T E C H N I C A L
R E P O R T

0337

SLEX ANALYSIS OF MULTIVARIATE NON-STATIONARY TIME SERIES

H. OMBAO, R. von SACHS and W. GUO

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

SLEX Analysis of Multivariate Non-Stationary Time Series

Hernando Ombao[1], Rainer von Sachs[2] and Wensheng Guo[3]

November 21, 2003

Abstract

We propose to analyze a multivariate non-stationary time series using the SLEX (**S**mooth **L**ocalized Complex **EX**ponentials) library. The SLEX library is a collection of bases; each basis consists of the SLEX waveforms which are orthogonal localized versions of the Fourier complex exponentials. In our procedure, we first build a family of multivariate SLEX models such that every model has a spectral representation in terms of a unique SLEX basis. The SLEX family provides a flexible representation for non-stationary random processes because every SLEX basis is localized in both time and frequency. The next step is to select a model using a penalized log energy criterion which we derive in this paper to be the Kullback-Leibler distance between a model and the empirical time series. In our procedure, we apply SLEX principal components analysis to obtain a decomposition of a possibly highly cross-correlated multivariate data set into non-stationary components with uncorrelated (non-redundant) spectral information. The best model is then selected by computing the log energy criterion based on the SLEX principal components. The proposed SLEX analysis for multivariate non-stationary time series closely parallels traditional Fourier analysis of stationary time series. Hence, our method gives results that are easy to interpret. Moreover, the SLEX method uses computationally efficient algorithms and hence can easily handle massive data sets. We illustrate the SLEX method by its application to a multivariate brain waves data set recorded during an epileptic seizure.

Keywords: Multivariate non-stationary time series; SLEX model; SLEX transform; Time-varying spectrum; Time-frequency analysis; Principal components analysis.

[1]H. Ombao is Assistant Professor, Department of Statistics and Department of Psychology, University of Illinois at Urbana-Champaign. Email: ombao@uiuc.edu. His research is supported, in part, by NSF DMS 0102511 and NIMH 62298.

[2]R. von Sachs is Professor, Institut de statistique, Université catholique de Louvain, Belgium. His research is supported, in part, by the Belgian government, via the "Projet d'Actions de Recherche Concertées" No. 98/03-217 and the IAP research network nr. P5/24.

[3]W. Guo is Assistant Professor, Department of Biostatistics, University of Pennsylvania. His research is supported, in part, by NIMH 62298.

1 Introduction

Many empirical multivariate time series are non-stationary. As an example, electroencephalograms (EEGs) or brain waves display amplitudes and frequency content that vary over time during an epileptic seizure. For illustration we present the plots of such a multichannel EEG in Figure 1, together with the scalp location of these EEGs, given in Figure 2. Obviously, these EEGs cannot be adequately analyzed using classical Fourier methods. Thus, we need to develop methods that can model the time-changing auto- and cross-spectral properties of multivariate non-stationary time series. Furthermore, we need to develop methods that are computationally efficient because typical EEG data sets are massive. It is not uncommon for EEG recordings to be taken at dozens of channels each with hundreds of thousands of observations. Finally, we would like to provide theoretical justifications for our method so that we can perform a rigorous analysis of empirical multivariate time series.

In this paper, we propose a new model for multivariate non-stationary processes that can capture the localized spectral and coherence features of the empirical time series. Our model explicitly defines a spectral density matrix whose elements, the spectra and cross spectra, are allowed to vary over time. Furthermore, we also present a computationally efficient method for estimating the unknown spectral quantities. Our entire procedure is based on the SLEX (**S**mooth **L**ocalized Complex **EX**ponentials) library (Ombao et al, 2001a). The SLEX library is a collection of bases; each basis consists of the SLEX waveforms which are time-localized versions of the Fourier complex exponentials, i.e. they are supported on smoothly-windowed blocks in time. The plot of the real and imaginary parts of the SLEX waveform is given in Figure 3. One very attractive feature of SLEX analysis is that it is a time-dependent analogue of Fourier analysis for stationary processes. Thus, it gives results that are easy to interpret. Moreover, the SLEX approach uses computationally efficient algorithms, thus it is capable of handling data sets that are massive.

The first step in the SLEX analysis is to build a family or collection of models. Each model in the SLEX family has a spectral representation in terms of a unique SLEX basis. Hence, each model, i.e. each basis, gives an explicit segmentation into time blocks for the underlying time series which is hence modelled to be nearly stationary within each time block. Our proposed approach selects a model from the SLEX family using a penalized log-energy criterion. One main contribution of this paper is that we derive the log-energy criterion as the Kullback-Leibler distance between a SLEX model and the unknown process that generated the empirical time series. In practice, the log energy criterion tends to select the model with very small time blocks. Thus, it has a tendency to split blocks that are almost stationary. To prevent this unnecessary splitting from happening too often, we will add a complexity penalty, resulting in a complexity penalized log energy criterion.

In this paper, we develop the theoretical motivation for the penalized log energy criterion. Our criterion explicitly takes into account both the auto-spectrum and the cross-spectrum. Thus our model selection criterion explicitly includes the time-varying cross-relationships between components of the multivariate time series. In addition, as second main contribution, we will approach the problem of high dimensionality in multivariate time series by applying the SLEX (frequency) domain principal components analysis (PCA). The SLEX PCA decomposes the multivariate time series into components (which we will call the SLEX principal components) that are non-stationary and uncorrelated in the frequency domain. Thus, the SLEX principal components contain non-redundant spectral information. The best model or the best segmentation will then be determined by the time-varying spectra of the SLEX principal components.

There has been a long history of analyzing high dimensional multivariate data by some sort of decomposition into uncorrelated components. Recently, independent component analysis (e.g., Cardoso, 1998) has gained some widespread attention particularly in computer science. In analyzing high dimensional data, frequency domain PCA plays some dominant role for treatment of dynamic models, such as vector autoregressive moving average (VARMA) or dynamic factor models in econometric applications; see e.g. Forni et al (2000). To our knowledge, however, there exists no statistical approach of explicitly modelling the deviation from stationarity in data with some time-varying auto- and cross-correlation structure. The work of Stock and Watson (1998) gives some very general framework for a possibly time-varying factor model description in economic index theory; some specific application to forecasting using time-domain principal components appeared in Stock and Watson (2002). ICA-based methods have proven useful for the purposes of "blind source separation" and signal extraction as they try to explain the data by components which are as closest to independent as possible. They are not designed to address the problem of high-dimensionality. In our approach we use PCA as a tool for finding the best segmentation, i.e. the best model, of a non-stationary multivariate time series of possibly high dimension. We finally point out the approach by West, Prado and Krystal (1999) which is complementary to our procedure. This approach, which explored latent structure in multivariate non-stationary time series, works well for short time series. However, this tends to be computationally costly and hence suggests subsampling the time series. Consequently, this can lead to a loss high-frequency information which is valuable when analyzing epileptic seizure EEGs.

To motivate now our multivariate SLEX model, we first take a look on the Cramér (spectral) representation of a stationary time series. A zero mean p-dimensional stationary random process $\mathbf{X}_t$ $(t = 0, \pm 1, \ldots)$ can be viewed as a sum of an infinite number of randomly weighted Fourier complex exponentials via the Cramér representation

$$\mathbf{X}_t = \int_{-1/2}^{1/2} \mathbf{A}(\omega)\ \exp(\ i2\pi\omega t)\mathbf{dZ}(\omega). \tag{1.1}$$

The elements in the Cramér representation are: the complex exponentials, which are the building blocks of this stochastic representation; the $p \times p$ transfer function matrix, $\mathbf{A}(\omega)$, that is Hermitian and constant in time; and the zero mean orthonormal random process, $\mathbf{dZ}(\omega)$. The spectrum of $\mathbf{X}_t$ is defined to be the matrix $\mathbf{f}(\omega) = \mathbf{A}(\omega)\mathbf{A}^*(\omega)$ where $\mathbf{A}^*$ denotes the complex conjugate of $\mathbf{A}$. For stationary processes, the spectral density matrix is constant in time. For non-stationary processes, Dahlhaus (2001) proposed the Cramér representation for a multivariate time series $\mathbf{X}_{t,T}$ (having length T) to be, approximately,

$$\mathbf{X}_{t,T} = \int_{-1/2}^{1/2} \mathbf{A}(t/T, \omega)\ \exp(\ i2\pi\omega t)\mathbf{dZ}(\omega). \tag{1.2}$$

The spectrum in the Dahlhaus model is defined at rescaled time $u = t/T$ and frequency ω is $\mathbf{f}_D(u, \omega) = \mathbf{A}(u, \omega)\mathbf{A}^*(u, \omega)$. Rescaling in time has been introduced as a purely theoretical device that allows to treat consistent estimation of the aforementioned spectrum which is supposed to have some regularity as a function in rescaled time u. This parallels nonparametric curve estimation, from an asymptotic point of view of getting more and more information about the spectrum as a curve on a finer and finer grid in rescaled time, as $T \to \infty$. The Dahlhaus class of these "locally stationary" time series includes VARMA models that have coefficients or parameters that change "slowly" over time.

