, worked very well and performed better than the Adak method. There is also a difference in the underlying principles for segmentation between the Adak and the Auto-SLEX methods. The Adak method searches for the best segmentation based on some well defined criterion, i.e., the cost computed for a parent block is a distance measure between the spectra of children blocks. The Auto-SLEX method, on the other hand, uses the BBA which searches for the basis that best represents the time series based on some well defined cost function. The best segmentation is implied in the best basis selected.

DMvS showed that consistent estimates for the time-varying covariances can be obtained by applying BBA using the squared cost function (the cost for a block is the sum of the squared periodograms in that block). DMvS also suggested using the log-energy cost with complexity penalty and provided theoretical justification. In addition to the theoretical motivation in DMvS, we implemented the log-energy cost because empirical results indicate that it is able to give a higher rate of correct segmentation than the squared cost.

Both the Adak method and the DMvS method (with log-energy as cost function) apply a segmentation procedure using raw periodograms. The Auto-SLEX procedure is more general because it includes raw periodograms (smoothing parameter set to zero) as a possible output of the automatic smoothing step. There is also empirical evidence that suggests BBA is able to give the correct segmentation at a higher rate when applied to smoothed SLEX rather than raw SLEX periodograms.

Finally, DMvS uses wavelet-denoising while the Auto-SLEX uses kernel-smoothing. Linear smoothers ensure that the estimate of the spectral density matrix is non-negative definite.

5 Consistency Theorem

In this section we prove consistency of the smoothed (univariate) SLEX periodogram as an estimator of the evolutionary spectrum $f_{XX}(u,\omega)$ defined in the Dahlhaus model of local stationarity given in Section 2. Our result shows that the effects of both the dyadic structure of the SLEX transform and the bell Ψ_- on the spectral estimator vanish asymptotically. Thus, the result addresses two possible criticism of the SLEX method.

To prove consistency of our SLEX spectral estimators, we modify existing theoretical results involving tapered Fourier periodograms, which are computed using just *one* smooth window (Dahlhaus 1997). Let us now set some necessary notation. Suppose that the univariate time series has length $T = 2^L$ for some integer L. Let $J = J_T$ be the finest level on which we consider the blocks $S(J,b)$ (as defined in Section 3.2) which are of length $M = M_T$; hence $M_T = T/2^J = 2^{L-J_T}$. For our asymptotics, we assume that, as $T \to \infty$, $M \to \infty$ but $M/T \to 0$.

Define a segment of observations centered at $[uT]$, called $S_{u;T} = \{[uT] - M/2, [uT] + M/2\}$. In Dahlhaus (1996), the windowed Fourier periodograms were computed on $S_{u,T}$. In our approach, we compute the SLEX periodograms $I_S(u,\omega_k)$ for frequencies k using observations on the same segment $S_{u,T}$.

As in the classical (Fourier) periodogram, we can obtain consistent estimators by smoothing the SLEX periodograms, across frequency, in each segment. We use a conventional kernel smoother:

$$\widehat{f}(u,\lambda) \;=\; b_f^{-1} \sum_{k=-M/2+1}^{M/2} K_f((\lambda-\omega_k)/b_f)\; I_S(u,\omega_k) \tag{16}$$

where $\omega_k = 2\pi k/M$; $K_f(\lambda)$ is the kernel function with compact support $[-1/2,1/2]$; $K_f(\lambda) = K_f(-\lambda)$ and $\int K_f(\lambda)d\lambda = 1$. Its bandwidth b_f is assumed to fulfill $b_f \to 0$ and $M\,b_f \to \infty$ as $T \to \infty$.

It is interesting to observe that in the resulting kernel estimator (16) we automatically include some sort of "smoothing" in rescaled time. We observe that in rescaled time, u, the shrinking length of the segment $|S_{u,T}|/T$ is proportional to $b_t = M_T/T$ which thus plays the role of a "bandwidth" in time.

Theorem 5.1 *(Consistency Theorem) Suppose $X_{t,T}$ is locally stationary process as defined in Definition 2.1, with mean zero and transfer function A whose partial and mixed derivatives up to second order are continuous. Then, for $k \neq 0$, $\omega \neq 0$,*

(i.) $E\ I_S(u,\omega_k) \ = \ f(u,\omega_k) \ + \ C_{\Psi,f_u''}\ b_t^2 \ + \ o\ (b_t^2) \ + \ O\ (\log(M_T)/M_T)$
(ii.) $E\ \widehat{f}(u,\omega) \ = \ f(u,\omega) \ + \ C_{\Psi,f_u''}\ b_t^2 \ + \ C_{K_f,f_\omega''}\ b_f^2 \ + \ o\ (b_f^2 + b_t^2 + \log(M_T)/M_T),$
(iii.) $\mathrm{var}(\widehat{f}(u,\omega)) \ = \ (b_f b_t T)^{-1} f^2(u,\omega)\ \widetilde{C}_{\Psi,K_f} \ + \ o\ ((b_f b_t T)^{-1})$

where $C_{\Psi,f_u''}$ denotes a constant which depends both on the window functions Ψ_+ and Ψ_- and the second derivative f_u'' of $f(u,\omega_k)$ with respect to time u; $C_{K,f_\omega''}$ is a constant which depends on the kernel function K_f and the second derivative f_ω'' of $f(u,\omega)$ with respect to frequency ω; and $\widetilde{C}_{\Psi,K_f}$ depends on the size of the windows Ψ_+ and Ψ_- and on the kernel function K_f.

Proof. Given in the appendix.

Remarks:
(1.) In Theorem 5.1 the smoothness conditions on the evolutionary spectrum, i.e., on the transfer function $A(u,\omega)$ could be substantially relaxed without losing the property of consistency (cf. Neumann and von Sachs (1997)).
(2.) The rate of consistency is $(b_f b_t T)^{-1/2} = (b_f M_T)^{-1/2}$ which is similar to the rate for smoothing a periodogram over M_T frequencies in a neighborhood of bandwidth b_f.
(3.) As in Dahlhaus (1996), optimal rates for b_t (i.e, for M_T) and b_f can be calculated as those which minimize the asymptotically leading term of the mean relative squared error $E\left(\widehat{f}(u,\lambda)/f(u,\lambda) - 1\right)^2$. Without going into details, following the usual technique of balancing the leading parts of squared bias and variance we arrive at $b_t \sim T^{-1/6}$ and $b_f \sim T^{-1/6}$.
(4.) The kernel estimator (16) has the same structure as a segmented periodogram-based statistic in the central limit theorem of Dahlhaus (1997, Theorem A.2). Hence with the same techniques applied on the higher cumulants of $\widehat{f}(u,\lambda)$, asymptotic normality of our kernel estimator can be shown. In fact, by a consequence of a central limit theorem, even for non-Gaussian time series $\{X_{t,T}\}$, the SLEX coefficients $\{d^{(X)}(\omega_k)\}$ are asymptotically normal.
(5.) Our simulation studies of Sections 6.2 and 6.3 support the theoretical results that the effect of the bell Ψ_- as well as the effect of the dyadic structure can be neglected for large enough sample size.
(6.) We have proven consistency given a fixed segmentation. A more general question arises concerning whether or not our method consistently estimates the spectrum when the segmentation is selected by BBA. Compare the approximation results of Section 2, in particular Theorem 2.1. We expect that our use of an orthogonal transform will be important in developing this more general theory. DMvS developed a related theory based on the orthogonal local cosine transform. Although we have not yet developed the more general theory, this argument provides additional support to our use of an orthogonal transform.

