I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0141

SMOOTHING SPLINE ANOVA FOR TIME-DEPENDENT SPECTRAL ANALYSIS

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

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

Smoothing Spline ANOVA for Time-Dependent Spectral Analysis

Wensheng Guo, Ming Dai, Hernando C. Ombao and Rainer von Sachs *

September 21, 2001

Abstract

In this paper, a locally stationary process is proposed using a Smooth Localized Complex Exponential (SLEX) basis, whose spectrum is assumed to be smooth in both time and frequency. A smoothing Spline ANOVA (SS-ANOVA) is used to estimate and make inference on the time-varying log-spectrum. This approach allows the time and frequency domains to be modeled in an unified approach and jointly estimated. Because the SLEX basis is orthogonal and localized in both time and frequency, our method has good finite sample performance. It also allows for deriving desirable asymptotic properties. Inference procedures such as confidence intervals and hypothesis tests proposed for the SS-ANOVA can be adopted for the time-varying spectrum. Because of the smoothness assumption of the underlying spectrum, once we have the estimates on a time-frequency grid, we can calculate the estimate at any given time and frequency. This leads to a high computational efficiency as for large data sets we only need to estimate the initial raw periodograms at a much coarser grid. We present simulation results and apply our method to an EEG data recorded during an epileptic seizure.

Key Words: *Spectral Estimation, Smoothing Spline, SLEX, Locally Stationary Process, Tensor Product*

1 Introduction

Spectral analysis has been an important tool for time series analysis with wide range of applications. Traditional methods are based on the assumption of second-order stationarity. However, this assumption seldom holds in real applications and is only approximately valid for series of very short duration. Recent developments relaxed this assumption to a locally stationary setting (Dahlhaus, 1997; Adak, 1998; Ombao *et al.*, 2001, 2002) in which the spectrum is assumed to be changing slowly over time, and the time series can be approximated by a piecewise stationary time series. The limitations of these methods are: first, their finite sample estimates are not smooth in time because their time varying

*Wensheng Guo (wguo@cceb.upenn.edu) is Assistant Professor, Ming Dai is Post Doctoral fellow, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104. Hernando C. Ombao is Assistant Professor, Departments of Statistics and Psychiatry, University of Pittsburgh, Pittsburgh, PA 15260-0001, U.S.A. Rainer von Sachs is Professor, Institut de Statistique, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium. This research is supported by NIH R01 grants NIMH62298 and CA84438 as well as by the contract 'Projet d'Actions de Recherche Concertées' No. 98/03-217 of the Belgian government.

spectrum is only asymptotically tied to a smooth slowly varying function in time; second, except for the parametric approach of Dahlhaus (1997), estimation in the frequency domain is conditional on the segmentation in the time domain, and therefore prevent the joint estimation in both domains in the nonparametric approaches; third, nonparametric inference is difficult.

In this paper, we propose a smoothing spline tensor product model for a time varying log-spectrum of a locally stationary process. The underlying transfer function and hence the spectrum is assumed to be smooth in time and frequency. This method allows simultaneous smoothing in both domains, and is computationally efficient. Inference procedures developed in the context of smoothing spline ANOVA (SS-ANOVA) (Gu and Wahba, 1993b; Guo, 2000) can be adopted for this model, offering the possibility for confidence regions and for tests (on stationarity, among others).

Our specific motivation comes from the interest in studying how seizures are generated in epileptic patients. We are interested in learning how the spectrum changes over time, which will provide us information to eventually predict the onset of a seizure. Figure 1.1 is an electroencephalograms (EEG) data recorded during an epileptic seizure. It can be seen that the power builds up until the seizure erupts, and the energy then gradually dies out. The assumption of stationarity does not hold in such applications, and a locally stationary model is more reasonable. Another challenge in modeling EEG data is the extreme computational demand, as tremendous amount of data are collected over time. Some initial data processing is usually needed to reduce the problem to a manageable scale. This also motivates us to model the spectrum of the underlying process as a continuous smooth function in both time and frequency. Then its interpretation remains the same across different subsamples in the time-frequency plane, whereas an estimate that is conditional on a chosen time segmentation has interpretation only on this chosen time-frequency grid. We will return to this application in Section 6.

Some of the recent developments in time-dependent spectral analysis can be classified as 1) local Fourier basis approaches (Dahlhaus, 1997; Neumann and von Sachs, 1997), and 2) orthogonal basis approaches (Ombao *et al.*, 2001, 2002), to name but a few. In the local Fourier basis approach, Dahlhaus (1997) extended the Cramér representation to include a time varying transfer function, and used a parametric approach to estimate its spectral function. Neumann and von Sachs (1997) used a tensor product wavelet model to estimate the spectrum of a locally stationary process. Their method is applied on very localized Fourier periodograms to avoid a preliminary segmentation of the time series.

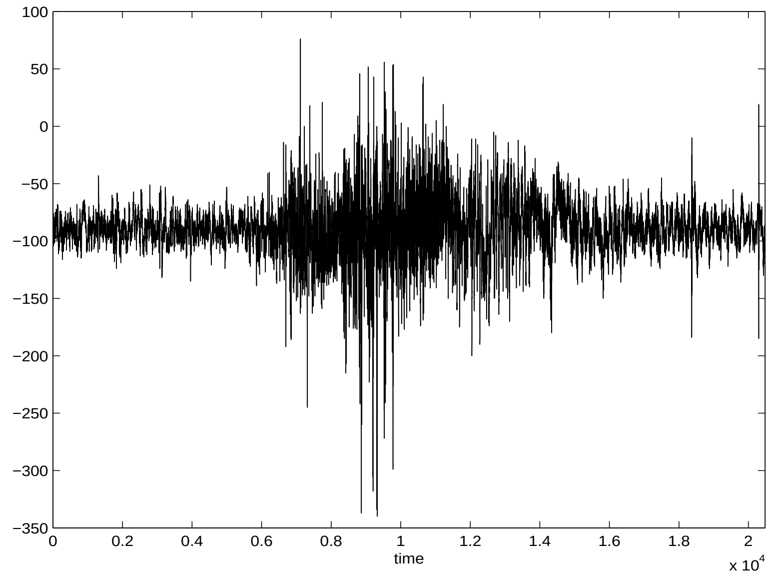

Figure 1.1: EEG data collected during an epileptic seizure. The sampling rate is 100 Hertz. The total number of time points T = 20480

Although this is conceptually very appealing, their final estimates suffer from a high variability (even for comparatively large sample sizes) and hence the lack of allowing for reliable inference.

In the orthogonal basis approach of Ombao *et al.* (2001, 2002), they proposed a discrete time SLEX model in which the Smooth Localized Complex Exponential (SLEX) basis was used to represent the time varying spectrum. A Best Basis Algorithm (BBA) was used to select the optimal segmentation. Each final segment is assumed to be stationary and the traditional frequency domain analysis can be done within a segment. SLEX bases being orthogonal in time and frequency allowed to establish asymptotic properties. Along this line of development, in this paper we propose a new SLEX process whose spectrum is smooth in both time and frequency and use smoothing spline ANOVA (SS-ANOVA) to estimate its log-spectrum. Our model is distinguished from Ombao *et al.* (2002) (termed discrete SLEX model) in the following aspects: 1) our spectrum is defined for any time-frequency, while the estimated spectrum of the discrete SLEX model only has interpretation at the final time-frequency grid given by the BBA; 2) we make the smoothness assumption explicitly and therefore our estimate is smooth even in finite samples, while the estimate in the discrete SLEX model is obtained at each stationary block in the selected time-frequency grid and hence is not smooth; 3) our approach models the log-spectrum instead of the spectrum, which guarantees positivity and produces stable estimates,

4) our approach allows joint nonparametric estimation in time and frequency, while in the existing SLEX method the estimate in the frequency domain is conditional on the segmentation in the time domain; and 5) existing inference procedures such as Bayesian confidence intervals (Wahba, 1983) and generalized likelihood ratio test (Guo, 2000) can be adopted for our approach, while inference in the discrete SLEX model is difficult and needs to resort to computationally intensive methods such as the bootstrap (Ombao *et al.*, 2000).

Our method can be viewed as an extension of Wahba (1980) to a locally stationary process. The key idea in Wahba (1980) is that the periodograms at different Fourier frequencies of a stationary time series are asymptotically independent, and the asymptotic distribution of a periodogram is the corresponding spectrum times a chi-square distribution. Therefore the log-periodograms can be modeled as a signal-plus-noise model, and the distribution of the errors is approximately Gaussian. Then smoothing splines can be used to estimate the log-spectrum. The direct extension of this approach to a locally stationary process based on ordinary Fourier basis functions is difficult as the Fourier basis is not local in time. We therefore need to first define a locally stationary time series model in terms of the SLEX basis that is orthogonal and localized in both time and frequency. This provides a way to explicitly calculate the local SLEX periodograms which are further modeled by a smoothing spline ANOVA model (Gu and Wahba, 1993a).

The rest of the paper is structured as follows: in Section 2, we define our models; in Section 3, we discuss the estimation of the log-spectrum. The inference and model selection are to be found in Section 4. Some simulation results are given in Section 5. An application to the EEG data is used as an illustration in Section 6. We conclude this paper with some remarks in Section 7. Proofs are deferred to the Appendix.