In this paper, we develop the SLEX model for multivariate non-stationary time series. Our approach is complementary to the Dahlhaus model but we place an important consideration

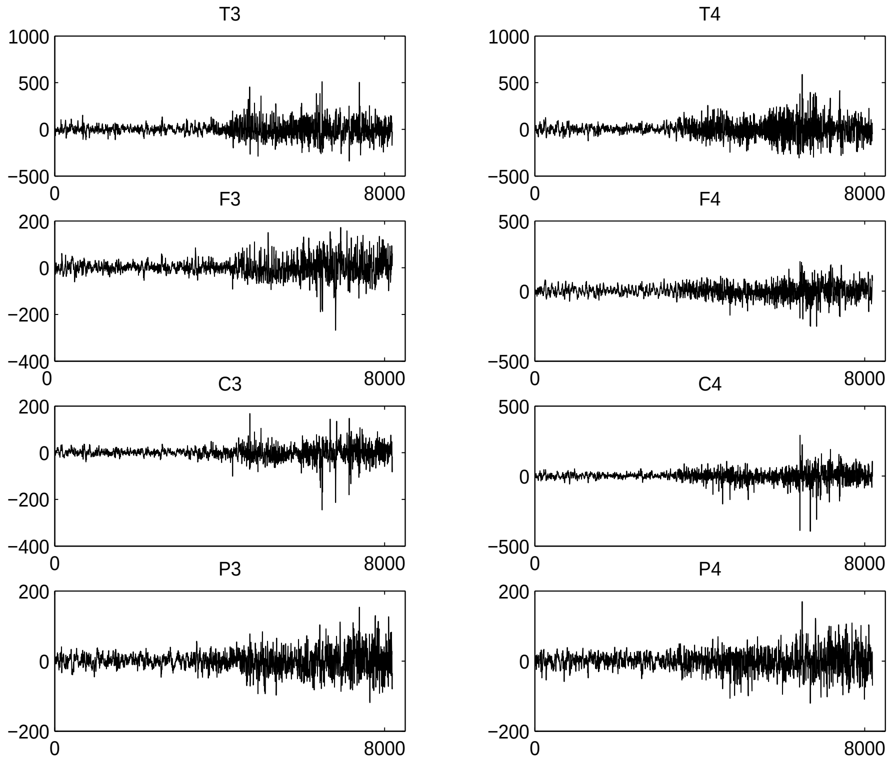

Figure 1: Electroencephalogram recorded during an epileptic seizure. $T = 1024$. Sampling rate is 100 Hertz. Plots on the left (right) column are EEGs recorded on the left (right) side of the brain. Only 8 EEG plots are shown although the analysis was conducted on the complete dataset that consists of $p = 18$ channels. The EEG plots on the *left column* are recordings from the *left side of the brain*: $T3$ (left temporal lobe); $F3$ (left parietal); $C3$ (left central) and $P3$ (left parietal). The plots on the *right column* are recordings from the *right side of the brain*: $T4$ (right temporal); $F4$ (right frontal); $C4$ (right central) and $P4$ (right parietal lobe).

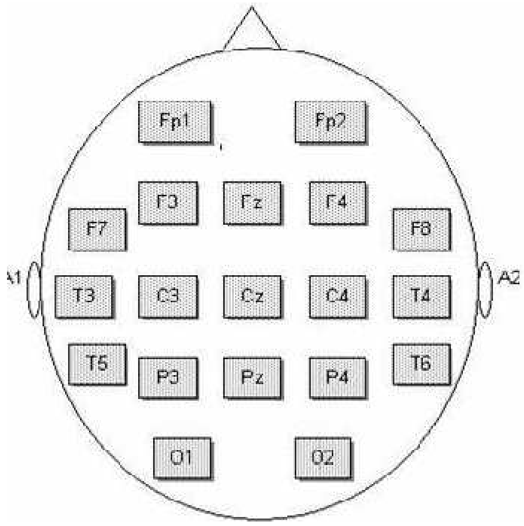

Figure 2: Recording locations of the electroencephalograms.

in developing a model that allows for computationally efficient methods of estimation. This is necessary in dealing with massive data sets which we frequently encounter in practice. The special feature of the SLEX model is that it uses a unique SLEX basis as the building blocks for the stochastic spectral representation. Basically we replace the complex exponentials in (1.2) by the SLEX basis functions. As such they are localized along a certain set of blocks in time which make up a partitioning of the rescaled time interval $[0, 1]$ and hence provide a segmentation of the time series into quasi-stationary blocks. The driving zero mean increment process is orthogonal over frequencies within each of these time blocks. The time (i.e. block)- dependent coefficients of the resulting SLEX spectral representation replace the time-varying transfer function. The spectrum on each time block and each frequency is again the (multivariate) square of these coefficients. For each fixed frequency, this SLEX spectrum is piecewise constant in time.

An essential distinction between the Dahlhaus and SLEX models is in the representation, i.e., Fourier versus SLEX basis functions. While the Dahlhaus model is more general but does not benefit from localized basis functions, the SLEX model is more specific in imposing an explicit segmentation structure in time. Consequently, an important feature of the SLEX model is that, for each T, it provides an explicit partitioning of the time-frequency plane. This enables the construction of estimators which are directly adapted to the model. But, furthermore, the SLEX model even allows for inference that is based on resampling (as explored in Ombao et al, 2000) because its discrete time-frequency representation or synthesis equation makes it easy to generate non-stationary time series following this model.

The rest of the paper is organized as follows. In Section 2, we recall the essential ideas of the SLEX library. Our new multivariate non-stationary SLEX model is presented in Section 3. In Section 4, we will develop the theoretical motivation for our model selection criterion. The inherent problem of high dimensionality and the SLEX PCA are discussed in Section 5. In Section 6, we present the complete algorithm for selecting a model (i.e. finding the segmentation of the time series) from data. In addition, we also discuss our procedure for estimating the unknown spectra and cross-spectra. Finally, in Section 7, we illustrate the

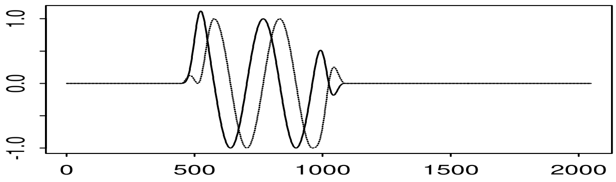

Figure 3: A SLEX waveform with support on block $S(2,1)$ oscillating at frequency index 2.

proposed SLEX analysis by application to an 18-channel epileptic seizure EEG data set.

2 The SLEX transform

2.1 The SLEX vectors

Fourier analysis is the classical approach to investigating the spectral properties of stationary time series. The Fourier basis functions, however, cannot adequately represent processes whose spectral properties evolve with time. The historical approach to studying non-stationary time series is to use tapered or windowed Fourier exponentials (Daubechies, 1992). A windowed Fourier function is of the form $\phi_F(u) = \Psi(u)\exp(i2\pi\omega u)$ where Ψ is a taper with compact support and $\omega \in (-1/2, 1/2]$. While the windowed Fourier functions are localized in time, they are in general non-orthogonal. In fact, the Balian-Low theorem says that does not exist any smooth window such that the windowed Fourier basis functions are simultaneously orthogonal and localized (Wickerhauser, 1994). Orthogonality is an important property. It makes models mathematically elegant and it facilitates the theoretical development of a model. Moreover, orthogonal transforms preserve the energy of the time series and allow the use of the best basis algorithm (BBA) of Coifman and Wickerhauser (1992) which is computationally efficient and hence facilitates the analysis of massive data sets.

The SLEX basis functions are simultaneously orthogonal and localized in time and frequency. They evade the Balian-Low obstruction because they are constructed by applying a projection operator, rather than a single taper, on the Fourier functions. The details on the construction of a projection operator are given in Wickerhauser (1994). It is shown in Ombao (1999) that the application of a projection operator is identical to applying two, rather than just a single, special smooth and compactly supported windows to the Fourier basis functions. A SLEX basis function $\phi_\omega(u)$ has the form

$$\phi_\omega(u) \quad = \quad \Psi_+(u)\exp(i2\pi\omega u) + \Psi_-(u)\exp(-i2\pi\omega u)$$

where $\omega \in (-1/2, 1/2]$ and $u \in [-\eta, 1+\eta]$ where $0 < \eta < 0.5$. The windows Ψ_+ and Ψ_- are constructed by using rising cut-off functions. These windows come in pairs, i.e., once Ψ_+ is specified, Ψ_- is determined. An example of the window pairs is plotted in Figure 4. The SLEX basis functions can be seen as generalized windowed Fourier functions. Note that if we set $\Psi_-(u) = 0$ for all u then the SLEX function in Equation (2.1) is reduced to the windowed Fourier function. Including the second window Ψ_- results in waveforms that are simultaneously localized and orthogonal.

SLEX basis functions generalize directly to orthogonal SLEX basis vectors for representing discrete 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

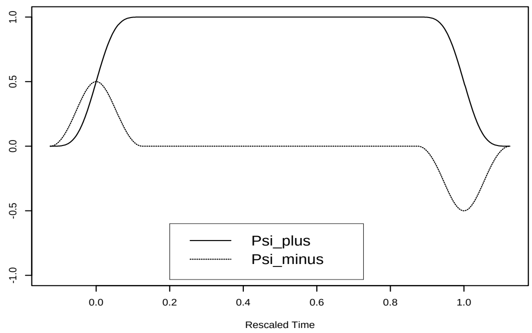

Figure 4: The bells Ψ_+ and Ψ_-.

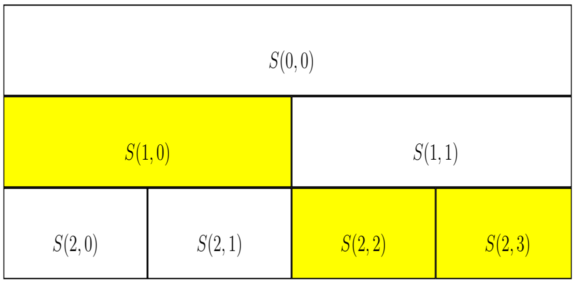

Figure 5: A SLEX library with level $J = 2$. The shaded blocks represent one basis from the SLEX library.

integer less than or equal to its argument. A SLEX basis vector with time support on S is

$$\phi_{S,\omega_k,t} = \Psi_+(\frac{t-\alpha_0}{|S|}) \exp[i2\pi\omega_k(t-\alpha_0)] + \Psi_-(\frac{t-\alpha_0}{|S|}) \exp[-i2\pi\omega_k(t-\alpha_0)]$$

where frequency $\omega_k = k/|S|$; $k = -|S|/2+1, \ldots, |S|/2$. The SLEX basis vector is defined on the set $t \in \{\alpha_0 - \epsilon, \ldots, \alpha_1 - 1 + \epsilon\}$ where ϵ is a small overlap between consecutive blocks.

2.2 The SLEX library

The SLEX library is a collection of bases; each basis consists of the SLEX waveforms which are localized, thus they are able to capture the local spectral features of the time series. The rich collection of bases in the SLEX library allows a flexible variety of possible representations for empirical time series. To illustrate these ideas, we construct a SLEX library in Figure 5 with level $J = 2$. There are 7 dyadic blocks in this library. These are: $S(0,0)$ which covers the entire time series; $S(1,0)$ and $S(1,1)$ which are the two half blocks and $S(2,0)$, $S(2,1)$, $S(2,2)$ and $S(2,3)$ which are the four quarter blocks. Note that in general, for each resolution level $j = 0, 1, \ldots, J$, there 2^j blocks each having length $T/2^j$. We will adopt the notation $S(j, b)$ to denote the block b on level j where $b = 0, 1, \ldots, 2^j - 1$.