6 Simulations

A good methodology should (1.) sufficiently often give the correct segmentation of the time series and (2.) give good spectral and coherence estimates at each of the stationary blocks.

We give the details of the implementation of the Auto-SLEX method. In the simulation studies reported here, we **(i.)** used the log-energy plus complexity term as cost function (Equation 14) in the search of the best segmentation; **(ii.)** computed the cost for all blocks down to level $J = 3$; **(iii.)** used the Daniell window for smoothing; **(iv.)** obtained bandwidths for every simulated data set by the GCV method.

We compared these simulations with the same set-up for the Adak method where we **(i.)** used the Kolmogorov-Smirnov distance measure without complexity penalty in the search of the best segmentation; **(ii.)** computed the cost for all blocks down to level $J = 3$; **(iii.)** used the bell Ψ_+ as the taper; **(iv.)** used the Daniell window for smoothing; **(v.)** obtained bandwidths by the GCV method. In preliminary numerical experiments, we implemented the Adak procedure with the complexity penalty but noted that the segmentation results were not different from the procedure without complexity penalty. Furthermore, we were not able to exactly replicate the results of Adak (1998) from the information given in the paper, and the computer programs used to generate Adak's results were not available.

In evaluating (1.), we performed simulation studies that determined how well the Auto-SLEX method gave the correct segmentation and then compared the performance with Adak's method. We evaluated (2.) by computing the average squared error (ASE) of the log of the time-dependent spectrum over frequency $[0, 1/2]$ as follows:

$$ASE = [T(M_J/2+1)]^{-1} \sum_{t=0}^{T-1} \sum_{k=0}^{M_J/2} \left\{ \log \widehat{f}(t/T, \omega_k) - \log f(t/T, \omega_k) \right\}^2 \qquad (17)$$

where $\widehat{f}(u, \omega_k)$ is an estimate of the true time-dependent spectrum $f(u, \omega_k)$.

6.1 Time Segmentation

A. Piecewise Stationary Process

We obtained $N = 5000$ simulated data sets, each of length $T = 1024$, from a piecewise stationary process $\{Y_t\}_{t=1}^{t=1024}$, with change point at $u = 1/2$, where Y_t is defined as:

$$Y_t = \begin{cases} X_{1,t} + \xi_t & \text{if } 1 \leq t \leq 512 \\ X_{2,t} + \xi_t & \text{if } 513 \leq t \leq 1024 \end{cases} \qquad (18)$$

where $\{X_{1,t}\}$ is AR(2) with parameters $\theta_1 = 1.69$ and $\theta_2 = -0.81$; $\{X_{2,t}\}$ is AR(2) with parameters $\theta_1 = 1.38$ and $\theta_2 = -0.81$; and the ξ_t are *i.i.d.* standard normal. We defined an AR(p) process in Equation (5). The same piecewise AR(2) processes were used in Adak (1998). Table 1 compares the Auto-SLEX and Adak methods. It gives the percentage of simulated time series data sets (out of $N = 5000$) that were divided according to different segmentations. For example, in Table 1, the Auto-SLEX method had 73.3% of the simulated time series data sets divided into two halves.

Table 1: **Piecewise Stationary Process** $N = 5000$ simulated data sets were obtained from a piecewise stationary process with change-point at $u = 1/2$. For each simulated data set, we applied the Adak and Auto-SLEX methods and counted the number of blocks that define the best segmentation chosen by each method. The **bold** numbers indicate the number of simulated data sets that gave the correct segmentation.

Number of blocks	Adak %	Ave. ASE	Auto-SLEX %	Ave. ASE
1	0	–	0	–
2	**43.7**	0.044	**73.3**	0.035
3	24.4	0.069	13.7	0.049
4	21.4	0.109	11.4	0.059
≥ 5	10.5	0.160	1.6	0.084

Table 2: **Stationary Process** $N = 5000$ simulated data sets were obtained from a stationary AR(2) process. For each simulated data set, we applied the Adak and Auto-SLEX methods and counted the number of blocks that define the best segmentation chosen by each method. The **bold** numbers indicate the number of simulated data sets that gave the correct segmentation.

Number of blocks	Adak %	Ave. ASE	Auto-SLEX %	Ave. ASE
1	**71.4**	0.023	**84.0**	0.022
2	14.9	0.046	8.2	0.037
3	6.7	0.071	2.6	0.053
4	5.3	0.110	4.8	0.063
≥ 5	1.7	0.150	0.4	0.095

B. Stationary Process

We obtained $N = 5000$ simulated data sets, each of length $T = 1024$, from a stationary process $\{Y_t\}_{t=1}^{t=1024}$ where Y_t is defined as: $Y_t = X_t + \xi_t$ where $\{X_t\}$ is AR(2) with parameters $\theta_1 = 1.69, \theta_2 = -0.81$. Table 2 summarizes the average ASE and segmentation performance of both the Adak and Auto-SLEX methods.

C. Discussion of segmentation results

(1.) The Auto-SLEX method is able to segment the time series reasonably well though it happens to divide a block even if it is stationary. This error, however, is more tolerable than not-splitting because because "under-splitting" can introduce non-stationarity bias. With "over-splitting", there is a chance of recovery from error if the spectral estimates at the splits of a stationary block are close. A negative effect of unnecessary over-splitting is that the variance of the periodograms are inflated because there would be fewer observations in a stationary block.

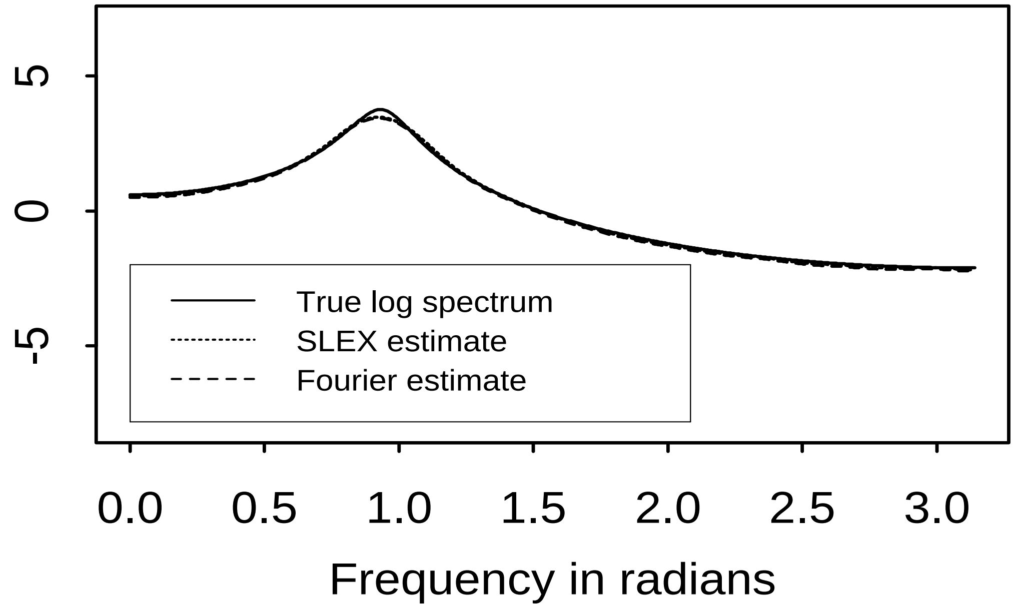

Figure 6: Log of the **mean** of Auto-SLEX estimates and the Fourier-based estimates of the spectrum in a stationary block, where the true process is AR(2)(1.07, -0.81). The means were computed from $N = 100$ simulated data sets each of length $T = 1024$.