2 The Model

In this section, we introduce our locally stationary time series model and the tensor product model for the log-spectrum.

2.1 Locally Stationary Process Model

The following definition is a modified version of Dahlhaus (1997) model where the original sequence of functions $A^0_{t,T}$ is replaced by a smooth two dimensional function $A^D(\omega, t/T)$ in frequency and time:

Definition 2.1 *A zero-mean stochastic process $\{X_t^D, t = 1, \cdots, T\}$ is called locally stationary if it has the following representation:*

$$X_t^D = \int_0^1 A^D(\omega, t/T)\exp(i2\pi\omega t)dZ(\omega), \tag{2.1}$$

where

1. *$Z(\omega)$ is a zero-mean orthogonal-increment process on [0, 1] with $Z(\omega) = \overline{Z(1-\omega)}$, $\overline{Z}$ is the conjugate of Z, cum$\{dZ(\omega_1, \cdots, dZ(\omega_k)\} = \Delta(\sum_{j=1}^{k}\omega_k)\Lambda_k(\omega_1, \cdots, \omega_{k-1})d\omega_1 \cdots d\omega_k$, where cum$\{\cdots\}$ denotes the cumulant of k-th order; $\Lambda_1 = 0, \Lambda_2(\omega) = 1, |\Lambda_k(\omega_1, \cdots, \omega_{k-1})| \leq C_k$, C_k is a constant and $\Delta(\omega) = \sum_{j=-\infty}^{\infty}\delta(\omega + j)$ is the period 2π extension of the Dirac delta function.*
2. *The transfer function $A^D(\omega, u)$ is continuous in $(\omega, u) \in [0, 1] \times [0, 1]$, $A^D(\omega, u) = \overline{A^D(1-\omega, u)}$. It is also a periodic function w.r.t. ω with period 1.*
3. *The time-dependent spectral function is defined as $f^D(\omega, u) = |A^D(\omega, u)|^2$ for $(\omega, u) \in [0, 1]\times[0, 1]$.*

This representation does not provide us a way to directly calculate time-varying periodograms because the Fourier basis are not localized in time. We therefore introduce the following locally stationary time series model using SLEX vectors.

First we partition the time interval into M blocks not necessarily of equal sizes. Let $1 = \alpha_1 < \alpha_2 \cdots < \alpha_M < \alpha_{M+1} = T$ be the partition points, $B_j = \{\alpha_j, \alpha_j + 1, \cdots, \alpha_{j+1} - 1\}$ be the jth block and $S_j = \alpha_{j+1} - \alpha_j$ be the size for the block B_j. Let $\{\epsilon_j, j = 1, \cdots, M+1\}$ be integers such that $(\alpha_j - \epsilon_j, \alpha_j + \epsilon_j), j = 1, \cdots, M+1$ are disjoint. For each j, the SLEX basis for the block B_j are defined as the following:

$$\phi_j(\omega, t) = b_+^j(t)\exp(i2\pi\omega t) + b_-^j(t, \omega)\exp(-i2\pi\omega t), \tag{2.2}$$

where

$$\begin{aligned} b_+^j(t) &= |r\{(t-\alpha_j)/\epsilon_j\}|^2|r\{(\alpha_{j+1} - t)/\epsilon_{j+1}\}|^2, \\ b_-^j(t, \omega) &= \overline{r}\{(t-\alpha_j)/\epsilon_j\}r\{(\alpha_j - t)/\epsilon_j\}\exp(4\pi i\alpha_j\omega) \\ &\quad -\overline{r}\{(t-\alpha_{j+1})/\epsilon_{j+1}\}r\{(\alpha_{j+1} - t)/\epsilon_{j+1}\}\exp(4\pi i\alpha_{j+1}\omega). \end{aligned}$$

and where $r(t)$ is a rising cutoff function. For more details about the construction of the SLEX vectors and rising cutoff function, see Wickerhauser (1994). In our simulations, we use

$$r(t) = \begin{cases} 0 & \text{if } t < -1, \\ \sin[\pi/4\{1 + \sin(\pi t/2)\}] & \text{if } -1 \le t \le 1, \\ 1 & \text{if } t > 1. \end{cases}$$

Both $b_+^j(t)$ and $b_-^j(t, \omega)$ have supports for $t \in [\alpha_j - \epsilon_j, \alpha_{j+1} + \epsilon_{j+1}]$, which means that the supports of $\phi_j(\omega, t)$ and $\phi_{j+1}(\omega, t)$ have an overlap of length $2\epsilon_{j+1}$. Although these supports are overlapped, the SLEX vectors are orthogonal over time.

The Balian-Low theorem states that there does not exist a smooth taper such that the tapered Fourier vectors are simultaneously orthogonal and localized in time and frequency. The SLEX vectors evade the obstruction, since they are obtained by applying a projection operation on the Fourier vectors, which is equivalent to applying two smooth windows on the Fourier vectors and its conjugate. However, the window $b_-^j(t, \omega)$ is not a regular taper, since in general it depends on both time and frequency.

For a given SLEX basis, we can then define a SLEX locally stationary process:

Definition 2.2 *A zero-mean stochastic process $\{X_t^S, t = 1, \cdots, T\}$ is called SLEX locally stationary if it has the following representation:*

$$X_t^S = \sum_{j=1}^{M} \int_0^1 A^S(\omega, t/T)\phi_j(\omega, t)dZ(\omega), \tag{2.3}$$

where

1. *the definition of $Z(\omega)$ is the same as defined in Definition 2.1.*

2. *The transfer function $A^S(\omega, u)$ is continuous in $(\omega, u) \in [0, 1] \times [0, 1]$, $A^S(\omega, u) = \overline{A^S(1 - \omega, u)}$. It is also a periodic function w.r.t. ω with period 1.*

3. *The time-dependent spectral function is defined as $f^S(\omega, u) = |A^S(\omega, u)|^2$ for $(\omega, u) \in [0, 1] \times [0, 1]$.*

Assumption 2.1 *The transfer function $A^S(\omega, u)$ has up to kth ($k \ge 1$) order continuous partial derivatives w.r.t. ω and u.*

Remarks

1. The Definition 2.2 explicitly defines a projection of a locally stationary process onto the SLEX basis, which is simultaneously orthogonal and localized in time and frequency. This provides a way to calculate the local periodograms (see Section 3 for details).

2. In our model, the transfer function and spectral function are continuous and smooth in both time and frequency. They have their interpretation on any given time-frequency grid. This is different from the discrete SLEX model (Ombao *et al.*, 2002), where, for finite sample size, the transfer function is piecewise constant in time and only has interpretation on the chosen time-frequency grid.

2.2 Relationship with Dahlhaus Model

In this section, we discuss the relationship between our SLEX model and the modified Dahlhaus model. Theorem 2.1 is an extension of Ombao *et al.* (2002) while Theorem 2.2 gives new insights.

Theorem 2.1 *(1) For a SLEX time series $\{X_t^S, t = 1, \cdots, T\}$ given by Definition 2.2 with transfer function $A^S(\omega, u)$, there exists a Dahlhaus locally stationary process $\{X_t^D, t = 1, \cdots, T\}$ with same transfer function such that*

$$T^{-1} \sum_{t=1}^{T} E|X_t^S - X_t^D|^2 = O\{\max(S_j)/T\} + O(T^{-1}) + O\{\max(\epsilon_j)/\min(S_j)\},$$

as $T \to \infty$, $\max(S_j)/T = o(1)$, $\max(\epsilon_j)/\min(S_j) = o(1)$.

(2) Conversely, given a Dahlhaus process $\{X_t^D, t = 1, \cdots, T\}$ with transfer function $A^D(\omega, u)$, there exists a SLEX time series $\{X_t^S, t = 1, \cdots, T\}$ with same transfer function, such that

$$T^{-1} \sum_{t=1}^{T} E|X_t^S - X_t^D|^2 = O\{\max(S_j)/T\} + O(T^{-1}) + O\{\max(\epsilon_j)/\min(S_j)\},$$

as $T \to \infty$, $\max(S_j)/T = o(1)$, $\max(\epsilon_j)/\min(S_j) = o(1)$.

Theorem 2.2 *Let $(X_t, t = 1, \cdots, T)$ be a locally stationary process having both the Dahlhaus and SLEX representations:*

$$\begin{aligned} X_t &= \int_0^1 A^D(\omega, t/T) \exp(i2\pi\omega t) dZ^D(\omega), \\ &= \sum_{j=1}^{M} \int_0^1 A^S(\omega, t/T) \phi_j(\omega, t) dZ^S(\omega). \end{aligned}$$

If the transfer functions $A^D(\omega, u)$ and $A^S(\omega, u)$ have continuous partial derivatives w.r.t. ω and u, then as $T \to \infty$, $\min(S_j) \to \infty$, $\max(\epsilon_j^2)/\min(S_j) \to 0$ and $\max(S_j^2)/T \to 0$,

$$|f^D(\omega, u) - f^S(\omega, u)| \to 0.$$

Remarks

1. Theorem 2.1 states that if the Dahlhaus process and the SLEX process are generated from the same transfer function, then they are asymptotically equivalent.

2. Theorem 2.2 states that if the time series is generated from a Dahlhaus process and we use a SLEX model to estimate the spectrum, the SLEX spectrum converges to the true (i.e. the Dahlhaus) spectrum.