The SLEX library above contains exactly 5 SLEX bases. Each basis is defined as consisting of blocks such that (i.) the union of blocks cover the support of the entire time series and (ii.) none of the blocks is covered by another in the same basis. We enumerate all the 5 bases in this library. First basis contains the stationary block $S(0, 0)$. The second basis contains blocks $S(1, 0), S(1, 1)$. The third basis contains $S(2, 0), S(2, 1), S(2, 2), S(2, 3)$. The fourth contains $S(2, 0), S(2, 1), S(1, 1)$. The fifth contains $S(1, 0), S(2, 2), S(2, 3)$ - which corresponds to the shaded blocks in Figure 5. we will We point out that the bases are allowed to have multi-resolution scales, i.e., a basis can have time blocks with different lengths. This is ideal for processes whose regimes of stationarity also vary with time.

In choosing the finest time scale (or deepest level) of the transform J, the statistician will need some advice from the scientific expert. For example, neurologists can give some guidance regarding an appropriate time resolution of EEGs. In general, we also adhere to the following statistical principles. The finest blocks should be sufficiently small in order to reduce bias due to non-stationarity. In other words, the blocks should be small enough so that we can be confident that the time series is stationary in these blocks. However, we should not make them smaller than what is necessary in order to control the variance of the spectral estimator.

2.3 Computing the SLEX transform

The SLEX transform is a collection of coefficients corresponding to the SLEX waveforms in the SLEX library. We will demonstrate that the SLEX coefficients can be computed using the fast Fourier transform (FFT). Let $\mathbf{V}_t$ be a p-variate time series with length $T = 2^L$ for some positive integer L. Let X_t be one component of $\mathbf{V}_t$. The SLEX coefficients (corresponding to X_t) on block $S(j, b)$ are defined as:

$$d_{j,b}^{(X)}(\omega_k) = (M_j)^{-1/2} \sum_t X_t \overline{\phi_{j,b,\omega_k,t}} \ , \tag{2.1}$$

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

3 The Multivariate SLEX Model

Every model in the SLEX family uses a unique SLEX basis as the building blocks of the stochastic spectral representation of the underlying time series to be modelled as such. In this section we will introduce our multivariate SLEX model, that is describe how to define one multivariate model for the non-stationary time series given a known SLEX basis. Let $\mathrm{V}_{t,T} = \{(X_{t,T}^1, \dots, X_{t,T}^p)', t = 0, \dots, T-1\}$ be a mean zero p-variate non-stationary time series. Let B_T be a basis in the library of SLEX basis that will be used in the spectral representation, and define $\bigcup_i S_i$ to be the blocks that make up the time supports of the SLEX basis vectors included in this basis B_T.

Definition 3.1. *For a given T and basis $\bigcup_i S_i \sim B_T$, the SLEX model of a zero mean p-dimensional non-stationary random process is:*

$$V_{t,T} \quad = \sum_{i:\bigcup S_i \sim B_T} \frac{1}{\sqrt{M_i}} \sum_{k_i=-M_i/2+1}^{M_i/2} \theta_{S_i,k_i,T} \ \overline{\phi_{S_i,\omega_{k_i},t}} \ \ \mathbf{z}_{S_i,k_i} \ , \tag{3.1}$$

where (i) $\theta_{S_i,k_i,T}$ is a sequence (in T) of fixed complex-valued $p \times p$ transfer function matrices defined on block S_i and frequency ω_{k_i} satisfying $\theta_{S_i,-k_i,T} = \overline{\theta_{S_i,k_i,T}}$ for all $k_i = 1, \dots, M_i/2-1$; (ii.) $\mathbf{z}_{S_i,k_i} = [z_{S_i,k_i}^1, \dots, z_{S_i,k_i}^p]'$ is a p-variate (complex-valued) random vector that satisfies, *for all $k_i, k_j \geq 0$, the following: $\mathbf{z}_{S_i,-k_i} = \overline{\mathbf{z}_{S_i,k_i}}$ and $Cov[\, \mathbf{z}_{S_i,k_i},\, \mathbf{z}_{S_j,k_j}] = \mathbf{I}_{p\times p}$ when $k_i/M_i = k_j/M_j$ and $\mathbf{0}_{p\times p}$ otherwise.*

Remarks.

(i.) The $p \times p$ SLEX spectral density defined at rescaled time $u = t/T \iff [uT] \in S_i$ and frequency ω_{k_i} is $\mathbf{f}_T(u, \omega_k) = |\theta_{S_i,k_i,T}|^2 \mathbf{1}_{S_i}([uT])$. The SLEX spectral density matrix varies with time but is constant within each block S_i.

(ii.) The random vector $\mathbf{z}_{S_i,k_i}$ has p components. When the frequency $\omega_{k_i} \neq 0, 1/2$, each component is complex valued whose real and imaginary parts are uncorrelated with zero mean and variance equal to 1/2. When $\omega_{k_i} = 0, 1/2$ then each component of is real-valued with zero mean and unit variance. Moreover, within each frequency k_i/M_i, the components are uncorrelated.

(iii.) For all i, j such that $k_i/|S_i| = k_j/|S_j|$, then Cov $(\mathbf{z}_{S_i,k_i}, \mathbf{z}_{S_j,k_j}) = \mathbf{I}_p$ ($p \times p$ identity matrix); and it is the zero matrix otherwise.

(iv.) We denote the diagonal elements are the spectral density matrix to be $\mathbf{f}_{11}, \ldots, \mathbf{f}_{pp}$; the off-diagonal elements are the cross-spectra $\mathbf{f}_{\ell m}$ that satisfy $\mathbf{f}_{m\ell} = \overline{\mathbf{f}_{\ell m}}$; the phase spectrum is $\arg(\mathbf{f}_{\ell m})$; the coherency is $R_{\ell m} = |\mathbf{f}_{\ell m}|/\sqrt{\mathbf{f}_{\ell\ell}\mathbf{f}_{mm}}$.

(v.) Analogous to the Cramér representation of stationary process, the basic components of the SLEX model are the SLEX basis vectors $\phi_{S,k,t}$, the time-evolutionary transfer function $\theta_{S,k,T}$ which is discrete in both time and frequency and the complex-valued random variables $z_{S,k}$.

(vi.) Our approach to estimating $\mathbf{f}_T(u, \omega_k)$ is parallel to the classical approach. We first compute the vector of SLEX coefficients at time block S_i and frequency ω_{k_i} which is denoted as $\mathbf{d}_{S_i}(\omega_{k_i}) = [d^1_{S_i}(\omega_{k_i}), \ldots, d^p_{S_i}(\omega_{k_i})]'$. Then we form the SLEX periodogram matrix $\mathbf{I}_{S_i}(\omega_{k_i}) = \mathbf{d}_{S_i}(\omega_{k_i})\mathbf{d}^*_{S_i}(\omega_{k_i})$, and finally smooth the auto and cross periodograms over frequency to obtain mean-square consistent estimators. We discuss this further in Section 6.

(vii.) One appealing feature of the SLEX model, which to our knowledge is not shared by other spectral models, is that one can easily simulate time series realizations using a simple synthesis equation. To generate a SLEX time series with a given SLEX basis B_T and transfer function θ, one only needs to generate the random variables $\mathbf{z}$'s. The special synthesis feature also allows for resampling-based inference which was explored in Ombao et al. (2000).

(viii.) Although in this paper we do not elaborate on asymptotic theory (i.e., letting T tend to ∞), we like to mention that this is possible by a straightforward analogy to our univariate approach in Ombao et al (2002). We briefly summarize this idea. When we let $T \to \infty$, we embed the SLEX model into a sequence of models with a spectral density matrix that approaches, in the limit, a smoothly varying function $\mathbf{f}(u, \omega)$ which does not depend on T, and which we call the "evolutionary SLEX spectral density matrix". Every element of the evolutionary SLEX spectral matrix, will be, for a fixed frequency ω, defined via a Haar wavelet representation as a function of time, which when truncated to some fixed scale in the

Haar wavelet representation, J_T, gives the SLEX spectral density matrix for the time series of length T. Haar wavelets have been chosen simply to formally represent indicator functions over time to give the link between the limiting smoothly in time varying spectrum $\mathbf{f}(u, \omega)$ and $\mathbf{f}_T(u, \omega_k)$, for each finite T.

4 The Model Selection Criterion

We now develop the theoretical foundation for the penalized log energy criterion which is our criterion for selecting the best SLEX model for the empirical time series. In this section, we will first state our criterion and then show that it is derived as the Kullback-Leibler distance between two multivariate random processes. We defer the details of the algorithm to Section 6.

4.1 The complexity penalized criterion

The complexity penalized log-energy cost at block $S(j, b)$ is defined to be

$$\text{Cost}(j, b) = \sum_{k=-M_j/2+1}^{M_j/2} \log \det \widetilde{\mathbf{I}}_{j,b}(\omega_k) + \ \beta_{j,b}(p)\sqrt{M_j}\ , \tag{4.1}$$

where $\widetilde{\mathbf{I}}_{j,b}$ is the smoothed periodogram matrix and $\beta_{j,b}(p)$ is the data-driven complexity penalty for block $S_{j,b}$. Let $h_{j,b}$ be the bandwidth used in smoothing the SLEX periodogram matrix. A simple version of the complexity parameter that we will use is $\beta_{j,b}(p) = p\,\beta_{j,b}$ where $\beta_{j,b}$ takes the form

$$\beta_{j,b} = \beta_{j,b}(h_{j,b}) = \log_{10}(\mathrm{e})/\sqrt{h_{j,b}}\ \sqrt{2\ \log M_j} \tag{4.2}$$

The cost for a particular model $\bigcup_i S_i \ \sim \ B_T$ is the *sum* of the costs for all blocks in B_T. In Figure 5, the shaded blocks represent one orthonormal basis (or one model). The cost for this particular basis is the sum of the costs at each of the shaded blocks. In the actual search for the best model, we will use the best basis algorithm (BBA) of Coifman and Wickerhauser (1992) which is an algorithm that has been shown to be computationally efficient. The basic idea is that if the cost at a parent block $S(j, b)$ is smaller than sum of the cost in the children blocks $S(j + 1, 2b - 1)$ and $S(j + 1, 2b)$ then we choose the parent block.