(2.) The Auto-SLEX method gives a higher rate of correct segmentation than the Adak method. Furthermore, it has a lower ASE than the Adak method.

6.2 Estimation Under Stationarity

6.2.1 Estimation of the spectrum

The Auto-SLEX method was also evaluated for stationary processes in the preceding section. We noted that in Table 2, it gave the correct segmentation (i.e, it gave the undivided time series as the segmentation) sufficiently often. In this subsection, we present simulation results that compared the smoothed SLEX periodograms and coherence estimates with the smoothed Fourier periodograms and coherence estimates of a bivariate series. We also performed a simulation study on how well SLEX estimates the lag between two components of a bivariate stationary time series.

We simulated $N = 100$ time series data sets, each of length $T = 2048$, from an AR(2) ($\theta_1 = 1.07$, $\theta_2 = -0.81$) process. For each simulated data set, we obtained the estimate

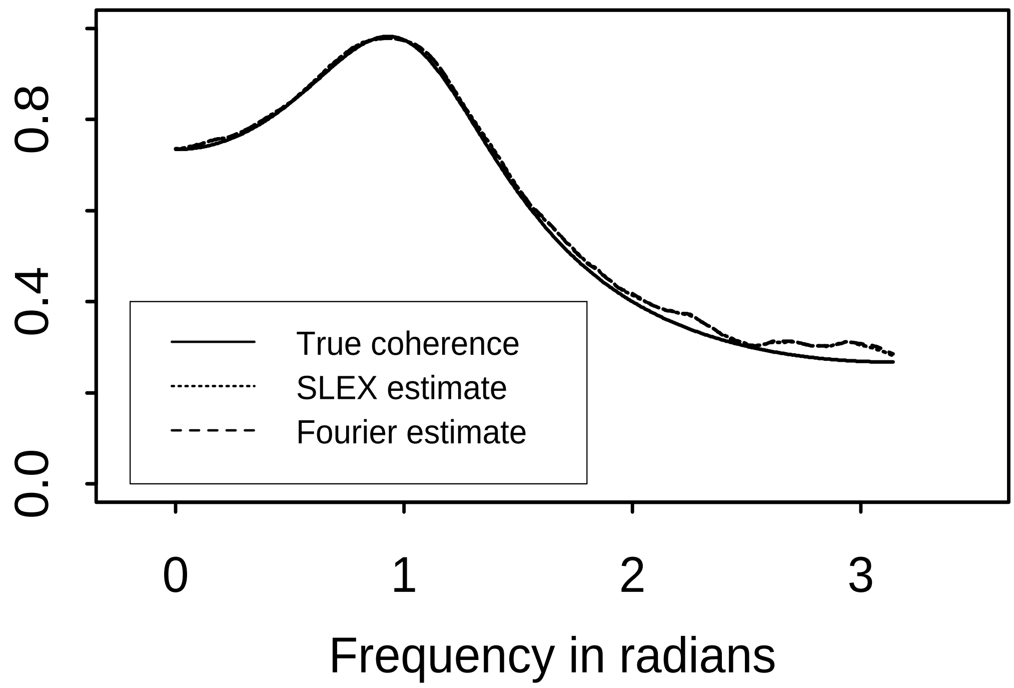

Figure 7: **Mean** of the Auto-SLEX estimates and Fourier-based estimate of the **coherence** in a stationary block, where the true bivariate process $\{X_t, Y_t\}$ is such that $\{X_t\}$ is AR(2)(1.07, -0.81) and $Y_t = 0.8X_t + \xi_t$. The means were computed from $N = 100$ simulated data sets each of length $T = 1024$.

of the log spectrum using (1.) the SLEX transform computed at level $J = 0$ only and (2.) the windowed Fourier transform with taper Ψ_+. Both the SLEX and windowed Fourier periodograms were smoothed using the Daniell window with the bandwidths automatically chosen by GCV. The means (out of $N = 100$) of the log-spectral estimates for each of the two methods are plotted in Figure 6. Note that the mean of the SLEX estimates is fairly close to both the true log-spectrum and the mean of the windowed Fourier estimates. This simulation study suggests that the difference between the SLEX periodograms and the windowed Fourier periodograms is negligible. It supports the theoretical results in Section 5 where we showed that the effect of the bell Ψ_- vanishes asymptotically.

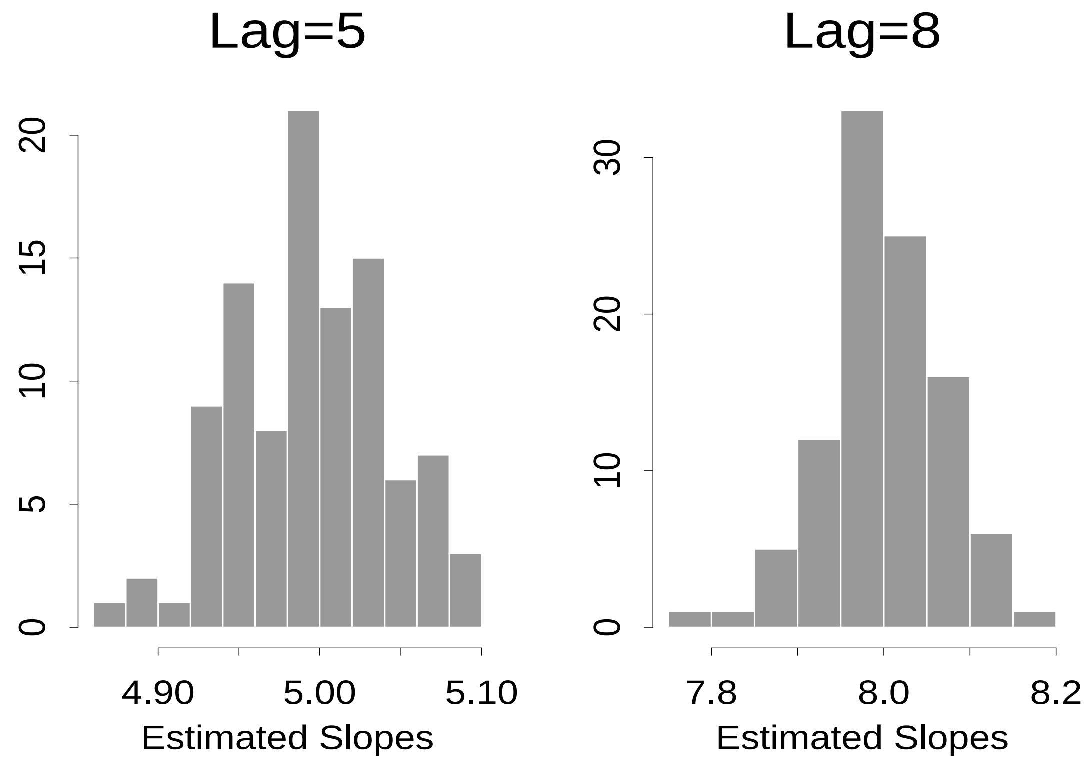

Figure 8: Auto-SLEX estimates of the lags when the true lags between the components of a bivariate stationary process are $\tau = 5$ and $\tau = 8$.