3. From Theorem 2.2, although we can choose different rising cutoff functions and different partitions of the time series to obtain the SLEX vectors, the spectral functions will always converge to the true spectral function, which is unique (Dahlhaus, 1996).

For the rest of the paper, we will use the model of a SLEX process and use the SS-ANOVA for estimation and inference of its log-spectrum. We suppress the superscript S when there is no confusion.

2.3 Tensor Product Model for Logarithm Spectrum

We next introduce the tensor product model for the log-spectrum. For simplicity, we focus on cubic splines, although our method works for splines of any orders. Similar to Wahba (1980), we use the logarithm of spectrum $\log\{f(\omega, u)\} = g(\omega, u)$, which is periodic w.r.t.ω with period 1, $g(\omega, u) = g(1 - \omega, u)$ and has kth ($k = 2$ for cubic splines) order continuous partial derivatives w.r.t. ω and u.

To deal with the curse of dimensionality, we adopt the tensor product model proposed by Gu and Wahba (1993a). We first need to define the reproducing kernels (RKs) for time and frequency domains, see Aronszajn (1950) for details of the reproducing kernel Hilbert space (RKHS)).

In the frequency domain, since $g_1(\omega) = g(\omega, \cdot)$ is periodic with period 1, it has the following Fourier expansion:

$$g_1(\omega) \sim a_0 + \sum_{j \neq 0} a_j \exp(i2\pi\omega j).$$

So the RKHS W_1 for ω can be decomposed as

$$W_1 = \{1\} \oplus H_1,$$

where the reproducing kernel (RK) R_1 for H_1 is the following:

$$R_1(\omega_1, \omega_2) = -k_4(\omega_1 - \omega_2),$$

where $k_4(x) = B_4(x)/4!, B_4(\cdot)$ is the fourth order Bernoulli polynomial (see Wahba (1990) for details).

In the time domain, $g_2(u) = g(\cdot, u)$ has a continuous second derivative. Its corresponding RKHS W_2 has the following decomposition:

$$W_2 = \{1\} \oplus \{u - 0.5\} \oplus H_2.$$

The RK R_2 for RKHS H_2 is defined as follows:

$$R_2(u_1, u_2) = k_2(u_1)k_2(u_2) - k_4(u_1 - u_2),$$

where $k_2(x) = B_2(x)/2!$ and $B_2(\cdot)$ is the second order Bernoulli polynomial.

The full tensor product RKHS for (ω, u) is given by:

$$\begin{aligned} H &= W_1 \otimes W_2, \\ &= [\{1\} \oplus H_1] \otimes [\{1\} \oplus \{u - 0.5\} \oplus H_2], \\ &= \{1\} \oplus \{u - 0.5\} \oplus H_1 \oplus H_2 \oplus \{H_1 \otimes (u - 0.5)\} \oplus \{H_1 \otimes H_2\}. \end{aligned}$$

The RK $R_3(.,.)$ for $H_3 = H_1 \otimes (u - 0.5)$ is given by

$$R_3\{(\omega_1, u_1), (\omega_2, u_2)\} = R_1(\omega_1, \omega_2)(u_1 - 0.5)(u_2 - 0.5),$$

and RK R_4 for $H_4 = H_1 \otimes H_2$ is given by

$$R_4\{(\omega_1, u_1), (\omega_2, u_2)\} = R_1(\omega_1, \omega_2)R_2(u_1, u_2).$$

Correspondingly, the logarithm of the spectral function has the following ANOVA decomposition:

$$g(\omega, u) = \beta_1 + \beta_2(u - 0.5) + g_1(\omega) + g_2(u) + g_3(\omega, u) + g_4(\omega, u), \tag{2.4}$$

where $\beta_1 + \beta_2(u - 0.5)$ is the linear trend, $g_1(\omega)$ (whose RK is $R_1(.,.)$) is the smooth periodic main effect for frequency, $g_2(u)$ (whose RK is $R_2(.,.)$) is the smooth main effect for time, g_3 (whose RK is $R_3(.,.)$) is smooth in frequency and linear in time, and g_4 (whose RK is $R_4(.,.)$) is the interaction between the two smooth terms in time and frequency. This model includes time varying AR and MA models as special cases, and we refer to a concrete example in Section 5.

For a particular application, a full model (2.4) may not be needed and a more parsimonious model can be obtained by setting some terms to zero. This can be done through sequentially testing one term at a time using the GML ratio test proposed by Guo (2000), which will be reviewed in Section 4.2.

3 Estimation

To obtain the estimate of the underlying spectrum, we use a two-stage approach: first, obtain the local periodograms using the SLEX transformation, and then use a SS-ANOVA model to estimate the log-spectrum which is continuous and smooth in both time and frequency.

3.1 SLEX Periodograms

The SLEX coefficients on the block B_j are defined as:

$$\hat{\theta}_{j,k,T} = \frac{1}{\sqrt{S_j}} \sum_{t=1}^{T} X_t \overline{\phi_j(\omega_k, t)}. \tag{3.1}$$

Further, the SLEX periodogram is defined as

$$\hat{\alpha}_{j,k,T} = |\hat{\theta}_{j,k,T}|^2, \tag{3.2}$$

where $\omega_k = k/S_j, k = 1, \cdots, S_j$.

Because of the local stationarity, the time series is approximately stationary in each small block. The spectrum at the middle point of each block can be estimated by the average of the time-varying spectrum in the block. This is the same as treating the time series in the block as stationary in calculating the periodogram for the middle point.

The initial blocking is only used to obtain the raw periodograms, which is not critical in the final estimate because of the second stage estimation using SS-ANOVA and the smoothness assumption on the underlying spectrum. To calculate the SLEX periodograms, we first divide the time series into adjacent disjoint blocks which can have different lengths. The chosen frequencies in each block can be a subset of the Fourier frequencies. The middle points of the blocks and the chosen frequencies form a time-frequency grid, whose periodograms are calculated using (3.2). The next lemma is similar to Theorem 5.2 and Corollary 5.1 in Ombao *et al.* (2002).

Lemma 3.1 *Under the SLEX model (2.3), the periodograms $\hat{\alpha}_{u,\omega,T}$ given by 3.2 are asymptotically independent and distributed as*

$$\hat{\alpha}_{j,k,T} \sim \begin{cases} f(\omega_k, u_j)\chi_2^2/2 & \text{if } \omega_k \neq 0, 1/2, \\ f(\omega_k, u_j)\chi_1^2 & \text{if } \omega_k = 0, 1/2, \end{cases} \tag{3.3}$$

as $T \to \infty$, $\min(S_j) \to \infty$, $\max(\epsilon_j^2)/\min(S_j) = o(1)$ and $\max(S_j^2)/T = o(1)$, where u_j is the middle point of the block B_j. □

This result is the same as for the classical Fourier transform of a stationary process. So when taking logarithm on both sides of Equation (3.3),

$$\log(\hat{\alpha}_{j,k,T}) \approx \log\{f(\omega_k, u_j)\} + \delta_{jk},$$

where δ_{jk} are asymptotically independent with mean C_{ω_k} and variance $\sigma^2 = \pi^2/6$, where C_{ω_k} is the Euler Mascherni constant, and $-C_{\omega_k} = \gamma = 0.57721\cdots$ for $\omega_k \neq 0, 1/2$ and $C_0 = C_{1/2} = -(\ln 2 + \gamma)/\pi$.

Let $Y_{j,k} = \log(\hat{\alpha}_{j,k,T}) - C_{\omega_k}$. Then

$$Y_{j,k} \approx g(\omega_k, u_j) + \epsilon_{jk}, \tag{3.4}$$

where ϵ_{jk} are asymptotically independent with zero mean and variance $\pi^2/6$.

3.2 Spline Estimation

Let $\gamma = (\omega, u)$ be a frequency-time point, $\mathbf{\Gamma} = (\gamma_i)_{i=1}^n$ be the selected frequency-time grid to calculate the observations $\mathbf{Y} = (Y_i, i = 1, \cdots, n)^T$ given by (3.4). As the estimation is the same for any sub-models by setting some terms of the full model (2.4) to zero, we only describe the estimation for the full model. The estimation is to find $g \in H$ which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i - g(\gamma_i)\}^2 + \sum_{j=1}^{4}\lambda_j||P_j g||^2, \tag{3.5}$$

where $\lambda_j, j = 1, \cdots, 4$ are smoothing parameters and $P_j, j = 1, \cdots, 4$ are projection operators on RKHSs $H_j, j = 1, \cdots, 4$.

Let $\lambda_j = \lambda/\theta_j, j = 1, \cdots, 4$, $Q^\theta = \sum_{j=1}^{4} \theta_j R_j(\mathbf{\Gamma}, \mathbf{\Gamma})$. Gu (1989) and Gu and Wahba (1993a) gave the solution to (3.5). Conditional on the smoothing parameters, the estimates $\hat{g}(\mathbf{\Gamma})$ at these design points $\mathbf{\Gamma}$ are given by

$$\hat{g}(\mathbf{\Gamma}) = W\mathbf{d} + Q^\theta \boldsymbol{c}, \tag{3.6}$$

where $W = (1, u_i - 0.5)_{i=1}^n$. The coefficients $\boldsymbol{c}$ and $\boldsymbol{d} = \{\beta_1, \beta_2\}^T$ are determined by

$$\begin{aligned} (Q^\theta + n\lambda I)\boldsymbol{c} + W\boldsymbol{d} &= \boldsymbol{Y}, \\ W^T\boldsymbol{c} &= 0. \end{aligned} \tag{3.7}$$

The estimate of any given time-frequency point $\gamma' = (\omega', u')$ is given by

$$\hat{g}(\gamma') = W'\boldsymbol{d} + \boldsymbol{c}^T \sum_{j=1}^{4} \theta_j R_j(\mathbf{\Gamma}, \gamma'), \tag{3.8}$$

where $W' = (1, u' - 0.5)$.

