

Nonparametric Spectral Analysis With Applications to Seizure Characterization Using EEG Time Series

Li Qin

Program in Biostatistics

Vaccine and Infectious Diseases Institute

Division of Public Health Sciences

Fred Hutchinson Cancer Research Center, Seattle, Washington 98109

lqin@fhcrc.org

Yuedong Wang

Department of Statistics and Applied Probability

University of California, Santa Barbara, California 93106

yuedong@pstat.ucsb.edu

May 14, 2008

Summary

Understanding the seizure initiation process and its propagation pattern(s) is a critical task in epilepsy research. Characteristics of the pre-seizure electroencephalograms (EEGs) such as oscillating powers and high-frequency activities are believed to be indicative of the seizure onset and spread patterns. In this article, we analyze epileptic EEG time series using nonparametric spectral estimation methods to extract information on seizure-specific power and characteristic frequency (or frequency band(s)). Because the EEGs may become non-stationary before seizure events, we develop methods for both stationary and local stationary

processes. Based on penalized Whittle likelihood, we propose a direct generalized maximum likelihood (GML) and generalized approximate cross-validation (GACV) methods to estimate smoothing parameters in both smoothing spline spectrum estimation of a stationary process and smoothing spline ANOVA time-varying spectrum estimation of a locally stationary process. We also propose permutation methods to test if a locally stationary process is stationary. Extensive simulations indicate that the proposed direct methods, especially the direct GML, are stable and perform better than other existing methods. We apply the proposed methods to the intracranial electroencephalograms (IEEGs) of an epileptic patient to gain insights into the seizure generation process.

Key Words: EEG, epilepsy, GACV, GML, locally stationary process, permutation test, smoothing parameter, smoothing spline, SS ANOVA

1 Introduction

Roughly 1% of the population in developed nations suffers from epilepsy. Of these about 30% have medically refractory epilepsy, where the most devastating feature is seizure. The only hope for relieving these patients from the disabling seizures is resective surgery, while the surgical success rate varies between less than 25% to 70% depending on how well the seizure initiation zone could be removed or the seizure propagation path could be disconnected (Schiller, Cascino and Sharbrough 1998a). Therefore, a better understanding of the seizure initiation process and its propagation patterns is crucial to the success of a surgery. Epileptic patients usually undergo presurgical evaluations where they are monitored over time with EEGs recorded by electrodes placed at different locations of the brain called channels. The

locations of the EEG channels are believed to give the best coverage of the epileptogenic zone, determined by the neurosurgeons. While EEGs from 128 or 256 channels are recorded, only a few of them cover the origin or critical path of the seizure generation. Therefore, the identification of the seizure focus and its spread patterns would improve clinical judgement on where to resect that would render the patients seizure free. Characteristics of the pre-seizure (“pre-ictal”) EEGs including oscillating powers and high-frequency activities are believed to be indicative of the seizure onset and spread patterns. They behave differently from those characteristics of the baseline (“inter-ictal”) EEGs (Winterhalder, Taiwald, Voss, Aschenbrenner-Scheibe, Timmer and Schulze-Bonhage 2003).

Spectral analysis of EEG time series plays a central role in epilepsy research. Due to the large heterogeneity in pathology between patients, seizure-specific characterization has to be initially performed within each subject. Identifying seizure-specific power changes and seizure-characteristic frequency band(s) is a key step of this endeavor and the main focus of this paper. Methods proposed in this paper provide a systematic tool for this key step. As shown in the simulations, the proposed direct GML and GACV methods are more robust and efficient than existing methods.

Spectral analysis is an important field in time series analysis (Brillinger 1981, Shumway and Stoffer 2000). The spectrum is often used to describe the power distribution and stochastic variation of a time series. For stationary time series, harmonic trends and power evolution can be explained by the spectrum at different frequencies. It is well known that stationarity does not always hold for EEG time series (Lopes da Silva 1978). Locally stationary processes have been proposed to approximate the non-stationary time series (Dahlhaus 1997, Ombao, Raz, von Sachs and Guo 2002, Guo, Dai, Ombao and von Sachs 2003). The time-varying spectrum of a locally stationary time series characterizes changes of the stochastic variation