6.2.2 Estimation of the coherence

In simulation study, we simulated $N = 100$ time series data sets from a bivariate stationary process $\{X_t, Y_t\}_{t=1}^{t=1024}$ where $\{X_t\}$ is AR(2)($\theta_1 = 1.07$, $\theta_2 = -0.81$), $Y_t = 0.8X_t + \xi_t^{(Y)}$ and $\{\xi_t^{(Y)}\}$ are *i.i.d.* standard normal. The mean of the SLEX and the windowed Fourier estimates of coherence are compared in Figure 7. The SLEX coherence estimates are very close to the true coherence and are comparable to the windowed Fourier estimates.

6.2.3 Estimation of the lag

We also performed a simulation study where we investigated the performance of the SLEX transform in estimating the time lag between two components of a bivariate stationary time series. We simulated $N = 100$ time series data sets from a bivariate stationary process $\{X_t, Y_t\}_{t=1}^{t=1024}$, where $\{X_t\}$ is AR(1), ($\theta_1 = 0.9$); $Y_t = 0.9X_{t+\tau} + \xi_t^{(Y)}$; lag $\tau = 5$; and $\{\xi_t^{(Y)}\}$ are *i.i.d.* $N(0, 1)$. A standard approach in estimating the lag between two processes is to estimate their phase spectrum (phase part of the cross-spectrum) and then plot the estimated phase spectrum against frequency. The estimated *slope* of the phase plot is the

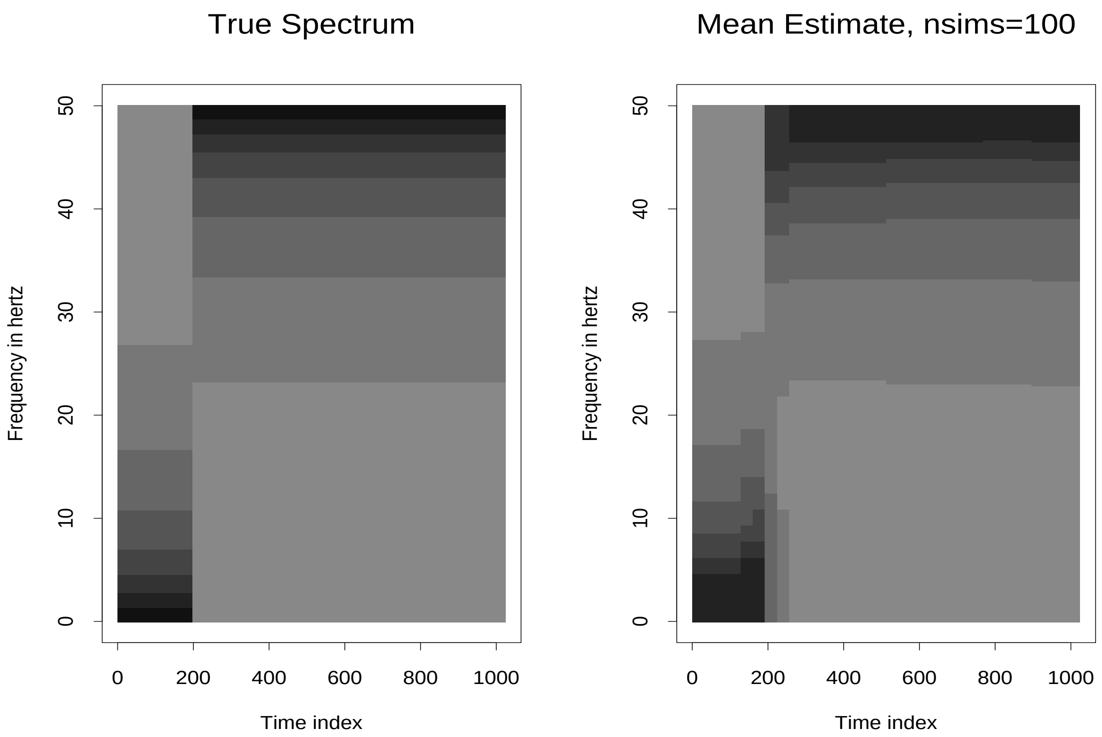

Figure 9: Left: True time-varying log-spectrum of a non-dyadic process. Right: Mean of the Auto-SLEX estimates from $N = 100$ simulated data sets.

estimated *lag*. In the simulation study, a linear model was fit for the phase plot of each simulated data set and the histogram of all the estimated slopes is plotted in Figure 8. The 2.5^{th} and 97.5^{th} percentiles for the slope (or lag) obtained were 4.9 and 5.1 respectively.

In another simulation study where the true lag was $\tau = 8$, the 2.5^{th} and 97.5^{th} percentiles for the slope (or lag) obtained were 7.9 and 8.1 respectively.

6.3 Non-dyadic segmentation

In practice it is somewhat unlikely that the segmentation of a piecewise stationary process follows a dyadic structure. An example of such a process is $\{X_t\}_{t=1}^{t=1024}$, where

$$X_t = \begin{cases} X_{1,t}, & \text{if } 1 \leq t \leq 197 \\ X_{2,t}, & \text{if } 198 \leq t \leq 1024 \end{cases} \tag{19}$$

where $X_{1,t}$ and $X_{2,t}$ are AR(1) processes with parameters $\theta_1 = 0.9$ and $\theta_1 = -0.9$ respectively. In Figure 9, we show the true piecewise spectrum and the mean Auto-SLEX estimate based on $N = 100$ simulated data sets. Note that the Auto-SLEX method selects segmentations with finer blocks close to the non-dyadic change-point. The Adak and DMvS methods