The smoothing parameters λ and $\theta_j, j = 1, \cdots, 4$ can be chosen by GCV (generalized cross validation), GML(generalized maximum likelihood) or URE (unbiased risk estimate) to minimize corresponding criteria. For GCV, the criterion is

$$V(\lambda) = [1/n\boldsymbol{Z}^T(Q^\theta + n\lambda I)^{-2}\boldsymbol{Z}]/[1/n\,\mathrm{tr}(Q^\theta + n\lambda I)^{-1}]^2;$$

for GML, it is

$$M(\lambda) = [\boldsymbol{Z}^T(Q^\theta + n\lambda I)^{-1}\boldsymbol{Z}/n]/[\det(Q^\theta + n\lambda I)^{-1}]^{1/(n-2)};$$

for URE, it is

$$U(\lambda) = ||\{I - A(\lambda)\}y||^2 + 2\sigma^2\,\mathrm{tr}\,A(\lambda)/n,$$

where $\boldsymbol{Z} = F_2^T\boldsymbol{Y}$, F_1, F_2, T is the QR decomposition for W: $W = (F_1, F_2)\left(\begin{array}{c} T \\ 0 \end{array}\right)$. In our simulations, GML outperforms the other criteria, and therefore is recommended.

4 Inference and Model Selection

In this section, we discuss the inference and model selection for our model.

4.1 Bayesian Confidence Intervals

Using the Bayesian model (Wahba, 1983), Gu and Wahba (1993b) derived the Bayesian confidence intervals for SS-ANOVA. For any given time-frequency point γ', $\hat{g}(\gamma')$ is the posterior mean of $\mathrm{E}\{g(\gamma')|\boldsymbol{Y}\}$ with $\boldsymbol{d} = \{\beta_1, \beta_2\}^T$ having diffuse priors. The corresponding posterior variance $\mathrm{var}\{g(\gamma')|\boldsymbol{Y}\}$ can be used to construct Bayesian confidence intervals. It was shown that

$$\begin{aligned} \mathrm{var}\{g(\gamma')|\boldsymbol{Y}\} = \; & \sigma^2/(n\lambda)\{v^\theta + W'A(W')^T - 2W'AW^TM^{-1}R^\theta \\ & -(R^\theta)^T(M^{-1} - M^{-1}WAW^TM^{-1})R^\theta\}, \end{aligned} \tag{4.1}$$

where $v^\theta = \sum_{j=1}^{4} \theta_j R_j(\gamma', \gamma')$, $A = (W^TM^{-1}W)^{-1}$, $M = Q^\theta + n\lambda I$ and $R^\theta = \sum_{j=1}^{4} \theta_j R_j(\boldsymbol{\Gamma}, \gamma')$. So the Bayesian confidence interval for $g(\gamma')$ is given by $\hat{g}(\gamma') \pm z_{\alpha/2}\sqrt{\mathrm{var}\{g(\gamma')|\boldsymbol{Y}\}}$, where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

Although the Bayesian confidence intervals were derived as point-wise, Nychka (1988) pointed out that they are in fact curve-wise confidence intervals, which means:

$$1/n \sum_{i=1}^{n} \Pr[g(\gamma_i) \in \hat{g}(\gamma_i) \pm z_{\alpha/2}\sqrt{\mathrm{var}\{g(\gamma_i)|\boldsymbol{Y}\}}] \approx 1 - \alpha.$$

4.2 Testing Hypothesis and Model Selection

The advantage of using SS-ANOVA to model the log-spectrum (2.4) is that it decomposes the time-varying spectrum into meaningful additive terms. For example, if $g_3(\omega, u)$ and $g_4(\omega, u)$ are set to zero, then the log-spectrum is simply an additive model, which means that the time series can be modeled by a stationary time series times a time-varying amplitude (termed modulated stationary time series model, also see the example used in our simulation for example). Under this additive model, we can further test $H_0 : \beta_2 = 0, g_2(u) = 0$, which is equivalent to testing whether the time series is stationary. Guo (2000) showed that in the SS-ANOVA decomposition such as (2.4), testing $g_j(.) = 0$ is equivalent to testing $\theta_j = 0$, and under the null hypothesis the GML ratio follows a mixture of chi-square distributions. The connection can be seen from (3.8). The asymptotic distribution is due to the fact that $\theta_i \geq 0$ and the null hypothesis is on the boundary, which follows from the result of Self and Liang (1987). This result can also be used for model selection by sequentially testing one term at a ti Guo (2000) also proposed using maximum likelihood ratios to simultaneously test a non-parametric term and a parametric term together, which again follows from the result of Self and Liang (1987).

Without going into details, we summarize the direct application of the results of Guo (2000) in the following two lemmas.

Lemma 4.1 *In model (2.4), testing $H_0 : g_j(.) = 0$ (j=1,..., 4) is equivalent to testing $\theta_j = 0$, and under the null hypothesis, the asymptotic distribution of two times the logarithm of the GML ratio of the full model and the nested model is a 50:50 mixture of χ_1^2 and χ_0^2, where χ_v^2 is the central chi-square distribution with v degrees of freedom.*

Lemma 4.2 *Under the additive model ($g_3(\omega, u) = 0$ and $g_4(\omega, u) = 0$), testing whether the time series is stationary is equivalent to testing $H_0 : \beta_2 = 0, \theta_2 = 0$, and under the null hypothesis, the asymptotic distribution of two times the logarithm of the ML ratio of the full model and the nested model is a 50:50 mixture of a χ_1^2 and a χ_2^2.*

5 Simulations

To evaluate the finite sample performance of our method, we conduct a small set of simulations. We generate a time series using the following model $y_t = \exp\{0.5\sin(2\pi t/T)+2\}x_t$, where $x_t = 0.5x_{t-1}+\epsilon_t$ and $\epsilon_t \sim N(0,1)$ for $t = 1,\ldots,T$. The true log-spectrum is

$$g(\omega, u) = 4 + \sin(2\pi u) - \log\{1.25 - \cos(2\pi\omega)\},$$

for $\omega, u \in [0, 1]$. Figure 5.1 shows the true log-spectrum. This corresponds to the model (2.4) with the two interaction terms ($g_3(\omega, u)$, $g_4(\omega, u)$) set to zero.

In our simulations, we use the additive model to estimate the log-spectrum and calculate the Bayesian confidence intervals. For $T = 512, 1024, 2048, 4096$, we each generate 100 replicates and calculate the coverage of the Bayesian confidence intervals and mean square errors, which are summarized in Figure 5.3 and Figure 5.4.

In these simulations, the time series is partitioned into blocks of equal lengths. The time points chosen to calculate the local periodograms are the middle points of these blocks, and the frequency points for all of these blocks are same and the number of the frequency points is chosen such that the total number of time-frequency points is less than 2000 for computational efficiency. For $T = 512$, the time series is partitioned into 16 blocks and the number of the selected frequency points on each block

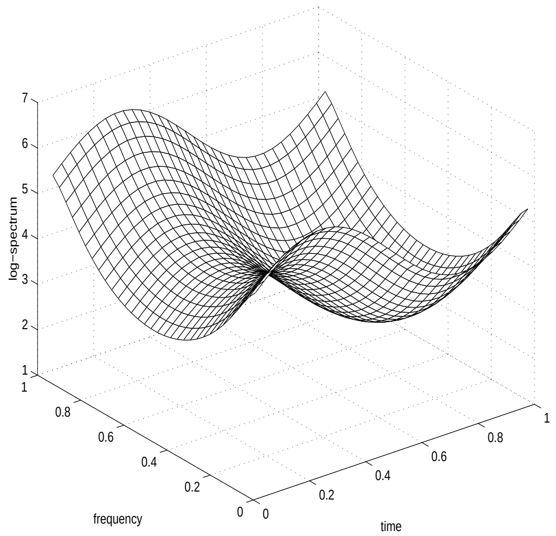

Figure 5.1: True log-spectrum

is 32. For $T = 1024, 2048$, the number of block is 32 and the number of the selected frequency points is also 32. For $T = 4096$, the number of blocks is 64 and the number of selected frequency points is 32.

The estimate of the first simulated time series for each of the four settings is shown in Figure 5.2. Figure 5.3 shows the box plots of the coverage of the 95% confidence intervals and the mean square errors are given in Figure 5.4. From these plots, it can be seen that as the sample size increases, the coverage gets better (closer to 95%) and the mean squared errors become smaller. For $T = 4096$, which is still a small sample size compared with real applications in time-frequency analysis, our method already performs very well in terms of coverage and MSE.