Note that our multivariate cost (4.1) is a direct generalization of the cost for the univariate time series X_t^ℓ:

$$\sum_{k=-M_j/2+1}^{M_j/2} \log(\widetilde{I}_{j,b}^{\ell,\ell}(\omega_k)) + \beta_{j,b}\sqrt{M_j}.$$

Note also that in Ombao et al. (2001a), the cost used was the sum of the univariate costs for each of the bivariate components. This simplification would work in case where the components of the multivariate time series are uncorrelated since the spectral matrix would be diagonal. However, in many practical situations, the components of multivariate time series show a high degree of correlation. Thus, any model selection criterion must take into account the cross-relationships between the components. We now motivate the log energy and the complexity penalty parts of our criterion.

4.2 Motivation of the log energy part

Donoho, Mallat and von Sachs (2000) proposed a log energy criterion for analyzing univariate locally stationary processes using the local cosine transform. The goal in Donoho et al. is quite distinct from our own pursuits. Thus, we will borrow the conceptual ideas and adapt them to our goal in this paper which is to analyze non-stationary time series that (i.) are multivariate and (ii.) admit SLEX spectral representations. Suppose now that we have a non-stationary Gaussian process with probability law P_Σ and unknown covariance matrix Σ. Our aim is to approximate Σ by a block-stationary covariance matrix that is the Fourier back transform of the SLEX spectrum (which is piecewise constant in time). This amounts to approximating the given non-stationary Gaussian process by an appropriate Gaussian SLEX process.

Let $f(x)$ be the density of P_Σ, and define the Shannon Entropy

$$H(P_\Sigma) \;=\; -\int f(x)\;\log f(x)\; dx\;.$$

We sketch now that the basis minimizing the log-entropy cost among all bases in the SLEX library of bases gives rise to finding the block-stationary covariance matrix which minimizes the Shannon Entropy. The link between these two minimization problems is given by searching among all covariance matrices of Gaussian probability laws which are *diagonal* in some basis $\mathcal{B}$ in the given library of orthogonal Fourier-like bases, here the SLEX bases. Note that the classical Fourier basis can be seen as diagonalizing the covariance matrix of stationary circular processes, hence the analogue for block-stationary approximations of locally stationary processes.

This minimizing covariance matrix, called Γ in the sequel, yields a Gaussian process which, among all processes with covariance matrix diagonalized by a basis in the library, is a best approximation to the covariance matrix Σ of the underlying Gaussian process. This approximation is actually in the sense of the Kullback-Leibler distance, expressed in terms of the (non-symmetric) discrepancy between the two Gaussian probability measures P_Σ and Q_Γ, with densities f and g, respectively:

$$K(Q_\Gamma|P_\Sigma) \;=\; \int f(x)\;\log[f(x)/g(x)]\; dx\;.$$

In the Appendix Section 9 we show that

$$2\;K(Q_\Gamma|P_\Sigma) = \log\det\Gamma - \log\det\Sigma = 2H(Q_\Gamma) - 2H(P_\Sigma)\;, \tag{4.3}$$

due to the special structure of Γ being an approximation to Σ and being diagonal in $\mathcal{B}$. (In fact, $\Gamma = \mathcal{B}\; Diag(\mathcal{B}'\Sigma\;\mathcal{B})\;\mathcal{B}'$, i.e. conceptually first Σ is rotated into the basis $\mathcal{B}$, then its off-diagonals are set to zero, then the result is back-rotated.) Minimizing the Kullback-Leibler distance $K(Q_\Gamma|P_\Sigma)$ of Q_Γ to P_Σ is hence the same as minimizing $H(Q_\Gamma)$. This follows from $H(P_\Sigma)$ being the entropy of the true unknown probability measure P_Σ and hence a constant with respect to the minimization problem.

Now, equation (9.2) in the Appendix Section 9 shows that, up to constants not depending on Γ,

$$2H(Q_\Gamma) = \sum_k \log(\lambda_k)\;,$$

where the λ_k are the eigenvalues of Γ. Note first that for the univariate case of SLEX, these eigenvalues are approximately equal to the spectrum $f_T(\omega_k)$ on the block $S_{(j,b)}$ of stationarity of the SLEX process. More specifically, still first for the univariate case, it can be shown that, approximately, for the covariance matrix Γ_M of dimension $M \times M$ of one block of length M, say,

$$2\ H(Q_{\Gamma_M}) = \log\det(\Gamma_M) \approx T \int \log f_T(\lambda)\ d\lambda \approx \sum_{k=-M/2+1}^{M/2} \log f_T(\omega_k)\ ,$$

again, up to constants not depending on Γ_M. Now, for our multivariate case, we get

$$2\ H(Q_\Gamma) = \log\det(\Gamma) \approx \sum_k \log\det(\mathbf{f}_T(\omega_k))\ .$$

This has been shown in a similar context by Dahlhaus (2001), Proposition 2.5, which gives rates of this approximation even for the general locally stationary situation. On the empirical level, of course, the theoretical quantities in the spectral matrix $\mathbf{f}_T$ of one block need to be replaced by the estimated quantities $\widetilde{\mathbf{I}}$.

4.3 The complexity penalty term

In Donoho et al (2000), Section 14, the log of the periodograms resulting from the local cosine transform were smoothed by some Haar wavelet soft thresholding. We note that this approach is equivalent to adding the value of the threshold to the (non-thresholded) coarsest scaling coefficient of this Haar expansion. Unlike Donoho et al. we do not apply wavelet thresholding as an estimation procedure. Rather, we use kernel smoothing. However, we can modify the concept in developing our own complexity penalty term. For simplicity, we will sketch our arguments only for the univariate situation. In particular, we will justify by some inequality why we can use the same principle here for logs of kernel smoothed SLEX periodograms. The multivariate penalty has actually the same form, apart from its slight modification which takes the dimensionality into account.

The penalized criterion for block (j, b) takes the form

$$\sum_{k\in\Omega_j} \log I_{j,b}(\omega_k)\ +\ \beta_j\ \sqrt{M_j}\ .$$

The term $\sqrt{M_j}$ can be motivated by projecting $\log I_{j,b}(\omega_k)$ onto a Haar wavelet vector as a function of frequencies $\omega_k, k \in \Omega_j$, supported on block (j, b), i.e. $\mathbf{\Psi}_j = 1/\sqrt{M_j}$. Let $\mathbf{I}_j = (I_{j,b}(\omega_k))_{k\in\Omega_j}$. Then,

$$\sum_{k\in\Omega_j} \log I_{j,b}(\omega_k) =< \mathbf{\Psi}_j, \log\mathbf{I}_j >\ \sqrt{M_j}\ .$$

Let $\tilde{\mathbf{I}}_j$ be the corresponding vector where each component is smoothed by Haar wavelet soft-thresholding using a threshold β_j. By straightforward application of the wavelet representation of $\tilde{\mathbf{I}}_j$ it is easy to show that

$$\sum_{k\in\Omega_j} \log \tilde{I}_{j,b}(\omega_k) =< \mathbf{\Psi}_j, \log\tilde{\mathbf{I}}_j >\ \sqrt{M_j} = (< \mathbf{\Psi}, \log\mathbf{I}_j > +\beta_j)\ \sqrt{M_j}\ ,$$

which ends this motivation.

Observe now that

$$\sum_{k\in\Omega_j} \log \widetilde{f}_{j,b}(\omega_k)\ \geq\ \sum_{k\in\Omega_j} \widehat{\log(I)}_{j,b}(\omega_k)\ ,$$

where $\widehat{\log(I)}_{j,b}(\omega_k)$ denotes, as intermediate reference, the kernel smoother over log-periodograms which would correspond to our actual kernel smoother of periodograms here. Note that the inequality holds due to the concavity of the logarithm.

Second observe that due to linearity and the normalization of the kernel weights to sum up to unity,

$$\sum_{k\in\Omega_j} \widehat{\log(I)}_{j,b}(\omega_k) \;\approx\; \sum_{k\in\Omega_j} \log I_{j,b}(\omega_k) \ ,$$

where $\approx$ stands for equality if we neglect the different behavior of the periodograms at frequencies zero and π.

Both observations together show that

$$\sum_{k\in\Omega_j} \log \widetilde{f}_{j,b}(\omega_k) \;\geq\; \sum_{k\in\Omega_j} \log I_{j,b}(\omega_k)$$

which means that, with high probability, it is larger than the population analogue, the " true cost" based on the unknown spectrum. Hence it is conservative to use the same penalty as in the original derivation (cf. the explanations in the previous Section 4.2) in the sense that it splits a block when it it non-stationary. This is because the cost is never smaller than the Donoho et al cost which asymptotically controls under-splitting.

4.4 The data-driven complexity penalty parameter

We propose a data-driven penalty parameter estimation which is based on the theoretical concepts developed for universal thresholds in wavelet thresholding as well as in complexity-penalisation problems [Antoniadis and Fan (2001) and Donoho (1997)]. Clearly we can picture our BBA search algorithm over block partitions in time as a (univariate) variant of the Dyadic CART algorithm of Donoho (1997). Hence it is justified to interpret our procedure as a Haar wavelet hard thresholding procedure with the special constraint of keeping or killing a complete set of wavelet coefficients which correspond to the scaling coefficients of the very block under consideration. In other words, the constraint is such that a Haar wavelet coefficient in a particular scale j is kept only if the Haar scaling coefficient on the coarser scale $j-1$ above has been kept. This guarantees indeed a Haar wavelet reconstruction which corresponds to a complete block partition of the whole observed interval $1,\ldots,T$ in time.

The universal threshold for keeping or killing wavelet coefficients over a block of length M_j is proportional to the standard deviation of these coefficients, times the well-known factor of $\sqrt{2\log M_j}$ for protecting against large deviations of Gaussian variables. As the variance of Haar wavelet coefficients is actually equal to the one of Haar scaling coefficients, we only need to calculate the latter one.