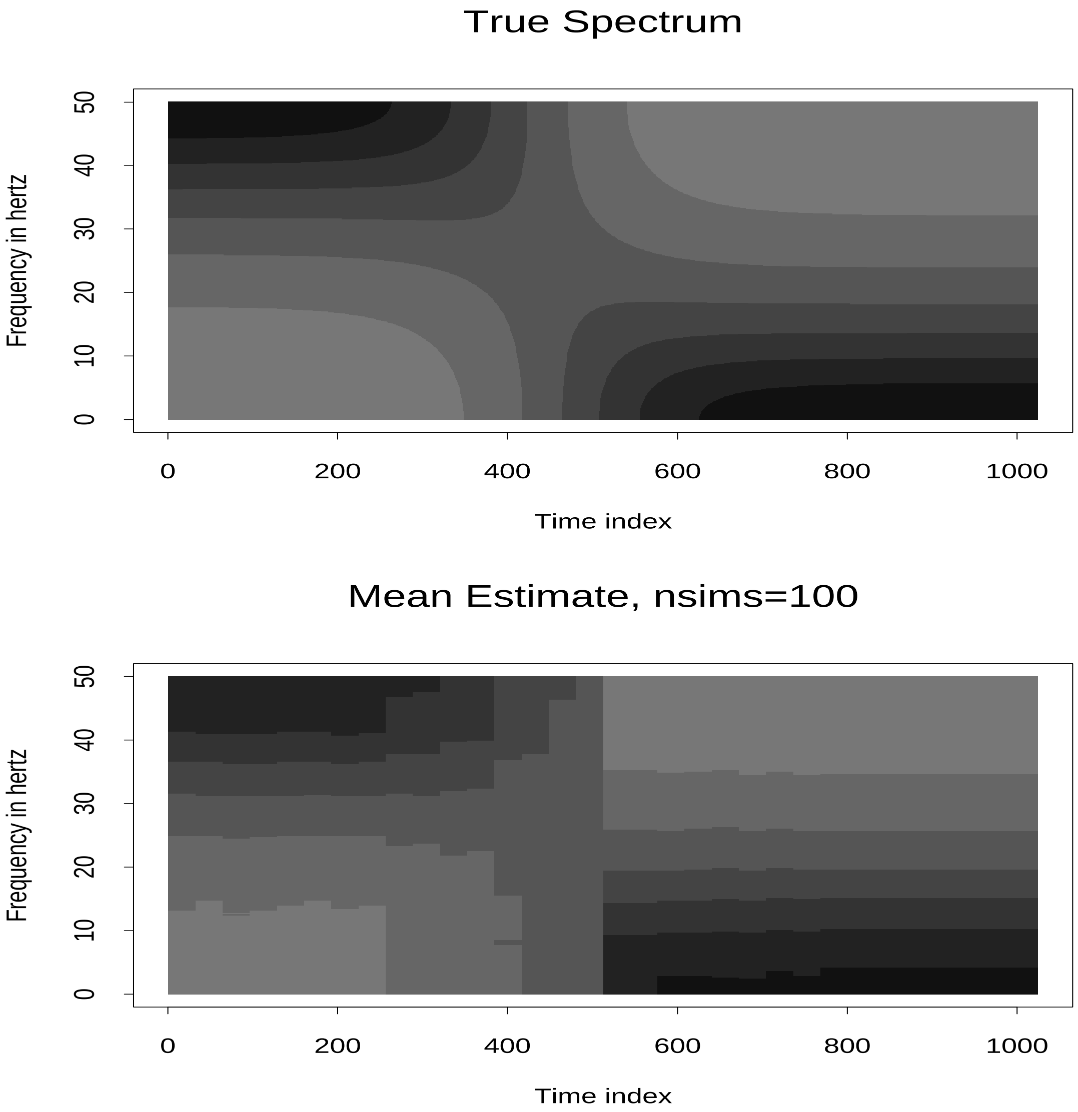

Figure 10: Left: The time-varying log-spectrum of a slowly changing AR(1) process. Right: Mean of the Auto-SLEX estimates from $N = 100$ simulated data sets.

both have this limitation as well. We believe, however, that the simplicity, flexibility, and computational efficiency of the BBA, as well as the advantages of working with a library of orthonormal bases, outweigh the disadvantages of the restriction to dyadic segmentations. Furthermore, we have shown in Section 5 that the effect of the dyadic segmentation on the estimation of the spectrum vanishes asymptotically. Future work, however, may lead to generalizing our approach to allow for non-dyadic segmentations. A possible approach is to include non-dyadic segmentation in the computation of the SLEX transform. Another possible approach is to use the cycle-spinning approach of Coifman and Donoho (1995).

6.4 Slowly-changing AR(1) process

Some processes change very slowly at each time point like the slowly-changing AR(1) process $Y_t = a_t Y_{t-1} + \epsilon_t$, where $a_t = 1.2\,(v_t/(1+v_t) - 0.5)$; $v_t = \exp\,(10t/1024 - 5)$. When t is close to 0, a_t is close to -0.6 and the power of the time-varying spectrum is concentrated on higher frequencies. When t is close to 1024, a_t is close to $+0.6$ and the concentration of power is shifted to lower frequencies. The true time-varying log spectrum and the mean of the smoothed estimates computed from $N = 100$ simulated data sets are shown in Figure 10.

7 Analysis of a bivariate EEG data set

We present an application of the Auto-SLEX method to EEG recordings at two channels, namely the left temporal ($T3$) and left parietal ($P3$), during an epileptic seizure. The EEG recordings are shown in Figure 1. Each EEG time series has length $T = 32768$ and was observed at the sampling rate of 100 Hertz. Our neurologist states that seizure, for this particular episode, started at the left temporal lobe at about time $t = 185$ seconds.

The Auto-SLEX estimates of the time-varying log spectra, plotted in Figure 11, showed a change in the power at about $t = 185$ seconds – a result consistent with our neurologist's observation. One can note that prior to seizure, power was concentrated at the lower frequencies. During seizure, however, power spread to all frequencies, consistent with the recruitment of large populations of neurons with diverse firing frequencies. As seizure drew to an end, the concentration of power was slowly restored to lower frequencies. Another interesting observation is that dramatic changes in the estimate of the log-spectrum at $P3$ came *later* than $t = 185$ seconds. This is also consistent with the fact that the patient was diagnosed with left temporal lobe epilepsy. Hence, seizure for this patient is expected to start at $T3$ and then eventually spread to other regions.

The Auto-SLEX estimates of the time-varying coherence (cf. Equation (15)), plotted in Figure 12, showed that the EEGs were strongly coherent at the alpha frequency band of about 9 to 12 Hertz prior to the onset of seizure. Furthermore, there was strong coherence at a narrow band of about 40 Hertz. One can observe too that the EEGs were strongly coherent at higher frequencies at the start of seizure. This is consistent with known clinical facts that the seizure, which started at $T3$ has not spread to $P3$ until a few seconds later. Thus, the EEGs were *not in synchrony* at the beginning of seizure. While dramatic changes in the electrical activity at the $T3$ channel were taking place at $t = 185$ seconds, there was "normal" electrical activity at the $P3$ channel at that time. An important observation is