6 Application to EEG Data

In this section, we apply our method to the EEG data shown in Figure 1.1. The length of the time series we used is 20480. The time series is partitioned into 40 disjoint intervals of equal lengths. On each block, there are 512 observations. The frequency points selected to calculate the SLEX periodograms are same on all these blocks and the number is 32. The final estimates and confidence intervals are calculated on 100×80 time-frequency grid points.

We first need to choose a best model. As the time series is clearly nonstationary, we start with the

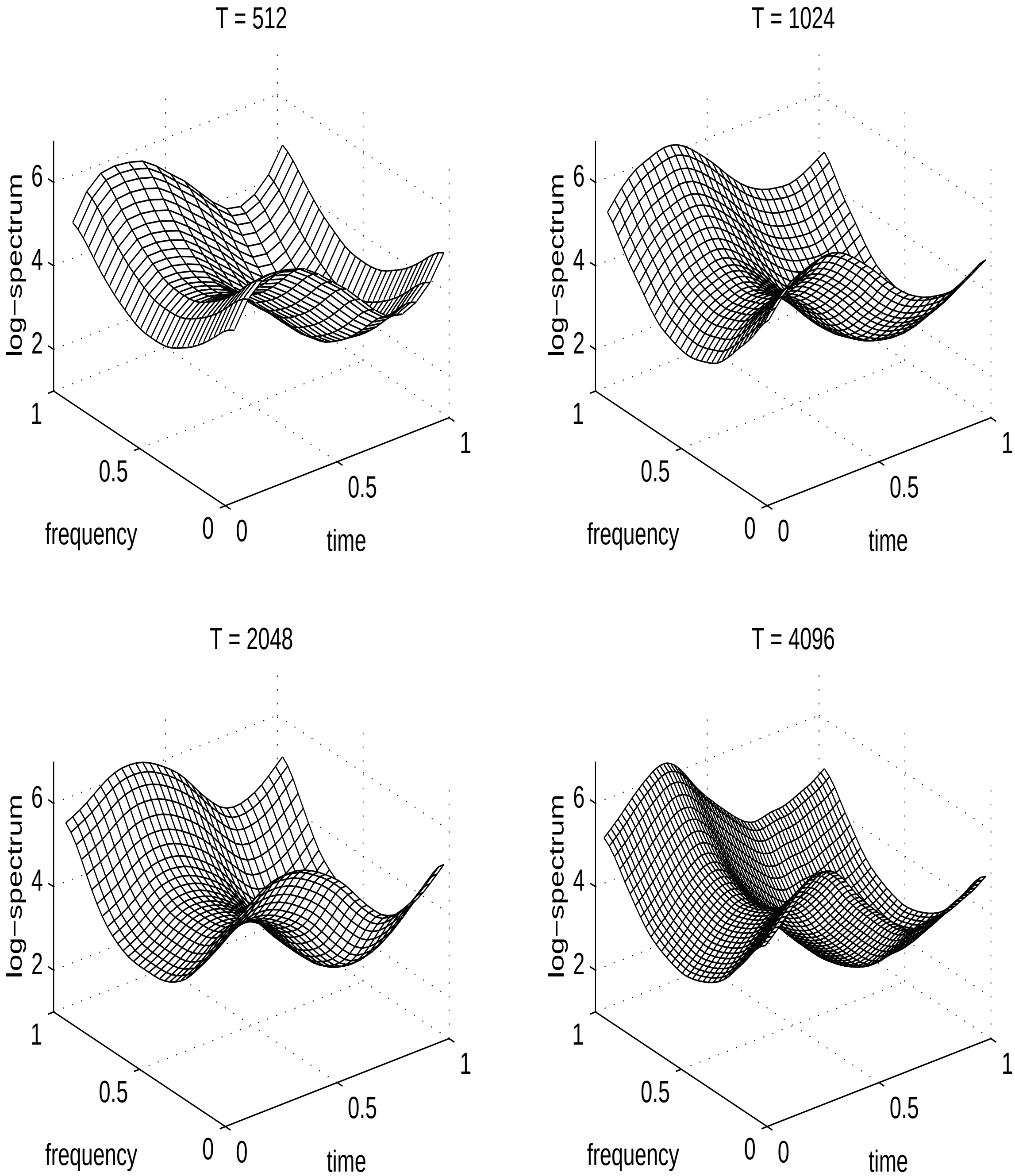

Figure 5.2: The estimated log-spectrum of the first of the 100 replicates in each setting.

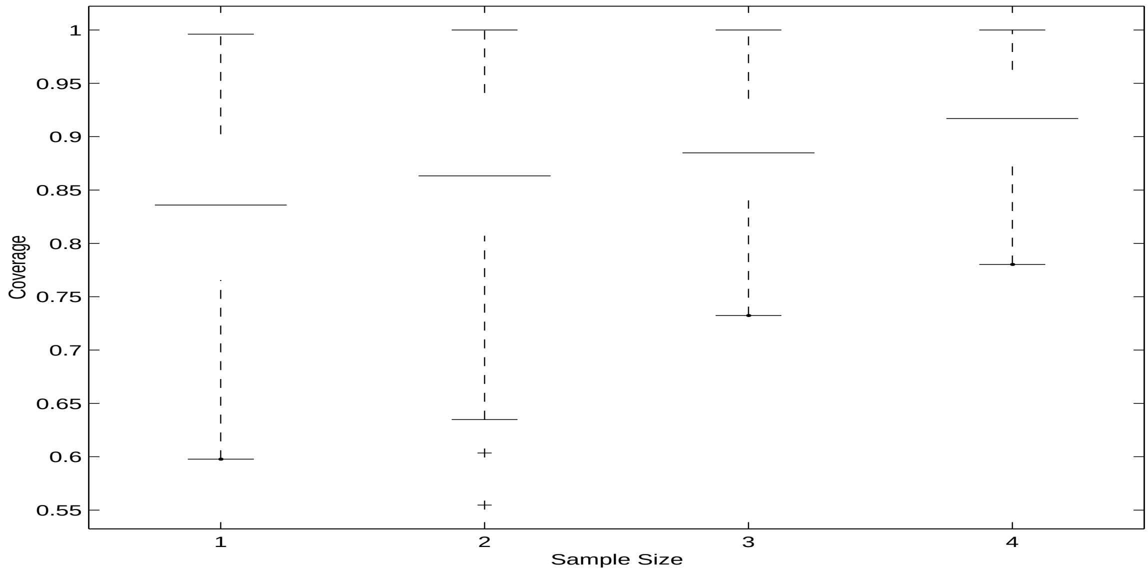

Figure 5.3: Coverage of the 95% Bayesian confidence intervals. From Left to right: $T = 512, 1024, 2048, 4096$.

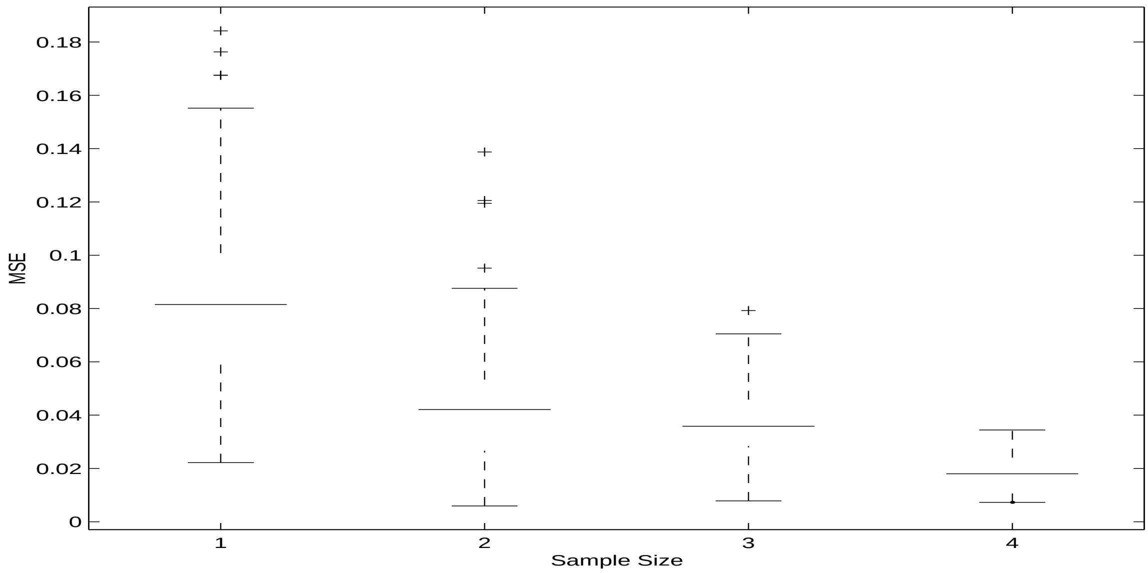

Figure 5.4: Mean Square Errors. From left to right: $T = 512, 1024, 2048, 4096$.