Following Brillinger (1981), Corollary 5.6.3, we observe that the asymptotic variance of the (univariate) log kernel-smoothed spectral estimate $\log \tilde{f}_{h_{j,b}}(\omega_k)$ at frequency ω_k in the block $S(j,b)$ of length M_j, smoothed by a bandwidth $h_{j,b}$, is equal to $(\log_{10}(e))^2/(M_j\ h_{j,b})$. Summing over M_j frequencies in this block (compare the form of our cost function) and treating all individual spectral estimates as (approximately) independent over Fourier frequencies in this block, the total (asymptotic) variance is M_j times the individual one derived above. Taking square roots, this motivates our choice of the scale and block dependent $\beta_{j,b}$ which is

$$\beta_{j,b} \;=\; \log_{10}(e)\ \sqrt{(h_{j,b})^{-1}\ 2\ \log(M_j)}\ .$$

In our approach, we use one bandwidth $h_{j,b}$ in smoothing all auto-spectra in order to ensure that our spectral matrix estimates are non-negative definite. As a result, the contribution of each component of the multivariate time series to the complexity penalty term is the same. Thus, in our multivariate situation, we simply use the complexity penalty parameter $\beta_{j,b}(p) = p\,\beta_{j,b}$.

5 Model Selection Based on Non-Redundant Spectral Information

In practice, we frequently encounter multivariate time series that exhibit a high degree of multi-collinearity. This suggests the need to extract non-redundant pertinent spectral information, from a massive high dimensional data set, that are to be used in model selection and further analysis. In this paper, we propose to apply principal components analysis on the SLEX (spectral) domain. We decompose the multivariate empirical time series into non-stationary components that have zero coherency. The essential idea is to use the non-redundant time-varying spectral information of the SLEX principal components in selecting the best model (or best segmentation). Once the best segmentation is selected, we continue our analysis by estimating the auto-spectra, cross-spectra and coherence. In addition, we may also investigate the most interesting non-stationary principal components.

Principal components analysis is traditionally used as a tool to reduce a p-variate time series into q principal components. In our approach to model selection, we do not directly use PCA to reduce dimensionality. Our criterion is a weighted average of the log spectra of the p SLEX PC's where the weights are proportional to the eigenvalues. Thus, our criterion suppresses the role of those SLEX PCs that have low eigenvalues in model selection. One can view our criterion as being implicitly based on a reduced dimension because only the SLEX PCs that significantly contribute to the variance play a significant role in model selection. In this section, we will motivate the SLEX PCA for multivariate non-stationary time series and state our criterion for model selection.

5.1 Principal components for stationary time series

Brillinger (1981) motivates frequency domain PCA for stationary time series in the following way. Suppose that we have a p-variate time series $\mathbf{V}_t = [X_1(t), \ldots, X_p(t)]'$, with zero mean and spectral density matrix $f_V(\omega)$, which we now want to approximate by a $q-$variate process $\mathbf{R}(t) = [R_1(t), \ldots, R_q(t)]$ $(q \leq p)$, whose components have zero coherency, defined by

$$\mathbf{R}(t) = \sum_{\ell=-\infty}^{\infty} \mathbf{c}_{t-\ell}' \mathbf{V}(\ell) \tag{5.1}$$

where $\{\mathbf{c}_r\}$ is a $p \times q$ valued filter satisfying $\sum_{r=-\infty}^{\infty} |\mathbf{c}_r| < \infty$. The filter coefficients $\{\mathbf{c}_r\}$ are obtained using an optimality criterion which we now describe. Suppose that we want to be able to reconstruct the $\{\mathbf{X}_t\}$ from the $q-$ variate $\mathbf{R}(t)$ by $\widehat{\mathbf{X}}(t) = \sum_{\ell=-\infty}^{\infty} \mathbf{b}_{t-\ell}\mathbf{R}(\ell)$ where the $p \times q$ filter $\mathbf{b}_r$ that satisfies $\sum_{r=-\infty}^{\infty} |\mathbf{b}_r| < \infty$. We want $\widehat{\mathbf{X}}(t)$ to be such that the mean square approximation error $E[(\mathbf{X}_t - \widehat{\mathbf{X}}_t)^*(\mathbf{X}_t - \widehat{\mathbf{X}}_t)]$ is minimized.

To ease our discussion, we suppose that the eigenvalues of $f_X(\omega)$ are unique and let $v^1(\omega), v^2(\omega), \ldots, v^q(\omega)$ be the eigenvalues arranged in decreasing magnitude. Denote the corresponding eigenvectors to be $V^1(\omega), V^2(\omega), \ldots, V^q(\omega)$. The solution to this problem is to choose

$$\mathbf{c}_\ell = \int_{-1/2}^{1/2} \mathbf{c}(\omega)\exp(i2\pi\ell\omega)d\omega \tag{5.2}$$

where $\mathbf{c}(\omega)$ is the matrix consisting of eigenvectors $V^1(\omega), \ldots V^q(\omega)$. It turns out that the spectrum of the $m-$th principal component $R_m(t)$ at frequency ω is the $m-$th largest eigenvalue $v^m(\omega)$. For an excellent discussion on the applications of frequency domain PCA in stationary time series, we refer the reader to Shumway and Stoffer (2000).

5.2 SLEX principal components for multivariate non-stationary time series

We now propose a time-dependent generalization of the frequency domain principal components analysis. In the non-stationary case, the filter coefficients are allowed to vary over time. Equivalently, the spectra of the SLEX principal components, which are the eigenvalues of the time-varying spectral density matrix, also evolve with time.

Our goal is to first decompose the multivariate non-stationary time series into the SLEX principal components which are non-stationary components that have zero coherency. The time-varying filter and PC spectra at block $S(j, b)$ are obtained by performing an eigenvalue-eigenvector decomposition of the estimated spectral density matrix $\widetilde{\mathbf{f}}_{j,b}(\omega_k)$ for each ω_k. The best model (or best segmentation) is obtained by applying the penalized log energy criterion on the SLEX PC. The spectra of the SLEX PCs are simply the eigenvalues of the spectral density matrix. Denote $v^1_{j,b}(\omega_k), \ldots, v^p_{j,b}(\omega_k)$ be the p eigenvalues arranged in decreasing magnitude. If we have a prior knowledge on the reduced dimension $q \leq p$, then the penalized log energy criterion (4.1) at block $S(j, b)$ can be defined in terms of the q most significant SLEX PCs to be

$$\text{Cost}(j, b) = \sum_{k=-M_j/2+1}^{M_j/2} \sum_{d=1}^{q} \log(v^d_{j,b}(\omega_k)) + q\ \beta_{j,b}\ \sqrt{M_j}. \tag{5.3}$$

In practice, however, q is rarely known. Moreover, even in the stationary situation there is no consensus on the best approach to selecting q. Here, we only mention two possible approaches. First, we could apply, at each frequency, some hard thresholding or significance test to the p eigenvalues. This would both allow for an objective choice of the reduced dimension q as well as exclude the $p - q$ non-significant eigenvalues from our criterion (5.3). For a "threshold" value, one may use some high percentile of the asymptotic distribution of the estimated log eigenvalues in block $S(j, b)$. Another approach is to apply a universal threshold coming from wavelet theory for spectral estimation problems. The issue of selecting q will be a topic of our future research.

In this paper, we will adopt a very simple approach which will not require the user to specify q. The basic idea is to weigh each SLEX PC's contribution to the cost function in proportion to its variance contribution. Essentially, SLEX PCs with larger eigenvalues are given more weight and those smaller ones are given smaller weights. The weight $w^d_{j,b}(\omega_k)$ of the SLEX PC with the d-th largest eigenvalue is defined to be

$$w^d_{j,b}(\omega_k) \quad = \quad v^d_{j,b}(\omega_k) / \sum_{c=1}^{p} v^c_{j,b}(\omega_k) \tag{5.4}$$

which is simply the proportion of variance explained by the d-th component at frequency ω_k, in block $S(j,b)$. The cost at block $S(j,b)$ is

$$\mathrm{Cost}(j,b) = \sum_{k=-M_j/2+1}^{M_j/2} \sum_{d=1}^{p} w_{j,b}^{d}(\omega_k)\ \log v_{j,b}^{d}(\omega_k)\ + \beta_{j,b}\ \sqrt{M_j}\ , \tag{5.5}$$

where, as before, $\log(v_{j,b}^{d}(\omega_k))$ is the logarithm of the spectrum of the $d-$th principal component at frequency ω_k in block $S(j,b)$ having applied PCA to the optimally smoothed periodogram matrix. Note that this cost is based on the "weighted" eigenvalues. Hence, we do not need the factor q in the complexity penalty term.

One attractive feature of weighting the eigenvalues is that the "optimal" number q need not be explicitly specified. It implicitly already renders as irrelevant those components that do not contribute much to the variance. From a numerical point of view it also avoids computational problems since the term $w^d \log(v^d)$ is assigned the value "zero" when v^d and w^d are both close to 0, i.e., when the absolute and relative contribution to variance respectively are small.

6 The Algorithm of the Multivariate SLEX Method

The algorithm of the SLEX method which we now present incorporates the theoretical derivations in Section 4 and the practical considerations on extracting non-redundant spectral information discussed in the preceding section. The outline of the algorithm is: Step (1.) Build the family of SLEX models by computing the SLEX transform, up to a pre-specified level J. Then compute the SLEX periodogram matrix, i.e., the SLEX raw periodograms and cross-periodograms. Step (2a.) Smooth the elements of this matrix, within each block $S(j,b)$, across frequency by a kernel smoother using the same bandwidth. This block-specific bandwidth is automatically selected by generalized cross-validation GCV. Step (2b.) Compute the eigenvalues of the optimally smoothed SLEX periodogram matrix. Step (3.) Select the best model using the complexity penalized log energy criterion applied on the spectra of the SLEX PC (or the eigenvalues). Step (4.) Estimate the spectral density matrix, phase and coherence by smoothing periodograms and cross-periodograms at blocks that are defined by the SLEX model selected.