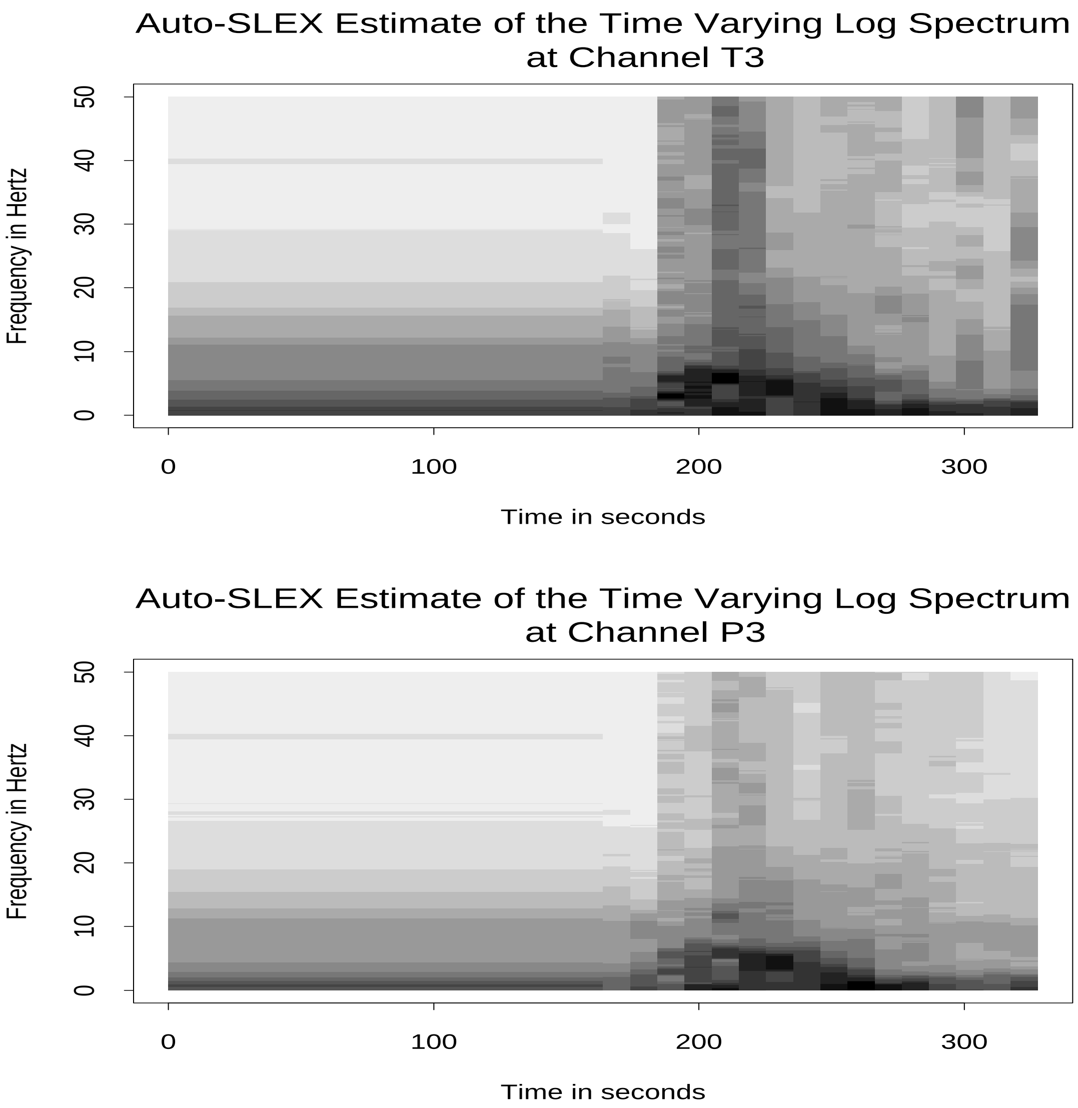

Figure 11: Auto-SLEX estimate of the time-varying log spectra of EEG taken at the left temporal channel ($T3$) and left parietal channel ($P3$).

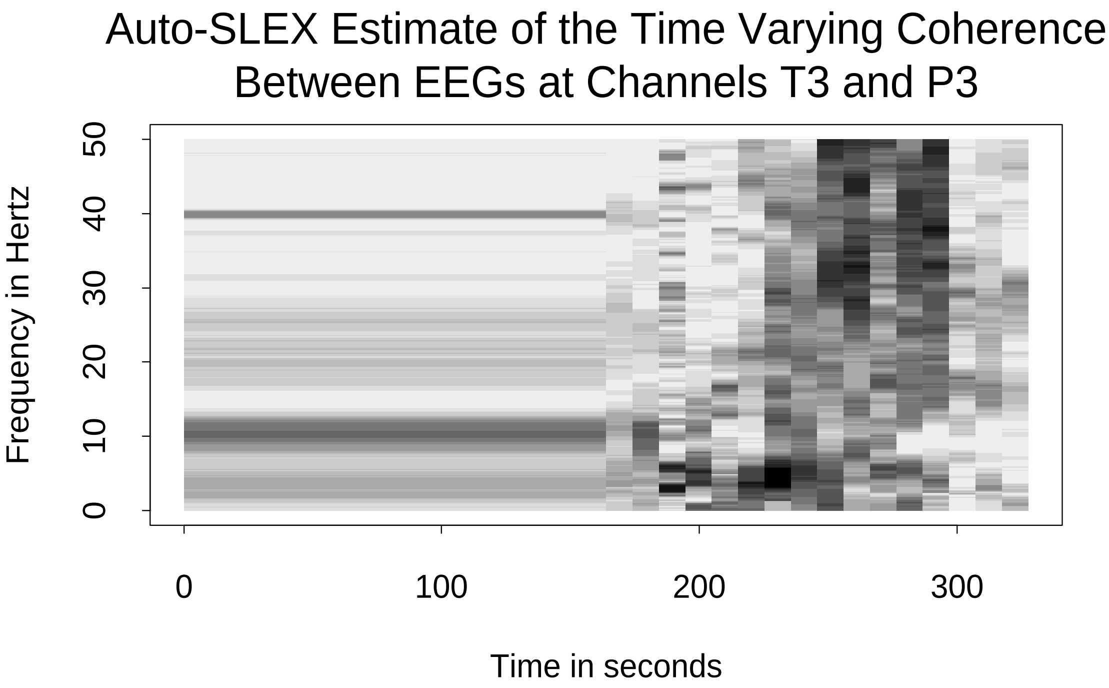

Figure 12: Auto-SLEX estimate of the time-varying coherence between EEGs taken at left temporal ($T3$) and left parietal channel ($P3$).

that the EEGs at $T3$ and $P3$ were strongly coherent at low frequencies at the onset towards the middle of the seizure period but the strong coherence shifted to higher frequencies at the end of the seizure. This is again consistent with the known fact that seizure is characterized by the evolution of waves with increasing frequencies. The two EEGs were *in synchrony* during seizure: both started with lower frequency waves evolving into higher frequency waves at the end of seizure.

8 Conclusion

The existing methods for analyzing non-stationary time series are built on the assumption that these non-stationary time series can be modeled by processes that are stationary within a small time interval. We have shown that the Auto-SLEX method is attractive both from practical and theoretical points of view. It is a procedure that is developed for bivariate time series and hence can give estimates of the spectra and coherence. It uses the SLEX transform which is localized in both time and frequency and hence is ideal for representing non-stationary processes. The method is also computationally efficient as it takes advantage of existing powerful algorithms: the fast Fourier transform and BBA. We have also given strong theoretical justifications for the Auto-SLEX method. We have rigorously shown that the smoothed SLEX periodograms are consistent estimators for the time-varying spectrum of the Dahlhaus locally stationary process. This important theoretical result tells us that the effect of the dyadic structure as well as the inclusion of another window, Ψ_{-}, vanish asymptotically. Finally, the Auto-SLEX method yields results that are easy to interpret because they parallel existing methods for stationary processes.

The Auto-SLEX method is also being further developed in both practical and theoretical directions. In practice, EEGs can be recorded at 128 channels. The method (both on the segmentation and estimation) has a natural extension from the bivariate to the multivariate setting. We are working on implementing SLEX (frequency) domain principal components analysis to extract the important information from data sets with high dimensionality. Furthermore, we are working on a SLEX model of locally stationary processes where the spectral representation is not in terms of the Fourier complex exponentials, as in Equation (1), but in terms of the SLEX vectors themselves. We have begun theoretical work on the open questions of consistency of the Auto-SLEX estimator when the segmentation is selected by the BBA and when our multivariate estimator is considered. We expect that our new SLEX-based model of locally stationary processes will facilitate our theoretical development.