Table 1: Results of the model selection procedure

Step	Model	-2 log-GML	p-value
1	$\beta_1 + \beta_2(u - 0.5) + g_1(\omega) + g_2(u)$	500.25	NA
2(Best)	$\beta_1 + \beta_2(u - 0.5) + g_1(\omega) + g_2(u) + g_3(\omega, u)$	476.37	$<$0.0001 (2 vs 1)
3	Full Model	476.18	0.4537 (3 vs 2)

additive model and then add one term at a time up to the full model of (2.4). Applying the result of Lemma 4.1, we can choose the best model. Table 1 summarizes the results. The second column describes the model, the third column is -2 times log-GML. Finally, the fourth column is the p-value calculated using -2 times log-GML ratio compared with a $50 : 50$ mixture of χ_1^2 and χ_0^2. We conclude that the model with $g_4(.) = 0$ is the best model. This means that the EEG data can not be adequately modeled by a modulated stationary time series model because of the presence of $g_3(.)$ in the final model.

Figure 6.1 shows the estimates, together with the 95% Bayesian confidence intervals estimated using the best model. From the plot, we can see that at all times power at the lowest frequency is about 4 times greater that the highest frequency. Before the seizure, the spectrum is almost stationary. Then we start seeing some ripples in the log-spectrum which turn into a huge surge. Energy in all frequencies increases by almost 5 times. This result is similar to the result reported in Ombao *et al.* (2001). However, modeling the log-spectrum instead of the spectrum allows us to quantify more details in the higher frequencies which may otherwise be dominated by the more dramatic changes in the lower frequencies. After the seizure, the power begins to decrease, but by the end of the monitored period, it has not returned to the level before the seizure. This can not be directly identified in the time domain plot showed in Figure 1.1.

7 Conclusion

We have proposed to use SS-ANOVA and SLEX transformation together as a general tool for time-dependent spectral analysis, which enables joint estimation in time and frequency domains. Inference procedures such as confidence intervals and hypothesis testing proposed in the SS-ANOVA context can be adopted for inference on the time-varying spectrum. Unlike most of the existing methods that focus on estimating a discrete time-varying spectrum which is only asymptotically tied to a smooth

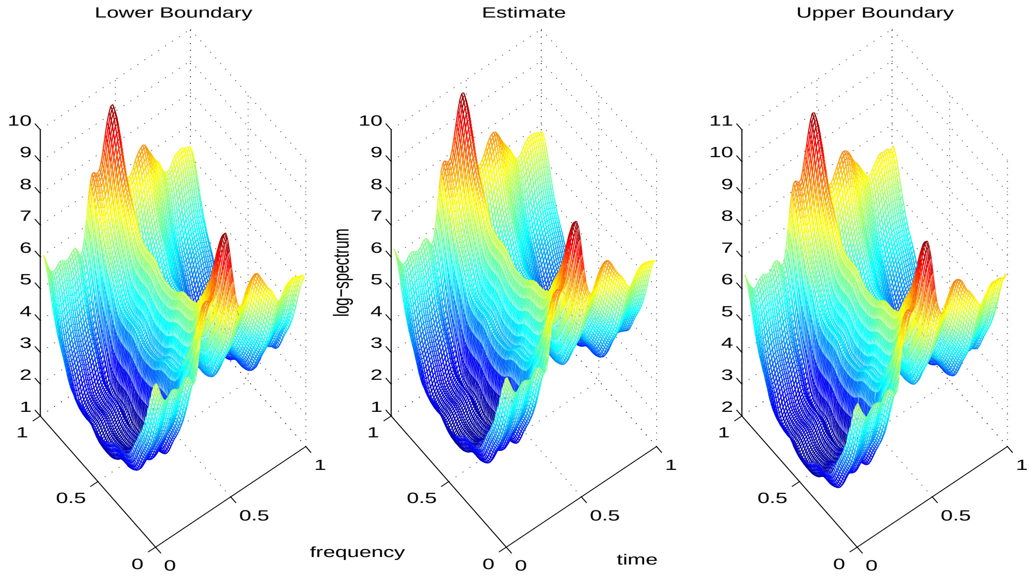

Figure 6.1: The left and the right are the lower and upper boundaries of the 95% confidence intervals. The middle is the smoothed estimate of the log-spectrum.

object, we directly model the continuous underlying spectrum, which is assumed to be smooth and has interpretation at any time-frequency point, even in finite samples.

We proposed a two-stage approach to estimate the underlying spectrum. In order to obtain initial estimates of the time-varying periodograms, we first need to partition the time series into small blocks that are approximately stationary. The second stage analysis uses SS-ANOVA to borrow strength from neighboring blocks to reconstruct a smooth continuous evolutionary spectrum. Unlike most of the existing methods that estimate the discrete spectrum conditional on the partition, our initial partition is only to obtain initial estimates of the periodograms and therefore is not critical to the final estimates. Because of the smoothness assumption on the underlying spectrum and the SS-ANOVA framework, once we obtain the estimate for a given time-frequency grid, we can obtain an estimate of the spectrum at any given time-frequency point. This leads to great computational efficiency especially for large data sets, as we only need to calculate the initial estimates of the periodograms at a coarser time-frequency grid.

One potential limitation of our proposed method is its smoothness assumption which can not handle abrupt jumps in the time-varying spectrum. In contrast, the discrete SLEX model (Ombao *et al.*, 2002) can handle abrupt jumps of finite heights at the break points of the dyadic segmentation. It is possible to combine the two methods to estimate a time varying spectrum that is smooth except for a few abrupt jumps, which we will pursue in our future research.

Appendix A

Proofs of the Theorem 2.1, 2.2 and Lemma 3.1

To prove Theorems 2.1 and 2.2, we need the following propositions. The first is about the orthogonal property of the SLEX vectors, which can be found in Wickerhauser (1994).

Proposition A.1

$$\sum_{t=1}^{T} \phi_l(\omega_1, t)\overline{\phi_m(\omega_2, t)} = \begin{cases} 0 & \textit{if } l \neq m, \\ \sum_{t \in B_l} \exp\{i2\pi(\omega_1 - \omega_2)t\} & \textit{if } l = m. \end{cases} \tag{A.1}$$

Proposition A.2 *If the transfer function $A(\omega, u)$ satisfies the Assumption 2.1, then it is Hölder continuous w.r.t. ω and u.*

Proposition A.3 *For a locally stationary SLEX process $\{X_t, t = 1, \cdots, T\}$ and the spectral function $f(\omega, u)$ which has continuous partial derivatives w.r.t. ω and u, then:*

$$T^{-1} \sum_{t=1}^{T} Var(X_t) = \int_0^1 \int_0^1 f(\omega, u) d\omega du + O\{\max(S_k)/T\}. \tag{A.2}$$

Proof:

From Equation 2.3 and the fact that the SLEX vectors are compactly supported and $f^S(\omega, u)$ is Hölder continuous w.r.t. u, we have that:

$$\begin{aligned}
1/T \sum_{t=1}^{T} Var(X_t) &= 1/T \sum_{t=1}^{T} \sum_{k,l=1}^{M} \int_0^1 f^S(\omega, t/T)\phi_k(\omega, t)\overline{\phi_l(\omega, t)} d\omega, \\
&= 1/T \sum_{k=1}^{M} \sum_{t=1}^{T} \int_0^1 f^S(\omega, t/T)\phi_k(\omega, t)\overline{\phi_k(\omega, t)} d\omega \\
&\quad + 1/T \sum_{|k-l|=1} \sum_{t=1}^{T} \int_0^1 f^S(\omega, t/T)\phi_k(\omega, t)\overline{\phi_l(\omega, t)} d\omega, \\
&= 1/T \sum_{k=1}^{M} \sum_{t=1}^{T} \int_0^1 f^S(\omega, u_k)\phi_k(\omega, t)\overline{\phi_k(\omega, t)} d\omega \\
&\quad + 1/T \sum_{|k-l|=1} \sum_{t=1}^{T} \int_0^1 f^S(\omega, u_k)\phi_k(\omega, t)\overline{\phi_l(\omega, t)} d\omega + \mathrm{O}\{\max(S_k)/T\}, \\
&= 1/T \sum_{k=1}^{M} \sum_{t=1}^{T} \int_0^1 f^S(\omega, u_k)\phi_k(\omega, t)\overline{\phi_k(\omega, t)} d\omega + \mathrm{O}\{\max(S_k)/T\}, \\
&= 1/T \sum_{k=1}^{M} S_k \int_0^1 f^S(\omega, u_k) d\omega + \mathrm{O}\{\max(S_k)/T\},
\end{aligned}$$

where u_k is the middle point of the block B_k. The last two equations are from Proposition A.1. Now replace the sum over k by integration, the above equation is equal to:

$$= \int_0^1 \int_0^1 f^S(\omega, u) d\omega du + \mathrm{O}(\sum_{k=1}^{M} S_k^2/T^2) + \mathrm{O}\{\max(S_k)/T\}. \tag{A.3}$$

Now consider $\sum_{k=1}^{M} S_k^2/T^2$. We have

$$\sum_{k=1}^{M} S_k^2/T^2 \leq \max(S_k) \sum_{t=1}^{M} S_k/T^2 = \max(S_k)/T. \tag{A.4}$$

The Proposition follows from Equation A.3 and A.4. $\square$

Proof of Theorem 2.1 Given a SLEX locally stationary process $\{X_t^S, t = 1, \cdots, T\}$ with transfer function $A^S(\omega, u)$, we have

$$X_t^S = \sum_{j=1}^{M} \int_0^1 \phi_j(\omega, t/T) A^S(\omega, t/T) dZ^S(\omega). \tag{A.5}$$