The Algorithm

Step 1. Let $d_{j,b}^{\ell}(\omega_k)$ $(k = -M_j/2+1, \ldots, M_j/2)$ be the SLEX coefficients of time series X^{ℓ} $(\ell = 1, \ldots, p)$ on block $S(j,b)$. Next, we compute the SLEX raw periodograms and cross-periodograms: $I_{j,b}^{\ell,m}(\omega_k) = d_{j,b}^{\ell}(\omega_k)\overline{d_{j,b}^{m}(\omega_k)}$ where $\ell, m = 1, \ldots, p$. Denote the resulting periodogram matrix to be $\mathbf{I}_{j,b}(\omega_k)$.

Step 2a. Smooth the elements of the SLEX periodogram matrix at each block $S(j,b)$ using a kernel smoother with an optimal bandwidth that is selected by a criterion that is based on generalized cross-validation of the gamma deviance. Let $\widehat{I}_{j,b,h}^{\ell,\ell}$ be the smoothed SLEX periodograms in block $S(j,b)$ of X^{ℓ} , $\ell = 1, \ldots, p$, being smoothed using the block-specific bandwidth $h_{j,b}$. We denote the "error degrees of freedom" by $df_{j,b,h} = 1 - \mathrm{trace}(\mathrm{H}_h)/(M_j/2+1)$ where H_h is the smoother matrix. Then $h_{j,b}$ is the minimizer of the sum of the GCV functions for $\{X_t^\ell\}, (\ell = 1, \ldots, p)$:

$$GCV_{j,b}(h) = \sum_{\ell=1}^{p} GCV_{j,b}^{\ell}(h) \text{ where} \tag{6.1}$$

$$GCV_{j,b}^{\ell}(h) = \frac{2}{(df_{j,b,h})^2} \sum_{k=0}^{M_j/2} q_k \left\{ \frac{I_{j,b}^{\ell,\ell}(\omega_k)}{\widehat{I}_{j,b,h}^{\ell,\ell}(\omega_k)} - \log \frac{I_{j,b}^{\ell,\ell}(\omega_k)}{\widehat{I}_{j,b,h}^{\ell,\ell}(\omega_k)} - 1 \right\} \tag{6.2}$$

where $q_k = 1/2$ when $k = 0, M_j/2$ and $q_k = 1$ otherwise.

Step 2b. Compute the eigenvalues at each block $S(j,b)$ and frequency ω_k and arrange them in decreasing order $v_{j,b}^1(\omega_k), \ldots, v_{j,b}^p(\omega_k)$. Compute the weights $w_{j,b}^d(\omega_k),\ d = 1, \ldots, p$, that will be assigned to the eigenvalues according to Equation (5.4).

Step 3. Compute the cost for each block $S(j,b)$:

$$\text{Cost}(j,b) = \sum_{k=-M_j/2+1}^{M_j/2} \sum_{d=1}^{p} w_{j,b}^d(\omega_k)\ \log v_{j,b}^d(\omega_k)\ + \beta_{j,b}\ \sqrt{M_j}. \tag{6.3}$$

The cost for a particular model with basis $\bigcup_i S_i \sim B_T$ is the sum of the costs for all blocks in the basis B_T used for this model. The best basis (or equivalently, the best model or best segmentation) in the SLEX library is the one that minimizes the cost. In the actual search, we implement the BBA.

Step 4. Extract the optimally smoothed SLEX periodograms and the real and imaginary parts of the SLEX cross-periodograms (from Step 2) at each of the blocks $S(j,b)$ of the model chosen by the BBA.

We denote the estimate of the time-varying coherency and phase between X^ℓ and X^m to be

$$\widetilde{R}_{j,b}^{\ell,m}(\omega_k) = |\widetilde{I}_{j,b}^{\ell,m}(\omega_k)| / \sqrt{\tilde{I}_{j,b}^{\ell,\ell}(\omega_k) \tilde{I}_{j,b}^{(m,m)}(\omega_k)}\ , \tag{6.4}$$

and

$$\widetilde{\phi}_{j,b}^{\ell,m}(\omega_k) = \arg(\widetilde{I}_{j,b}^{\ell,m}(\omega_k))\ , \tag{6.5}$$

respectively.

Remarks.

(i.) We briefly describe the motivation and justification of our automatic bandwidth selector. The details of the method are reported in Ombao et al. (2001b). The SLEX periodograms behave asymptotically like the tapered Fourier periodograms. Thus, under suitable conditions given in Brillinger (1981), the SLEX periodograms at each block are approximately independently distributed as gamma (chi-squared) random variables with mean and variance involving the unknown spectrum. Thus, we can model the SLEX periodograms using generalized additive models as in Hastie and Tibshirani (1990) with the periodogram as the dependent variable, frequency as the predictor, the identity as the link function and a gamma as the random component. The same distributional properties also hold for the estimated eigenvalues $\widetilde{fd}_{j,b}(d, \omega_k)$, we again refer to Brillinger (1981), Section 9.4. Consequently, we can use a generalized cross validation function that is based on the deviance for gamma distributed random variables as a smoothing parameter selection criterion.

(ii.) Note that we use the same bandwidth in smoothing all components of the periodogram matrix in order to guarantee that the estimate of the spectral density matrix is positive definite (i.e., the coherence estimates all lie in the interval $[0, 1]$).

(iii.) Regarding the choice of the penalty for complexity, this follows exactly the same argument as in our Section 4 using Equation (9.4.18) in Brillinger (1981) for the asymptotic variance of the estimator of the $d-$th spectral eigenvalue. We observe that for a kernel smoother of bandwidth $h_{j,b}$ the asymptotic variance of the log kernel-smoothed estimate $\log \widetilde{f}^{d}_{h_{j,b}}(\omega_k)$ of the spectrum of the $d-$th principal components at frequency ω_k in the block $S(j, b)$ of length M_j is equal to $(\log_{10}(\mathrm{e}))^2/(M_j\ h_{j,b})$.

7 The SLEX Analysis of Multi-channel EEGs

We present an analysis of an 18-channel EEG recorded from a female patient diagnosed by Dr. Malow (neurologist at the University of Michigan) to be suffering from left temporal lobe epilepsy. The EEGs at 8 channels (out of the 18 channels used in the analysis) are shown in Figure 1. Each EEG time series has length $T = 8192$. They were recorded for about 82 seconds and then sampled at the rate of 100 hertz.

The SLEX analysis selected a segmentation with changes that occur at approximately 20, 30, 36, 38, 40, 61, 72, 73, 74 and 77 seconds from the starting time. This particular segmentation is reasonable according to the attending neurologist. Moreover, we have confidence that our method yields a reasonable segmentation of the EEGs because results from some preliminary numerical experiments with simulated data have 90% correct segmentation. In this paper, we choose to focus our discussion on the analysis of the EEG data set rather than the simulation studies.

The SLEX time-varying spectra are given in Figure 6. The shown change points are consistent with the expectations of the attending neurologist. The physical manifestations of seizure became evident at around 40 seconds from the start of recording. However, prior to this, changes in the electrical activity of the brain were already beginning to take place. The SLEX analysis of the EEGs was able to capture the change around 40 seconds as well as the other important change-points in the time series. Epileptic seizure is not a static process but one that continuously evolves over time. One can also note that a common trend exists in all the EEG channels that is prior to seizure, power was concentrated at the lower frequencies. During seizure, however, power spread to all frequencies. This is consistent with the known clinical fact that large populations of neurons with diverse firing frequencies are recruited during the seizure. As seizure drew to an end, the concentration of power was slowly restored to lower frequencies.

The coherence plots in Figure 7 are helpful to understanding connectivity between brain areas, i.e., how the neuronal activity in one brain area may influence another. One feature of these EEGs is that connectivities between brain areas are not static but rather evolve during an epileptic seizure. This patient has left temporal epilepsy ($T3$) so we focus our coherence analysis between the $T3$ channel and the other channels. It is fascinating to note that the coherence between $T3$ and those at the left side of the brain, namely, left parietal

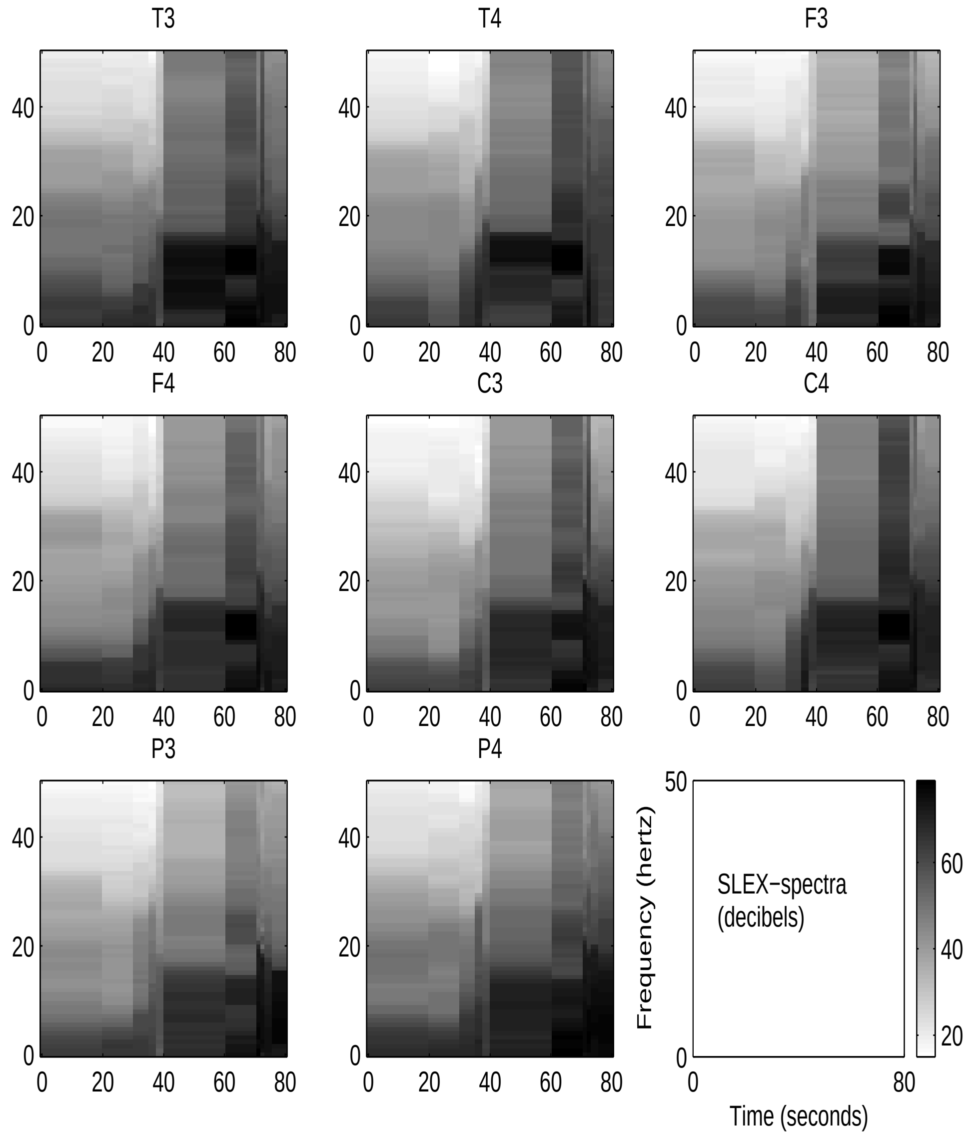

Figure 6: SLEX auto-spectral estimates at the 8 clinically most interesting channels (out of the 18 channels used in the analysis).