over time that may reflect important patterns of biological activity. In particular, our spectral analysis of the EEGs can be used to aid the clinical judgement in two ways: first, due to the one-to-one correspondence between the spectrum and the variance of the EEG time series, changes in brain power before, during and after a seizure can be characterized by the estimated time-varying spectra of the pre-ictal and inter-ictal EEGs; second and more importantly, from a signal processing point of view, the phase information between two or more channels reveals whether the signals are synchronized, and at what frequencies (or frequency bands) if they are. This information provides an important marker of seizure initiation and localization (Mormann, Andrzejak, Kreutz, Rieke, David, Elger and Lehnertz 2003). By comparing the pre-seizure time-varying spectrum with that of the baseline (seizure-free) segments within a channel, the channel-specific seizure-characteristic frequencies can be identified (if such characteristics exist). If the seizure-specific frequency band(s) in one channel overlaps with that in another channel, then there will likely be the signal coupling between the two locations. Resecting either of the location would help to reduce surgical failure. However if the frequency band(s) do not overlap, then there may be across-frequency interactions that associated with the seizure propagation patterns (Mormann, Lehnertz, David and Elger 2000).

As in most epileptic EEG studies, we analyze the intracranial EEGs (IEEGs) as the recordings have less artifacts. Figure 1 shows 5-minute IEEG segments from a patient with medicine-resistant mesial temporal lobe epilepsy right before a seizure’s clinical onset (upper panel) and at baseline (lower panel) extracted at least four hours before the seizure’s onset. The data were collected by the EEG Lab of University of Pennsylvania (D’Alessandro, Vachtsevanos, Esteller, Echauz and Litt 2001). It is visually unclear what the seizure-specific frequency band and its time-varying patterns are for both channels. We will describe analyses

of this data in Section 5.

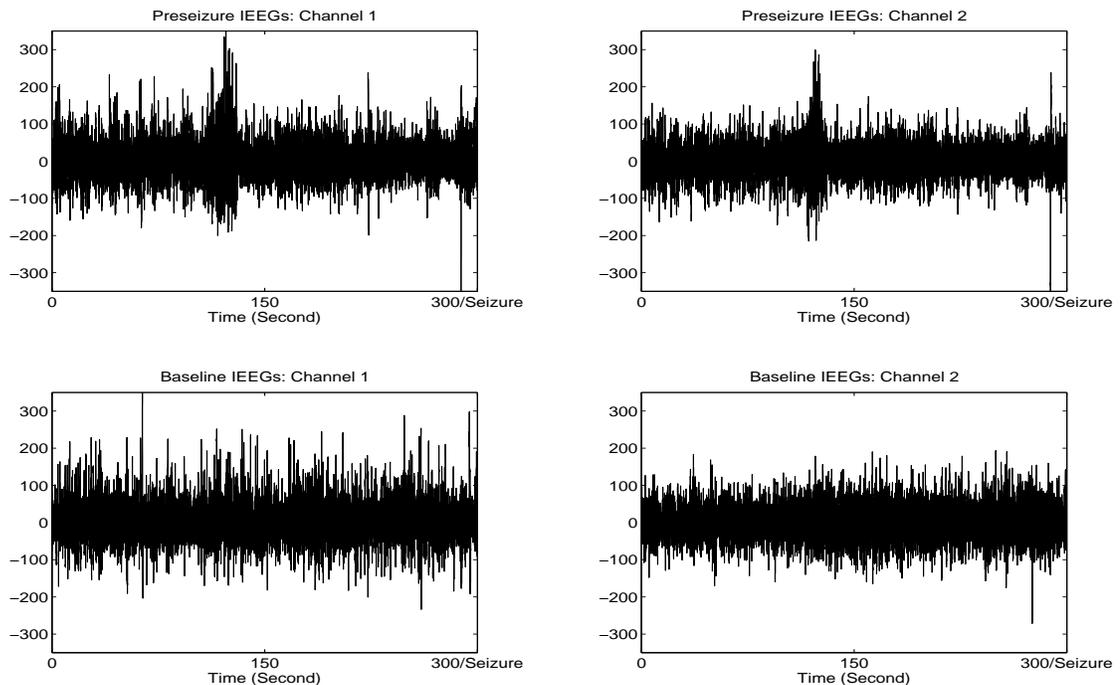


Figure 1: IEEG segments from two channels of an epileptic patient. The upper panels show 5 minute preseizure segments with the seizure’s onset at the 5th minute. The lower panels show 5 minute baseline segments collected hours away from the seizure. The sampling rate is 200 Hertz. The total number of time points is 60000 for each segment.