Let $A^D(\omega, u) = A^S(\omega, u) = A(\omega, u)$ and $Z^D(\omega), \omega \in [0, 1]$ be another increment process such that

$$\mathrm{cov}\{dZ^S(\omega), dZ^D(\lambda)\} = \delta(\omega - \lambda) d\omega. \tag{A.6}$$

Let a Dahlhaus locally stationary process $\{X_t^D, t = 1, \cdots, T\}$ be given by the following:

$$X_t^D = \int_0^1 \exp(i2\pi\omega t) A^D(\omega, t/T) dZ^D(\omega). \tag{A.7}$$

To prove this Theorem, we rewrite

$$1/T \sum_{t=1}^{T} E|X_t^S - X_t^D|^2 = 1/T \sum_{t=1}^{T} \{\mathrm{var}(X_t^S) + \mathrm{var}(X_t^D) - 2\,\mathrm{cov}(X_t^S, X_t^D)\}. \tag{A.8}$$

From Proposition A.3 and Equation A.7, we have:

$$\begin{aligned} 1/T \sum_{t=1}^{T} \mathrm{var}(X_t^S) &= \textstyle\int_0^1 \int_0^1 f(\omega, u) d\omega du + \mathrm{O}\{\max(S_k)/T\}, \\ 1/T \sum_{t=1}^{T} \mathrm{var}(X_t^D) &= \textstyle\int_0^1 \int_0^1 f(\omega, u) d\omega du + \mathrm{O}(1/T). \end{aligned} \tag{A.9}$$

Let $\tilde{B}_j = [\alpha_j - \epsilon_j, \alpha_{j+1} + \epsilon_{j+1}]$ be the support of the SLEX vector $\phi_j(\omega, \cdot)$. From Equation A.5 and A.7, we have:

$$\begin{aligned} 1/T \sum_{t=1}^{T} \mathrm{cov}(X_t^S, X_t^D) &= 1/T \sum_{t=1}^{T} \sum_{j=1}^{M} \textstyle\int_0^1 \phi_j(\omega, t) f(\omega, t/T) \exp(-i2\pi\omega t) d\omega, \\ &= 1/T \sum_{j=1}^{M} \sum_{t \in \tilde{B}_j} \textstyle\int_0^1 \phi_j(\omega, t) f(\omega, t/T) \exp(-i2\pi\omega t) d\omega. \end{aligned}$$

Since $f(\omega,u)$ is Hölder continuous w.r.t. u, so for any $t \in \tilde{B}_j$, $|f(\omega,t/T)-f(\omega,u_j)| \le C|t/T-u_j| \le C(S_j+\epsilon_j+\epsilon_{j+1})/(2T) \le C\max(S_j)/T$, where u_j is the middle point of block B_j. So

$$\begin{aligned}
&|1/T\sum_{t=1}^{T}\mathrm{cov}(X_t^S,X_t^D) - 1/T\sum_{j=1}^{M}\sum_{t\in\tilde{B}_j}\int_0^1 \phi_j(\omega,t)f(\omega,u_j)\exp(-i2\pi\omega t)d\omega| \\
&= 1/T|\sum_{j=1}^{M}\sum_{t\in\tilde{B}_j}\int_0^1 \phi_j(\omega,t)\{f(\omega,t/T)-f(\omega,u_j)\}\exp(-i2\pi\omega t)d\omega| \\
&\le 1/T\sum_{j=1}^{M}\sum_{t\in\tilde{B}_j}\int_0^1 |\phi_j(\omega,t)||f(\omega,t/T)-f(\omega,u_j)||\exp(-i2\pi\omega t)|d\omega \\
&\le 1/T\sum_{j=1}^{M}\sum_{t\in\tilde{B}_j}\int_0^1 C\max(S_j)/T|\phi_j(\omega,t)|d\omega \le 1/T\sum_{j=1}^{M}S_j 2C\max(S_j)/T = 2C\max(S_j)/T.
\end{aligned} \tag{A.10}$$

From the construction of the SLEX vectors, we have the following:

$$|1/S_j\sum_{t\in\tilde{B}_j}\phi_j(\omega,t)\exp(-2i\pi\omega t) - 1| \le 4\max(\epsilon_j)/\min(S_j).$$

Then we have:

$$\begin{aligned}
&|1/T\sum_{j=1}^{M}\sum_{t\in\tilde{B}_j}\int_0^1 \phi_j(\omega,t)f(\omega,u_j)\exp(-i2\pi\omega t)d\omega - 1/T\sum_{j=1}^{M}\int_0^1 S_j f(\omega,u_j)d\omega| \\
&= 1/T|\sum_{j=1}^{M}\int_0^1 f(\omega,u_j)\sum_{t\in\tilde{B}_j}\{\phi_j(\omega,t)\exp(-i2\pi\omega t)-S_j\}d\omega| \\
&\le 1/T\sum_{j=1}^{M}\int_0^1 S_j f(\omega,u_j)|1-1/S_j\sum_{t\in\tilde{B}_j}\phi_j(\omega,t)\exp(-2i\pi\omega t)|d\omega \\
&\le 1/T\sum_{j=1}^{M}\int_0^1 S_j\ f(\omega,u_j)\ d\omega\ 4\max(\epsilon_j)/\min(S_j) \le 4C'\max(\epsilon_j)/\min(S_j)\ ,
\end{aligned} \tag{A.11}$$

where $C' = \sup_{\omega,u}\{f(\omega,u)\}$.

From Equation A.10 and A.11, we have:

$$|1/T\sum_{t=1}^{T}\mathrm{cov}(X_t^S,X_t^D) - 1/T\sum_{j=1}^{M}\int_0^1 S_j f(\omega,u_j)d\omega| = \mathrm{O}\{\max(S_j)/T\} + \mathrm{O}\{\max(\epsilon_j)/\min(S_j)\}\ . \tag{A.12}$$

Replace the sum over the blocks by the integral for $u \in [0,1]$, we have:

$$1/T\sum_{j=1}^{M}\int_0^1 S_j f(\omega,u_j)d\omega = \int_0^1\int_0^1 f(\omega,u)d\omega du + \mathrm{O}\{\max(S_j)/T\}. \tag{A.13}$$

From Equation A.12 and A.13,

$$1/T\sum_{t=1}^{T}\mathrm{cov}(X_t^S,X_t^D) = \int_0^1\int_0^1 f(\omega,u)d\omega du + \mathrm{O}\{\max(S_j)/T\} + \mathrm{O}\{\max(\epsilon_j)/\min(S_j)\}. \tag{A.14}$$

The Theorem follows from Equations A.8, A.9 and A.14. $\square$

Proof of Theorem 2.2

From Assumption 2.1, the transfer function $A(\omega, u)$ and the spectrum function $f(\omega, u)$ are Hölder continuous w.r.t. ω, u.

Let the SLEX coefficient $\hat{\theta}_{j,k,T}$ given by 3.2, we have:

$$\begin{aligned}
E|\hat{\theta}_{j,k,T}|^2 &= 1/S_j \sum_{t,s=1}^{T} E(X_t X_s)\overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s), \\
&= 1/S_j \sum_{t,s=1}^{T} \sum_{l,m=1}^{M} \int_0^1 A^S(\lambda, t/T)\overline{A^S(\lambda, s/T)}\phi_l(\lambda, t)\overline{\phi_m(\lambda, s)}d\lambda \ \overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s).
\end{aligned}$$

Since the SLEX vectors are compactly supported, the summations for t, s and l, m in the above equation only need to be considered for t, s in the support of $\phi_j(\omega_k, \cdot)$ and $l, m = j-1, j, j+1$. Since $A^S(\lambda, \tau)$ is Hölder continuous w.r.t. τ, $A^S(\lambda, t/T) = A^S(\lambda, u_j) + \mathrm{O}(S_j/T)$ for any t in the support of the SLEX vectors $\phi_j(\omega_k, \cdot)$, where u_j is the middle point of the block B_j. Hence, from Proposition A.1, the above is equal to:

$$\begin{aligned}
&= 1/S_j \sum_{l,m=1}^{M} \sum_{t,s=1}^{T} \int_0^1 A^S(\lambda, u_j)\overline{A^S(\lambda, u_j)}\phi_l(\lambda, t)\overline{\phi_m(\lambda, s)}d\lambda \ \overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s) \\
&\quad + \mathrm{O}\{\max(S_m^2)/T\}, \\
&= 1/S_j \sum_{t,s=1}^{T} \int_0^1 |A^S(\lambda, u_j)|^2 \phi_j(\lambda, t)\overline{\phi_j(\lambda, s)}d\lambda \ \overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s) + \mathrm{O}\{\max(S_m^2)/T\}. \qquad \text{(A.15)} \\
&= 1/S_j \int_0^1 |A^S(\lambda, u_j)|^2 \sum_{t,s \in B_j} \exp\{i2\pi(\lambda - \omega_k)(t-s)\}d\lambda + \mathrm{O}\{\max(S_m^2)/T\}, \\
&= f^S(\omega_k, u_j) + \mathrm{O}(1/S_j) + \mathrm{O}\{\max(S_m^2)/T\}.
\end{aligned}$$

Since the time series $\{X_t, t = 1, \cdots, T\}$ also has the Dahlhaus representation (2.1), we have:

$$\begin{aligned}
E|\hat{\theta}_j(\omega_k)|^2 &= 1/S_j \sum_{t,s=1}^{T} E(X_t X_s)\overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s), \\
&= 1/S_j \sum_{t,s=1}^{T} \int_0^1 A^D(\lambda, t/T)\overline{A^D(\lambda, s/T)} \exp\{i2\pi\lambda(t-s)\}d\lambda\overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s), \qquad \text{(A.16)} \\
&= 1/S_j \sum_{t,s=1}^{T} \int_0^1 |A^D(\lambda, u_j)|^2 \exp\{i2\pi\lambda(t-s)\}d\lambda\overline{\phi_j(\omega_k, t)}\phi_j(\omega_k, s) + \mathrm{O}(S_j^2/T).
\end{aligned}$$

The last equation is because that $A^D(\omega, u)$ is Hölder continuous w.r.t. u.