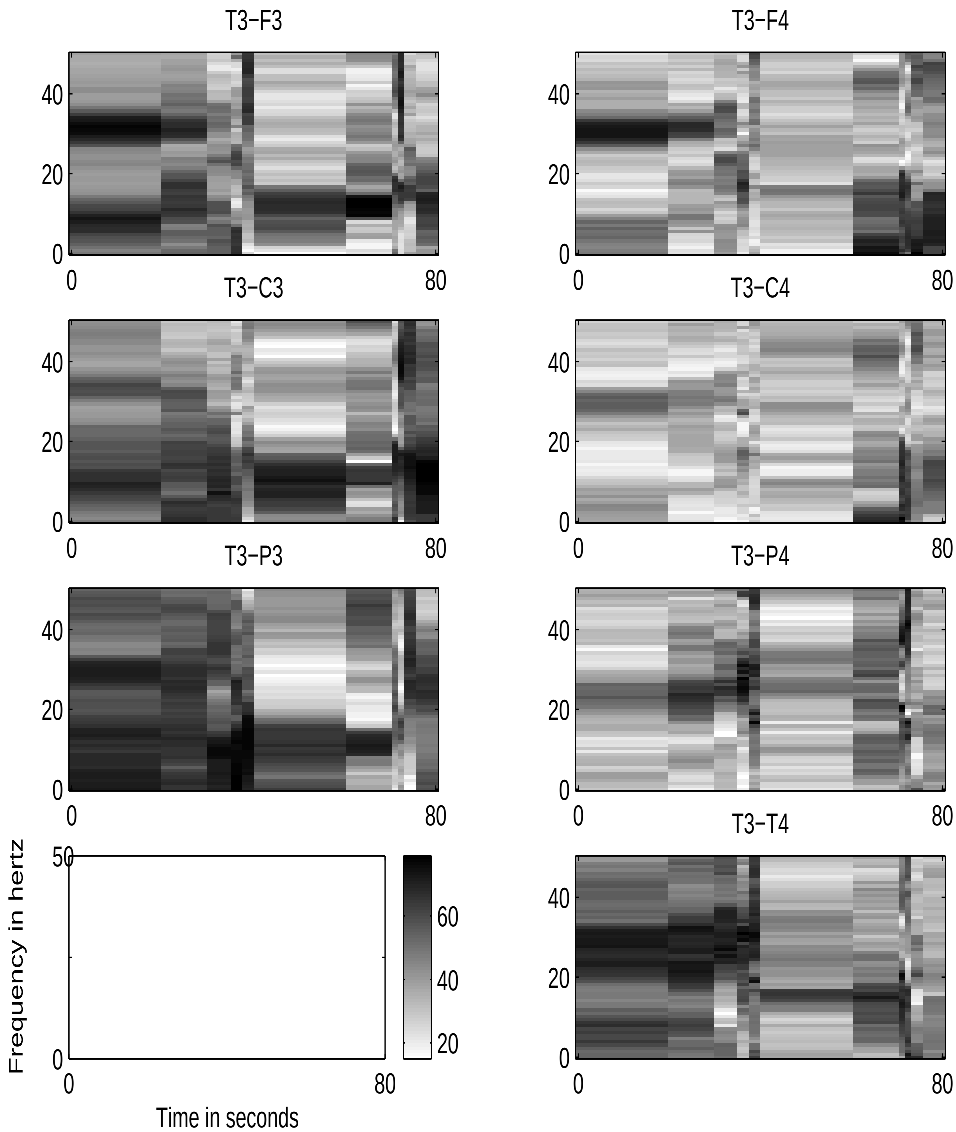

Figure 7: Left: SLEX coherence estimates between $T3$ and channels on the left side of the brain namely $F3, C3, P3$. Right: SLEX coherence estimates between $T3$ and the channels on the right side of the brain namely $F4, C4, P4, T4$.

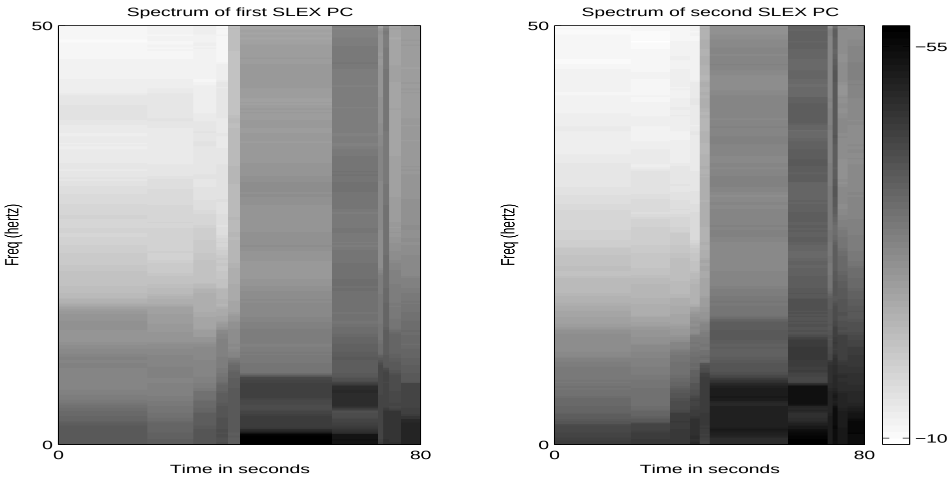

Figure 8: Time-varying spectra of the first and second SLEX principal components.

($P3$), left frontal ($F3$) and left central ($C3$) are very similar to each other. Moreover, they are very different from their counterparts on the right side ($P4$, $F4$ and $C4$ respectively). This suggests connectivities between brain areas that are on the *same* side of the brain of this subject evolve over time in very similar manner during an epileptic seizure.

We also note that prior to the start of the seizure ($t < 40$ seconds), $T3$ was strongly coherent with the left side channels at the alpha band (8-12 hertz) and the beta band (between 25-30 hertz). This was observed between $T3$ and its right hand counterpart the right temporal lobe ($T4$). During an epileptic seizure, there is strong coherence between the left side channels at the alpha band and at extremely high frequencies over 30 hertz.

The primary information that is conveyed in the spectral plots is that the distribution of power over frequency indeed evolves during the seizure process. We see that power at the lower frequencies is increased and that power is spread to middle and higher frequencies during seizure. It is very interesting to note that these important features of the multi-channel EEGs are captured by the first and second SLEX principal components. We report that the first and second SLEX PC account for approximately 70% of the variance in the EEGs. The time-varying spectra of the first SLEX PC (in Figure 8, left side) account primarily for the increase in power in the lower frequencies after the onset of seizure. The second SLEX PC (in Figure˜effig:pcspec, right side), on the other hand accounts for the spread of power from the delta band (1-4 hertz) to the alpha (8-12 hertz) band.

We also obtained the absolute values of the eigenvectors of the first and second SLEX PC at the delta (1-4 hertz), alpha (8-12 hertz) and beta (25-30 hertz) frequency bands. The absolute values that correspond to the first and second SLEX PC are given in figures 9 and 10. These can be interpreted as the time-varying weights at the EEG channels. In this paper, we report only the weights at the 8 most interesting channels. We note that, for the first SLEX PC, most of the weights are concentrated on the $T3$ (left temporal lobe) and $T4$ (right temporal lobe) channels. This is consistent with the fact that this patient is diagnosed as having left temporal lobe epilepsy. Thus, intense activity is expected at the temporal lobes. We observe that, at the beta frequency band, the weights at the $T3$ and $T4$ channels become even larger during an epileptic seizure. Neurologists believe that many neurons firing at diverse frequencies are recruited during the seizure process. This intense activity is being

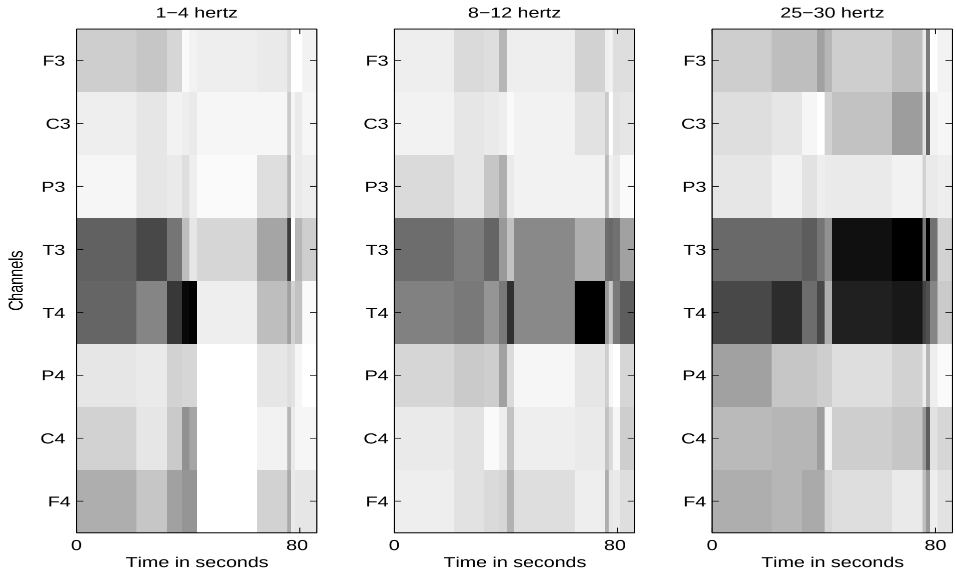

Figure 9: Time-varying weights of the **first** SLEX PC at the delta frequency band (1-4 hertz), alpha frequency band (8-12 hertz) and higher beta frequency band (25-30 hertz). Darker shades represent larger weights.