It is well known that the periodograms is an unbiased, but not consistent, estimator of the spectrum. Therefore, periodogram smoothing is a popular tool for nonparametric spectra estimation. Smoothing techniques including kernel smoother (Lee 1997, Ombao, Raz, Strawderman and von Sachs 2001, Hannig and Lee 2004), smoothing spline (Wahba 1980, Pawitan and O’Sullivan 1994, Guo et al. 2003), regression spline (Kooperberg, Stone and Truong 1995), local polynomial (Fan and Kreutzberger 1998) and wavelet (Gao 1997) have been applied to obtain consistent estimators. One may smooth periodograms directly

(Lee 1997, Ombao et al. 2001, Hannig and Lee 2004), smooth logarithmic periodograms (Wahba 1980, Guo et al. 2003), or use Whittle likelihood (Pawitan and O’Sullivan 1994, Guo et al. 2003). In this paper, we consider smoothing spline estimation based on the Whittle likelihood.

It is well-recognized that the choice of smoothing parameters is crucial to the performance of the smoothing methods. Many methods have been developed for kernel smoother (Lee 1997, Ombao et al. 2001, Hannig and Lee 2004) and smoothing spline (Wahba 1980, Pawitan and O’Sullivan 1994). Generalized cross-validation (GCV), generalized maximum likelihood (GML) and unbiased risk (UBR) criteria can be used to select the smoothing parameter when fitting smoothing spline models on the logarithms scale (Wahba 1990). The logarithmic transformation is the first order approximation which is less efficient than the Whittle likelihood (Pawitan and O’Sullivan 1994, Fan and Kreutzberger 1998, Guo et al. 2003). For fitting smoothing splines using penalized Whittle likelihood, Pawitan and O’Sullivan (1994) developed a criterion based on an estimate of the risk function for the selection of the smoothing parameter. Guo et al. (2003) used an indirect GML method to select the smoothing parameter. The indirect approach does not guarantee convergence and may have inferior performance than direct methods (Section 4).

We develop new direct data-driven methods for selecting smoothing parameters in the estimation of a spectral density and tests for the hypothesis that a locally stationary process is stationary. The rest of the paper is structured as follows. We present direct GML and GACV methods for stationary processes in Sections 2. We develop direct GML method, GACV method and stationarity tests for locally stationary processes in Sections 3. The simulation study is summarized in Section 4. The analysis of IEEG time series is presented in Section 5. We conclude with some remarks in Section 6.

2 Stationary Processes

2.1 Notations

Let $X_t, t = 0, \pm 1, \pm 2, \dots$, be a stationary time series with mean zero and covariance function $\gamma(u) = E(X_t X_{t+u})$. The second-order properties of X_t are equivalently described by the spectrum

$$f(\omega) = \sum_{u=-\infty}^{\infty} \gamma(u) \exp(-i2\pi\omega u), \quad \omega \in [0, 1], \quad (1)$$

where the imaginary unit $i^2 = -1$.

Let X_0, X_1, \dots, X_{T-1} be a finite sample of the stationary process and

$$y_k = T^{-1} \left| \sum_{t=0}^{T-1} X_t \exp(i2\pi kt/T) \right|^2, \quad k = 0, \dots, T-1, \quad (2)$$

be the periodogram at frequency $\omega_k = k/T$. Our goal is to estimate f based on observations in the form $\{(\omega_k, y_k), k = 0, \dots, T-1\}$. Under standard mixing conditions (Brillinger 1981)

$$y_k \approx f(\omega_k) \chi_{k,d_k}^2 / d_k, \quad (3)$$

where χ_{k,d_k}^2 are independent Chi-square random variables with degree of freedom $d_k = 1$ for $k = 0$ and $k = T/2$ (if T is even) and $d_k = 2$ for $1 \leq k \leq T/2 - 1$. As Pawitan and O'Sullivan (1994) and Guo et al. (2003), for simplicity, the slight difference in degrees of freedom will be ignored and all degrees of freedom are set to two.

2.2 Smoothing Spline Estimation

We model the logarithm of the spectrum $g = \log(f)$ using a periodic spline space (Wahba 1990)

$$W_2(per) = \{g : g \text{ and } g' \text{ are abs. cont.}, g(0) = g(1), g'(0) = g'(1), \int_0^1 (g''(\omega))^2 d\omega < \infty\}. \quad (4)$$

$W_2(per) = \{1\} \oplus W_2^0(per)$, where $\{1\}$ is the one-dimensional space consisting of all constant functions, and $W_2^0(per)$ is a reproducing kernel Hilbert space with reproducing kernel $R_1(\omega_1, \omega_2) = -B_4([\omega_1 - \omega_2])/24$, $[\omega_1 - \omega_2]$ is the fractional part of $\omega_1 - \omega_2$ and $B_4(\omega) = (\omega - 0.5)^4 - (\omega - 0.5)^2/2 + 7/240$ (Wahba 1990, Gu 2002). We note that Wahba (1980) and Pawitan and O’Sullivan (1994) essentially used the same model space with solutions approximated using cepstral coefficients.