From the construction of the SLEX vectors, we have that:

$$|\sum_{t=1}^{T} \phi(\omega_k, t)\exp(-i2\pi\lambda t) - \sum_{t \in B_j} \exp\{i2\pi(\omega_k - \lambda)t\}| = \mathrm{O}\{\max(\epsilon_l)\}.$$

By analogy to the consideration above on the derivation of Equation A.15, Equation A.16 is equal to:

$$\begin{aligned}&= \mathrm{O}(S_j^2/T) + 1/S_j \int_0^1 |A^D(\lambda, u_j)|^2 \sum_{t,s\in B_j} \exp\{i2\pi(\omega_k - \lambda)(t-s)\}d\lambda + \mathrm{O}[\max(\epsilon_l^2)\}/S_j^2],\\ &= \mathrm{O}(S_j^2/T) + f^D(\omega_k, u_j) + \mathrm{O}(1/S_j) + \mathrm{O}\{\max(\epsilon_l^2)/S_j\},\end{aligned} \tag{A.17}$$

From Equation A.15 and A.17, we have that

$$|f^D(\omega_k, u_j) - f^S(\omega_k, u_j)| = o(1). \tag{A.18}$$

Since $f^D(\omega, u)$ and $f^S(\omega, u)$ are Hölder continuous w.r.t. ω, u, then for any $\omega, u \in [0,1]$, let $[uT] \in B_j$ for some j and $k/S_j < \omega \le (k+1)/S_j$ for some $k \in \{1, \cdots, S_j\}$. We have:

$$\begin{aligned}|f^S(\omega, u) - f^S(\omega_k, u_j)| = \mathrm{O}(S_j/T) + \mathrm{O}(1/S_j),\\ |f^D(\omega, u) - f^D(\omega_k, u_j)| = \mathrm{O}(S_j/T) + \mathrm{O}(1/S_j).\end{aligned} \tag{A.19}$$

From Equation A.18 and A.19, we have $|f^S(\omega, u) - f^D(\omega, u)| = o(1)$ as $\max(S_j)/T \to 0$ and $\min(S_j) \to \infty$. □

Proof of Lemma 3.1 First we have:

$$\begin{aligned}\hat\theta_{j,k,T} &= 1/\sqrt{S_j}\sum_{t=1}^{T} X_t\overline{\phi_j(\omega_k, t)}\\
&= 1/\sqrt{S_j}\sum_{t=1}^{T}\sum_{l=1}^{M}\int_0^1 A^S(\lambda, t/T)\phi_l(\lambda, t)dZ(\lambda)\overline{\phi_j(\omega_k, t)}\\
&= 1/\sqrt{S_j}\sum_{t=1}^{T}\sum_{l=\max(j-1,1)}^{\min(j+1,M)}\int_0^1 A^S(\lambda, t/T)\phi_l(\lambda, t)dZ(\lambda)\overline{\phi_j(\omega_k, t)}\\
&= 1/\sqrt{S_j}\sum_{l=1}^{M}\sum_{t=1}^{T}\int_0^1 A^S(\lambda, u_j)\phi_l(\lambda, t)dZ(\lambda)\overline{\phi_j(\omega_k, t)} + r_1\\
&= 1/\sqrt{S_j}\sum_{t=1}^{T}\int_0^1 A^S(\lambda, u_j)\phi_j(\lambda, t)dZ(\lambda)\overline{\phi_j(\omega_k, t)} + r_1\\
&= 1/\sqrt{S_j}\sum_{t=\alpha_j+\epsilon_j}^{\alpha_{j+1}-\epsilon_{j+1}}\int_0^1 A^S(\lambda, u_j)\exp\{2\pi(\lambda-\omega_k)t\}dZ(\lambda) + r_1 + r_2.\end{aligned}$$

where

$$\begin{aligned}r_1 &= 1/\sqrt{S_j}\sum_{t=1}^{T}\sum_{l=\max(j-1,1)}^{\min(j+1,M)}\int_0^1\{A^S(\lambda, t/T) - A^S(\lambda, u_j)\}\phi_l(\lambda, t)dZ(\lambda)\overline{\phi_j(\omega_k, t)}\\
r_2 &= 1/\sqrt{S_j}\sum_{m=j}^{j+1}\sum_{t=\alpha_m-\epsilon_m}^{\alpha_m+\epsilon_m}\int_0^1 A^S(\lambda, u_j)\phi_j(\lambda, t)dZ(\lambda)\overline{\phi_j(\omega_k, t)}\end{aligned}$$

It is easy to see that $E(r_1) = E(r_2) = 0$. Similar to the proof of Theorem 2.2, using the Hölder continuity of $A^S(\omega, u)$ w.r.t. u and that the SLEX vectors are orthogonal and compactly supported, we get the following:

$$
\begin{aligned}
E|r_1|^2 = Var(r_1) &= 1/S_j \sum_{t,s=1}^{T} \sum_{l,m=\max(j-1,1)}^{\min(j+1,M)} \int_0^1 \{A^S(\lambda,t/T) - A^S(\lambda,u_j)\} \\
&\quad \overline{\{A^S(\lambda,s/T) - A^S(\lambda,u_j)\}}\phi_l(\lambda,t)\overline{\phi_m(\omega_k,s)}\phi_j(\omega_k,s)\overline{\phi_j(\omega_k,t)}d\lambda \\
&= 1/S_j \sum_{l,m=\max(j-1,1)}^{\min(j+1,M)} \sum_{t\in B_l, s\in B_m} \mathrm{O}\{\max(S_j)/T\} = \mathrm{O}\{\max(S_j^2)/T\}. \\
E|r_2|^2 = Var(r_2) &= 1/S_j \sum_{m=j}^{j+1} \sum_{t,s=\alpha_m-\epsilon_m}^{\alpha_m+\epsilon_m} \int_0^1 |A^S(\lambda,u_j)|^2 \phi_j(\lambda,t)\overline{\phi_j(\lambda,s)}d\lambda\overline{\phi_j(\omega_k,t)}\phi_j(\omega_k,s) \\
&= 1/S_j \sum_{m=j}^{j+1} \sum_{t,s=\alpha_m-\epsilon_m}^{\alpha_m+\epsilon_m} \mathrm{O}(1) = \mathrm{O}\{\max(\epsilon_k^2)/S_j\}.
\end{aligned}
$$

As $\max(S_j^2)/T = o(1)$ and $\max(\epsilon_j^2)/\min(S_j) = o(1)$, r_1, r_2 converge to 0 in probability. So $\hat{\theta}_{j,k,T}$ has the same asymptotic distribution as $1/\sqrt{S_j} \sum_{t=\alpha_j+\epsilon_j}^{\alpha_{j+1}-\epsilon_{j+1}} \int_0^1 A^S(\lambda,u_j)\exp\{i2\pi(\lambda-\omega_k)t\}dZ(\lambda)$. Since $A^S(\lambda,u_j)$ does not depend on time t, the Lemma then follows from the result for a stationary process. □

References

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

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68**, 337 – 404.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary process. *Stochastic Processes and their Applications*, **62**, 139 – 168.

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

Gu, C. (1989). Rkpack and its application: fitting smoothing spline models. *Proc. Statist. Comput. Sec: American Statistical Association*, pages 42 – 51.

Gu, C. and Wahba, G. (1993a). Semiparametric analysis of variance with tensor product thin plate splines. *Journal of the Royal Statistical Society, Series B*, **55**, 353 – 368.

Gu, C. and Wahba, G. (1993b). Smoothing spline ANOVA with components-wise Bayesian confidence intervals. *Journal of Computational and Graphical Statistics*, **2**, 97 – 117.

Guo, W. (2000). Inference in smoothing spline ANOVA. *Submitted for publication.*

Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics*, **25**, 38 – 76.

Nychka, D. (1988). Bayesian confidence intervals for smoothing spline. *Journal of the American Statistical Association*, **83**, 1134 – 1143.

Ombao, H., von Sachs, R., and Guo, W. (2000). Estimation and inference for time-varying spectra of locally stationary SLEX processes. *Discussion Paper 00/27, Institut de Statistique, UCL, Louvain-la-Neuve; in: 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. (2001). Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, **96**, 543 – 560.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002). The SLEX model of a nonstationary time series. *Discussion Paper 00/26, Institut de Statistique, UCL, Louvain-la-Neuve; to appear in: Ann. Inst. Stat. Math. (2002), Vol. 54, No. 1.*

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.

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

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, **45**, 133 – 150.

Wahba, G. (1990). *Spline Model for Observational Data.* CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM.

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