APPENDIX

Proof of Theorem 2.1 By Parseval's identity we have,

$$\sum_{t=1}^{T} E(X_{t,T} - \widehat{X}_{t,T})^2 = \sum_{t=1}^{T} \int_{-1/2}^{1/2} \left| A_{t,T}^{0}(\omega) - \sum_{b=1}^{B} W^{b}(t) A(u_b, \omega) \right|^2 d\omega \ . \tag{20}$$

Since $\sum_{b=1}^{B} W^b(t) = 1$ for all t, we have:

$$\begin{aligned}
\sum_{t=1}^{T} E(X_{t,T} - \widehat{X}_{t,T})^2 &= \sum_{t=1}^{T} \int_{-1/2}^{1/2} \Big| \sum_{b=1}^{B} W^b(t) A^0_{t,T}(\omega) - \sum_{b=1}^{B} W^b(t) A(u_b, \omega) \Big|^2 d\omega \\
&= \sum_{t=1}^{T} \int_{-1/2}^{1/2} \Big| \sum_{b=1}^{B} W^b(t) [A^0_{t,T}(\omega) - A(u_b, \omega)] \Big|^2 d\omega \\
&\leq \sum_{t=1}^{T} \int_{-1/2}^{1/2} \sum_{b=1}^{B} \Big| W^b(t) \Big|^2 \Big| A^0_{t,T}(\omega) - A(u_b, \omega) \Big|^2 d\omega \ .
\end{aligned}$$

We now fix $t = t^*$ and we assume that $t^*/T \in [u_{b^*}, u_{b^*+1})$. Recall that $M = M_T$ denotes the upper bound of the length of (any) segment which must fulfill $M/T \to 0$ as $T \to \infty$. Let ϵ denote the overlap between the segment b and $b+1$, and let $\Delta = \epsilon/T$ be the rescaled overlap. Only one of the following holds: $t^*/T \in [u_{b^*}, u_{b^*} + \Delta)$, or $t^*/T \in [u_{b^*} + \Delta, u_{b^*+1} - \Delta)$ or $t^*/T \in [u_{b^*+1} - \Delta, u_{b^*+1})$. When $t^*/T \in [u_{b^*} + \Delta, u_{b^*+1} - \Delta)$, $W^b(t) = W^{b^*}(t)\ \delta(b = b^*)$. Hence,

$$\begin{aligned}
\sum_{b=1}^{B} \Big| W^b(t) \Big|^2 \Big| A^0_{t,T}(\omega) - A(u_b, \omega) \Big|^2 d\omega &= \Big| W^{b^*}(t) \Big|^2 \Big| A^0_{t,T}(\omega) - A(u_{b^*}, \omega) \Big|^2 \\
&\leq \Big| A^0_{t,T}(\omega) - A(t/T, \omega) \Big|^2 + \Big| A(t/T, \omega) - A(u_{b^*}, \omega) \Big|^2 \\
&\leq (D/T)^2 \ + \ L^2 \ |t/T - u_{b^*}|^2 \\
&\leq (D/T)^2 \ + \ L^2 \ M^2/T^2 \qquad (21) \\
&\leq G \ (M/T)^2 \qquad (22)
\end{aligned}$$

where G is some constant. Here we have used Equation (2) for the first term in the sum in Equation (21). For the second term, we used the Lipschitz continuity (with constant $L > 0$) of $A(u, \omega)$ in u and the fact that $t/T = t^*/T$ has maximal distance M/T from u_{b^*}.

The argument for the two other (overlapping) cases is similar, we can bound the error from above by the sum over the contribution from the two segments. The key point here is that one can form an approximately stationary small interval around the change points. In fact, at a small neighborhood about $u_{b^*+1} : [u_{b^*+1} - \Delta, u_{b^*+1} + \Delta]$, the transfer function $A^0_{t^*,T}$ is "close" to both $A(u_{b^*}, \omega)$ and $A(u_{b^*+1}, \omega)$ when the locally stationary process is *slowly* changing. Finally,

$$\begin{aligned}
T^{-1} \sum_{t=1}^{T} E(X_{t,T} - \widehat{X}_{t,T})^2 &\leq T^{-1} \sum_{t=1}^{T} \int_{-1/2}^{1/2} \sum_{b=1}^{B} \Big| W^b(t) \Big|^2 \Big| A^0_{t,T}(\omega) - A(u_b, \omega) \Big|^2 d\omega \\
&\leq T^{-1} \sum_{t=1}^{T} \int_{-1/2}^{1/2} (G \ M/T)^2 d\omega \\
&= Q \ M^2/T^2
\end{aligned}$$

Proof of Theorem 5.1. The proof works by analogy to Dahlhaus (1996), Theorem 2.2, since the SLEX periodogram is analogous to a tapered (Fourier) periodogram where the windows $\Psi_{S,+}(t/T)$ and $\Psi_{S,-}(t/T)$ play the role of two taper functions. These fulfill the assumptions of classical smooth tapers. Furthermore, these two tapers are orthogonal: $\sum_{t\in S}\Psi_{S,+}(t/T)\Psi_{S,-}(t/T) = 0$. This is simply a consequence of the different parity of the two tapers. $\Psi_+ = \Psi_{S,+}$ is even while $\Psi_- = \Psi_{S,-}$ is odd on the segment S. Moreover, $\sum_{t\in S}(\Psi^2_{S,+}(t) + \Psi^2_{S,-}(t)) = M_T = T/2^{J_T}$. Note that the sum is independent of both t and ϵ and that our SLEX periodograms are correctly normalized by the sums of squares of the taper values, as is usual for tapered periodograms.

For clarity, we first compare how the SLEX periodograms differ from the tapered periodograms in the case of a stationary process. Under stationarity, Parseval's equality gives:

$$\begin{aligned} E\ I_S(\omega_k) &= M_T^{-1}\sum_t\sum_s \gamma(t-s)\phi_{S,k}(t)\overline{\phi_{S,k}(t)} && (23)\\ &= M_T^{-1}\int f(\alpha)\ \Delta(\omega_k,\alpha)\ d\alpha\ , && (24)\end{aligned}$$

where

$$\begin{aligned} \gamma(t-s) &= \operatorname{cov}(X_t,X_s) \quad \text{and} && (25)\\ \Delta(\lambda,\alpha) &= |\widehat{\Psi_+}(\lambda-\alpha)|^2 + |\widehat{\Psi_-}(\lambda-\alpha)|^2\ + 2\widehat{\Psi_+}(\lambda-\alpha)\widehat{\Psi_-}(\lambda+\alpha). && (26)\end{aligned}$$

Here, $\widehat{\Psi_+}$ and $\widehat{\Psi_+}$ denote the finite Fourier transforms of the windows Ψ_+ and Ψ_-, respectively. As known from classical results, e.g. Brillinger (1981), Theorem 4.3.2,

$$M_T^{-1}\int (|\widehat{\Psi_+}(\lambda-\alpha)|^2\ +\ |\widehat{\Psi_-}(\lambda-\alpha)|^2)\ f(\alpha)\ d\alpha\ =\ f(\lambda)\ +\ O\ (M_T^{-1})\ ,$$

since $|\widehat{\Psi_+}(0)|^2\ +\ |\widehat{\Psi_-}(0)|^2 = M_T$. Note that the regularity assumptions on both the window functions and the spectrum are fulfilled.

For the cross term we observe that, for $\lambda \neq 0$,

$$M_T^{-1}\int f(\alpha)\ \widehat{\Psi_+}(\lambda-\alpha)\widehat{\Psi_-}(\lambda+\alpha)\ d\alpha\ =\ o\ (M_T^{-1})\ , \tag{27}$$

as the Fourier transform of Ψ_+ asymptotically concentrates around λ whereas that of Ψ_- concentrates around $-\lambda$.