Because $g(\omega)$ is symmetric about $\omega = .5$, it suffices to estimate $g(\omega)$ for $\omega \in [0, .5]$. As in Wahba (1980) and Pawitan and O’Sullivan (1994), we use all y_k ’s even though $y_k = y_{T-k}$. This is to allow periodic smoothing. Let $g_k = g(\omega_k)$. Based on (3), we estimate g as the minimizer of the following penalized Whittle likelihood (Gu 2002)

$$\sum_{k=0}^{T-1} \{g_k + y_k \exp(-g_k)\} + T\lambda \int_0^1 \{g''(\omega)\}^2 d\omega, \quad (5)$$

where λ is a smoothing parameter balancing the goodness-of-fit and the smoothness of the function g .

For a fixed λ , the solution to (5) can be represented as (Wahba, Wang, Gu, Klein and Klein 1995, Gu 2002)

$$\hat{g}(\omega) = d + \sum_{k=0}^{T-1} c_k R_1(\omega_k, \omega). \quad (6)$$

We compute coefficients d and c_k ’s iteratively using the iterative reweighted penalized least squares (IRPLS) method (McCullagh and Nelder 1989, Wahba et al. 1995). Specifically, at each iteration, we compute the *working variable* $z_k = \tilde{g}_k + y_k \exp(-\tilde{g}_k) - 1$ and *weight* $w_k = 1$ where tilde indicates current estimates. Note that, as Pawitan and O’Sullivan (1994), we use the Fisher scoring method instead of the Newton-Raphson method employed in Wahba et al. (1995), Gu (2002) and Liu, Tong and Wang (2006). Our experience indicates that the Fisher scoring method is more stable in this situation. One may select λ using the GCV,

GML or UBR criterion at each iteration of the above IRPLS procedure (Wahba et al. 1995). However, such an indirect approach may lead to non-convergence of the algorithm (Section 4).

2.3 Direct GML and GACV Methods

We simply introduce the direct GML and GACV criteria in this section. Derivations can be found in supplementary materials.

Consider the following prior for g

$$G(\omega) = \alpha + (T\lambda)^{-\frac{1}{2}}W(\omega), \quad (7)$$

where $\alpha \sim N(0, a)$, $W(\omega)$ is a Gaussian process independent of α with $E\{W(\omega)\} = 0$ and $E\{W(\omega_1)W(\omega_2)\} = R_1(\omega_1, \omega_2)$. Gu (1992) showed that, as $a \rightarrow \infty$, the posterior distribution of $G(\omega)$ can be approximated by a Gaussian distribution with posterior mean equals the spline estimate \hat{g} . This connection between smoothing splines and Bayesian models has been exploited to develop methods for selecting smoothing parameters for estimating variance functions (Liu et al. 2006) and constructing confidence intervals (Gu 1992, Wahba et al. 1995). We use this connection to develop a direct GML criterion for selecting λ .

Let $\mathbf{y} = (y_0, \dots, y_{T-1})'$, $\mathbf{g} = (g_0, \dots, g_{T-1})'$, $S = (1, \dots, 1)$ be a vector of size T , $\Sigma = \{R_1(\omega_i, \omega_j)\}_{i,j=1}^T$, $(Q_1 \ Q_2)(R' \ \mathbf{0})'$ be the QR decomposition of S , and UDU' be the spectral decomposition of $Q_2'\Sigma Q_2$ where $D = \text{diag}(\delta_1, \dots, \delta_{T-1})$ and δ_ν are eigenvalues. Let $\hat{g}_k = \hat{g}(\omega_k)$, $u_k = 1 - y_k \exp(-\hat{g}_k)$, $\hat{\mathbf{g}} = (\hat{g}_0, \dots, \hat{g}_{T-1})'$, $\mathbf{u}_c = (u_0, \dots, u_{T-1})'$, $\mathbf{y}_c = \hat{\mathbf{g}} - \mathbf{u}_c$, and $\mathbf{z} = (z_1, \dots, z_{T-1})' = U'Q_2'\mathbf{y}_c$. Ignoring some constants, the negative log marginal likelihood of \mathbf{y} can be approximated by

$$\text{GML}(\lambda) = \sum_{k=0}^{T-1} \{\hat{g}_k + y_k \exp(-\hat{g}_k)\} - \frac{1}{2}\mathbf{u}_c'\mathbf{u}_c + \frac{1}{2} \sum_{\nu=1}^{T-1} \left\{ \ln(\delta_\nu/T\lambda + 1) + \frac{z_\nu^2}{\delta_\nu/T\lambda + 1} \right\}. \quad (8)$$