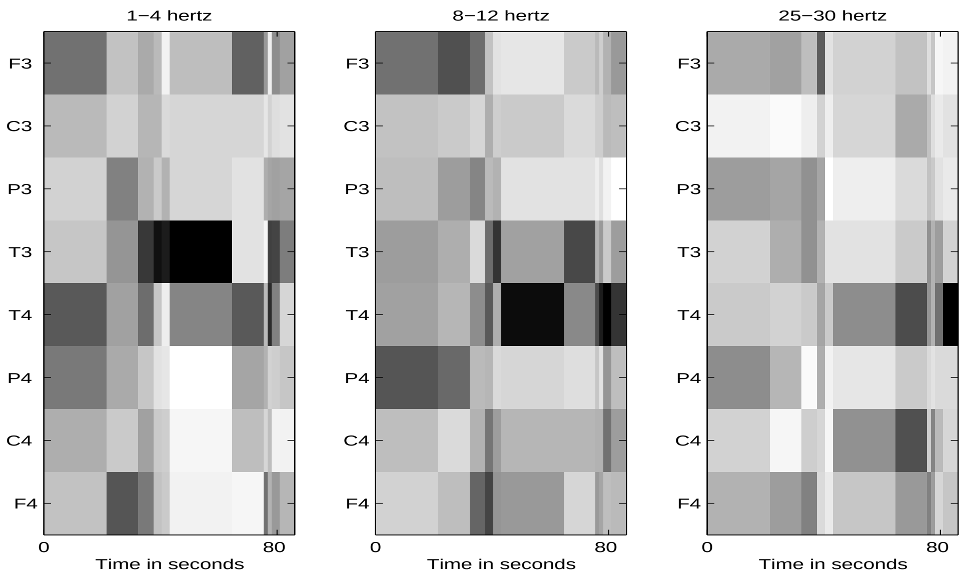

Figure 10: Time-varying weights of the **second** SLEX PC at the delta frequency band (1-4 hertz), alpha frequency band (8-12 hertz) and higher beta frequency band (25-30 hertz). Darker shades represent larger weights.

captured by the first and second SLEX PCs. Moreover, it is interesting to note that the SLEX PCs were able to capture the intense activity on the temporal lobes at both the left and right sides of the brain. The weights for the second SLEX PC shift and change between the $T3$ and $T4$ channels, again, indicating that the epileptic seizure is not a static but a dynamic process with possibly a feedback mechanism between the left and right temporal lobes.

8 Conclusion

We proposed to analyze a multivariate non-stationary time series using the SLEX library. We use the SLEX waveforms in representing the time-varying spectral and cross-spectral features of the multivariate non-stationary time series. The SLEX basis that we select to represent the empirical time series is the one which minimizes the Kullback-Leibler distance between the empirical time series and the models from the SLEX library. In practice, analyzing high dimensional multivariate time series can be computationally burdensome. We propose to decompose the multivariate time series into SLEX principal components. The SLEX PCs are non-stationary and are obtained by applying a time-variant filter on the empirical time series. The SLEX PCs are uncorrelated in the frequency domain and thus are able to capture non-redundant time-varying spectral information in the empirical time series. The spectra of the SLEX PCs capture the important features of the multivariate time series that are used in selecting the best model. After the segmentation step, we propose to estimate the time-varying SLEX spectral matrix by the kernel smoothed SLEX periodogram matrix. In this paper, we demonstrated our method by its application to multi-channel epileptic seizure brain waves.

Summarizing, we have combined the development of a new fast frequency-domain PCA based algorithm and a new multivariate model of piecewise stationarity: we use this model to justify our automated search for the segments of piecewise stationarity, in other words our algorithm proves useful in particular for the model class of SLEX processes that we propose in this paper. However, we like to have our particular model of nearly piecewise stationarity understood in the sense of giving a good approximative description of data which deviate from second-order stationarity in a broader sense. In particular, in our work Ombao et al (2002) on the univariate SLEX model, we have shown asymptotic equivalence between the Dahlhaus model of local stationarity and the SLEX model. Similar considerations are possible for the respective multivariate models. Once the best adapted segmentation is chosen, one can develop asymptotic theory of consistently estimating the model quantities (in as much as it is done in the aforementioned univariate treatment of Ombao et al, 2002). We decided not to include this here, in order to avoid overloading this methodological paper with such theoretical asymptotic developments. However, we like to emphasize that being able to derive consistent estimators is one main motivation for working with a rigorous multivariate model of non-stationarity.

9 Appendix: Derivation of equation (4.3)

Here we give the details on the proof of

$$2\ K(Q_\Gamma|P_\Sigma)\ = \log\det\Gamma - \log\det\Sigma\ = 2H(Q_\Gamma) - 2H(P_\Sigma)\ . \tag{9.1}$$

First we show that for a Gaussian time series $X_1, \ldots, X_T$ with probability law $P_\Sigma =$

$N(0,\Sigma)$,

$$H(P_\Sigma) = 1/2\ \log\det(\Sigma) + T/2\ \log(2\pi) + 1/2\ , \tag{9.2}$$

which proofs the second equality of equation (4.3):

$$H(P_\Sigma) = -\int f(x)\ \log(f(x))\ dx\ ,$$

where $f(x)$ denotes the density of the Gaussian law $P_\Sigma = N(0,\Sigma)$. Plugging this into the above integral and using the substitution $y = \Gamma^{-1/2}\ x$ immediately leads to the result.

For the first equality, we note that with $\Gamma = D_{\mathcal{B}}\Sigma = \mathcal{B}\ Diag(B'\Sigma B)\ B'$, and with

$$K(Q_\Gamma|P_\Sigma) = \int f(x)\ (\log(f(x)) - \log(g(x)))\ dx = E_f[\log(f(x)) - \log(g(x))]\ ,$$

we get with

$$2\ (\log(f(x)) - \log(g(x))) = \log\det\Gamma + x'\Gamma^{-1}x - \log\det\Sigma - x'\Sigma^{-1}x$$

and with $E_f(xx') = \Sigma$ that

$$2\ E_f[x'\Gamma^{-1}x - x'\Sigma^{-1}x] = (tr(\Gamma^{-1}E_f(xx')) - tr(\Sigma^{-1}E_f(xx'))\ \ = tr[\Gamma^{-1}\ \Sigma] - T\ .$$

Now,

$$tr[\Gamma^{-1}\ \Sigma] = T\ ,$$

since $tr[\Gamma^{-1}\ \Sigma] = tr[Diag(B'\Sigma B)^{-1}\ B\Sigma B] = T$
(as $B^{-1} = B'$, and since all diagonal elements in the resulting matrix are one).

Putting this together gives

$$2\ K(Q_\Gamma|P_\Sigma)\ \ =\ \ 2\ E_f[\log(f(x)) - \log(g(x))] = \log\det\Gamma - \log\det\Sigma\ ,$$

which ends the proof.

References

Antoniadis, A. and Fan, J. (2001). Regularization of Wavelet Approximations (with discussion). *Journal of the American Statistical Association,* **96**, 939-967.

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

Cardoso, J.F. (1998). Multidimensional independent component analysis. *Proc. ICASSP* 1998, Seattle.

Coifman, R. and Wickerhauser, M. (1992). Entropy based algorithms for best basis selection. *IEEE Transactions on Information Theory,* **32**, 712–718.

Dahlhaus, R. (2001). A likelihood approximation for locally stationary processes. *Annals of Statistics,* **28**, 1762–1794.

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

DONOHO, D. (1997). CART and Best-Ortho-Basis: A Connection. *Annals of Statistics,* **5**, 1870-1911.

DONOHO, D., MALLAT, S. AND VON SACHS, R. (2000). Estimating covariances of locally stationary processes: rates of convergence of best basis methods. *Technical Report 517*, Department of Statistics, Stanford University.

FORNI, M., HALLIN, M., LIPPI, M. AND REICHLIN. L. (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics* **82**, 540–554.

HASTIE, T. AND TIBSHIRANI, R. (1990). *Generalized Additive Models.* London: Chapman and Hall.

OMBAO, H. (1999). Theoretical Aspects of the SLEX. *Technical Report, Department of Statistics, University of Pittsburgh.*

OMBAO, H., VON SACHS, R. AND GUO, W. (2000). Estimation and inference for time-varying spectra of locally stationary SLEX processes. *Proceedings of the 2nd International Symposium on Frontiers of Time Series Modeling, Nara, Japan. December 14-17, 2000.*

OMBAO, H., RAZ, J., VON SACHS, R. AND MALOW, B. (2001a). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.* **96**, 543–560.

OMBAO, H., RAZ, J., STRAWDERMAN, R. AND VON SACHS, R. (2001b). A Simple GCV Method of Span Selection for Periodogram Smoothing. *Biometrika*, **88**, 1186-1192.

OMBAO, H., RAZ, J., VON SACHS, R. AND GUO, W. (2002). The SLEX Model of a Non-Stationary Random Process. *Annals of the Institute of Statistical Mathematics*, **54**, 171-200.

SHUMWAY, R.H. AND STOFFER, D.S. (2000). *Time Series Analysis and Its Applications.* New York: Springer.

STOCK, J.H. AND WATSON, M.W. (1998). Diffusion Indexes. *NBER Working Paper W6702*, National Bureau of Economic Research.

STOCK, J.H. AND WATSON, M.W. (2002). Forecasting principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**, 1167-1179.

WEST, M., PRADO, R. AND KRYSTAL, A. (1999). Evaluation and Comparison of EEG Traces: Latent Structure in Non-Stationary Time Series, *Journal of the American Statistical Association*, **94**, 1083-1094.

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