We continue with the variance of the SLEX-periodogram in the stationary case. For simplicity, consider the situation of a Gaussian process $\{X_{t,T}\}$, for which we can use Isserlis Theorem (Brillinger (1981), p. 21):

$$\operatorname{var}(I_S(\omega_k))\ =\ \operatorname{cov}(d_S(\omega_k),d_S(\omega_k))\operatorname{cov}(\overline{d_S}(\omega_k),\overline{d_S}(\omega_k) + \operatorname{cov}(d_S(\omega_k),\overline{d_S}(\omega_k))\operatorname{cov}(\overline{d_S}(\omega_k),d_S(\omega_k))\ .$$

Using the above result for the expectation $E\ I_S(\omega_k) = \operatorname{cov}(d_S(\omega_k),d_S(\omega_k)) = \operatorname{cov}(\overline{d_S}(\omega_k),\overline{d_S}(\omega_k))$, we observe that the first term in the sum tends to $f^2(\omega_k)$ with a rate of order M_T^{-1}. The second term tends to zero with this rate, as

$$\operatorname{cov}(d_S(\omega_k),\overline{d_S}(\omega_k))\ =\ E\ d_S^2(\omega_k)\ =\ M_T^{-1}\int f(\alpha)\ \widetilde{\Delta}(\omega_k,\alpha)\ d\alpha\ ,$$

with

$$\widetilde{\Delta}(\lambda,\alpha) \;=\; \widehat{\Psi_+}(\lambda+\alpha)\widehat{\Psi_+}(\lambda-\alpha) \;+\; \widehat{\Psi_-}(-\lambda-\alpha)\widehat{\Psi_-}(-\lambda+\alpha) \;+\; 2\;\widehat{\Psi_+}(\lambda-\alpha)\widehat{\Psi_-}(\lambda-\alpha)\;.$$

Here, the first two terms behave exactly as the cross term in equation (27) whereas for the final term we observe, by, e.g., once again the proof of Theorem 4.3.2 of Brillinger (1981), pp 402 f.,

$$M_T^{-1}\int f(\alpha)\,\widehat{\Psi_+}(\lambda-\alpha)\widehat{\Psi_-}(\lambda-\alpha) \;=\; M_T^{-1}\sum_t \Psi_+(t/T)\Psi_-(t/T)\,f(\lambda)\,d\alpha + O\left(M_T^{-1}\right) \;=\; O\left(M_T^{-1}\right),$$

using $\sum_t \Psi_+(t/T)\Psi_-(t/T) = 0$.

In total,

$$\operatorname{var}\left(I_S(\omega_k)\right) \;=\; f^2(\omega_k) \;+\; O\left(M_T^{-1}\right)\;.$$

Turning to our situation of local stationarity, we adapt the proof of part (i) of Dahlhaus (1996), Theorem 2.2, to the situation of our particular taper functions, to arrive at the same leading terms as in the stationary case. Note again, that the assumptions for the tapers of Dahlhaus (1997), Lemma A.5 and A.9 are fulfilled. Actually, a taper function of bounded variation would be sufficient. We only give a brief sketch.

Proof of (i): To treat $E\;I_S(u,\omega_k)$ we need to substitute in Equation (23) the term $\gamma(t-s)$ by

$$\begin{aligned}
\gamma(t,s) \;&=\; \operatorname{cov}(X_{t,T},X_{s,T}) \\
&=\; \int A^o_{[uT]-M/2+t,T}(\alpha)\;A^o_{[uT]-M/2+s,T}(-\alpha)\;\exp(-i\alpha(t-s))\;d\alpha\;.
\end{aligned}$$

Further, we replace $A^o_{[uT]-M/2+t,T}(\alpha)$ by $A(u-(M/2+t)/T,\alpha)$ and use the same techniques as in Dahlhaus (1996) to bound the replacement error by a term of order $O(\log(M_T)/M_T)$. We do the same for the factor with s instead of t, and integrate over α. Taylor expansion of $f(u-(M/2+t)/T,\omega_k)$ around u allows us to use the same ideas as in the stationary case and to finally arrive at the result of part (i) with

$$C_{\Psi,f''_u} \;=\; 1/2\int_0^1 (x-1/2)^2(\Psi_+^2(x)+\Psi_-^2(x))\;dx\;\,f''_u(u,\omega_k)\;.$$

Proof of (ii.): Part (ii.) follows from (i.) again by Taylor expansion of the spectrum, now around λ. Here,

$$C_{K_f,f''_\lambda} \;=\; 1/2\int \alpha^2\;K(\alpha)\;\;f''_\lambda(u,\lambda)\;d\alpha$$

Proof of (iii.): Part (iii.) is a consequence of the proof of Lemma A.9 in Dahlhaus (1997) (case of length of overlap $S=N$ following his notation), again with the necessary modification due to the presence of two terms Ψ_+ and Ψ_- in the taper function. It can be observed that again the principle applies as shown for the variance of $I_S(\omega_k)$ in the stationary case. Hence, for the factor which inflates the asymptotic variance of the kernel estimator $\widehat{f}(u,\lambda)$, we simply get

$$\lim_{T\to\infty} M_T^{-1}\sum_t(\Psi_+^2(t/T)\;+\;\Psi_-^2(t/T))^2 \;=\; \int_0^1(\Psi_+^2(x)\;+\;\Psi_-^2(x))^2\;dx$$

which equals $\widetilde{C}_{\Psi,K_f} \; / \int K^2(\lambda)\; d\lambda$. This ends the proof of the Theorem 5.1.

Literature Cited

Adak, S. (1998). Time dependent spectral analysis of non-stationary time series, *Journal of the American Statistical Association,* **93**, 1488-1501.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory,* Holden-Day, Oakland, CA.

Cohen, L. (1989). Time-frequency distributions: a review. *Proceedings of the IEEE,* **77**, 941-981.

Coifman, R. and Donoho, D. (1995). Translation-invariant de-noising. *Wavelets and Statistics,* Anestis Antoniadis and Georges Oppenheim, eds., Lecture Notes in Statistics **103**, pp. 125-150, New York, Springer-Verlag.

Coifman, R. and Wickerhauser, M. (1991). Entropy based algorithms for best basis selection. *IEEE Transactions on Information Theory,* **32**, 712-718.

Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In: *Athens Conference on Applied Probability and Time Series Analysis*, Vol. II. Lecture Notes in Statistics **115**, pp. 145-159. New York, Springer-Verlag.

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

Dahlhaus, R. (1999). A likelihood approximation for locally stationary processes. Beiträge zur Statistik 56, Universität Heidelberg.

Daubechies, I. (1992). *Ten Lectures on Wavelets,* Society for Applied and Industrial Mathematics, Philadelphia, PA.

Donoho, D., Mallat, S. and von Sachs, R. (1998). Estimating covariances of locally stationary processes: rates of convergence of best basis methods. *Technical Report 517, Statistics Department, Stanford University.*

Gao, H. (1997). Choices of Thresholds for the Wavelet Shrinkage Estimate of the Spectrum. *Journal of Time Series Analysis,* **18**, 231-251.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Neumann, M. H. and von Sachs, R. (1997). Wavelet Thresholding in Anisotropic Function Classes and Applications to Adaptive Estimation of Evolutionary Spectra. *Annals of Statistics,* **25**. 38-76.

Ombao, H., Raz, J. and von Sachs, R. (1999). A Simple GCV Method in Span Selection for Periodogram Smoothing. *Manuscript in preparation.*

Priestley, M. (1981). *Spectral Analysis and Time Series,* Academic Press.

Wickerhauser, M. (1994). *Adapted Wavelet Analysis from Theory to Software.* IEEE Press, Wellesley, MA.

Williams, W. (1997). Recent advances in time-frequency representations: some theoretical foundations. *Frequency and Wavelets in Biomedical Signal Processing,* Metkin Akay, ed., pp. 3-18. IEEE Press, Wellesley, MA.