The GML criterion (8) is new to the estimation of the spectrum. We minimize (8) to find an estimate of λ which is referred to as the direct GML estimate.

As Lin, Wahba, Xiang, Gao, Klein and Klein (2000), consider the comparative Kullback-Leibler criterion

$$\text{CKL}(g, \hat{g}) = \frac{2}{T} \sum_{i=0}^{T-1} \{\exp(g_i - \hat{g}_i) + \hat{g}_i\}. \quad (9)$$

The CKL criterion cannot be used directly to select the smoothing parameters since it depends on the true log-spectrum. We need a proxy for this criterion. Let $\hat{g}^{(-i)}$ be the estimate of g without the i th observation y_i . That is, $\hat{g}^{(-i)}$ is the minimizer of the penalized Whittle likelihood

$$\sum_{k \neq i} \{g_k + y_k \exp(-g_k)\} + T\lambda \int_0^1 \{g''(\omega)\}^2 d\omega.$$

Let $\hat{g}_k^{(-i)} = \hat{g}^{(-i)}(\omega_k)$. The leaving-out-one cross-validation estimate of the CKL criterion is

$$\text{CV}(\lambda) = \frac{2}{T} \sum_{i=0}^{T-1} \{y_i \exp(-\hat{g}_i^{(-i)}) + \hat{g}_i\}. \quad (10)$$

The function $\text{CV}(\lambda)$ may be used to select the smoothing parameter. However, the computation is usually expensive. An approximation of $\text{CV}(\lambda)$ is

$$\text{GACV}(\lambda) = \sum_{i=0}^{T-1} \{y_i \exp(-\hat{g}_i) + \hat{g}_i\} + \frac{\text{tr}H}{T - \text{tr}W_0^{1/2}HW_0^{1/2}} \sum_{i=0}^{T-1} y_i \exp(-\hat{g}_i) \{y_i - \exp(\hat{g}_i)\}, \quad (11)$$

where $H = (W + n\lambda\Omega)^{-1}V$, $W = \text{diag}(y_1 \exp(-\hat{g}_1)/2, \dots, y_n \exp(-\hat{g}_n)/2)$, $\Omega = Q_2(Q_2'\Sigma Q_2)^+Q_2'$, $+$ represents the Moore-Penrose generalized inverse, $V = \text{diag}(\exp(-\hat{g}_1)/2, \dots, \exp(-\hat{g}_n)/2)$ and $W_0 = \text{diag}(\exp(\hat{g}_1), \dots, \exp(\hat{g}_n))$. Equation (11) is referred to as the GACV criterion and a GACV estimate of λ is the minimizer of this criterion.

3 Locally Stationary Processes

3.1 Local Periodograms

Let $X_t, t = 0, \dots, T - 1$, be a finite sample of the following mean zero locally stationary process (Guo et al. 2003)

$$X_t = \int_0^1 A(\omega, t/T) \exp(i2\pi\omega t) dZ(\omega), \quad (12)$$

where $Z(\omega)$ is a zero-mean stochastic process on $[0, 1]$ and $A(\omega, u)$ denotes a transfer function with continuous second order derivatives for $(\omega, u) \in [0, 1] \times [0, 1]$.

The time-dependent spectrum $f(\omega, u) = \|A(\omega, u)\|^2$ is assumed to be a smooth function in both ω and u . For estimation, local periodograms are calculated on some pre-defined time-frequency grids. Specifically, the time domain is partitioned into J disjoint blocks $[b_j, b_{j+1})$ where $0 = b_1 < b_2 < \dots < b_J < b_{J+1} = 1$. Let $u_j = (b_j + b_{j+1})/2$ and ω_k 's be K frequencies in $[0, 1]$. Then the local periodograms are calculated as

$$\hat{I}_{kj} = \hat{I}(\omega_k, u_j) = \frac{|\sum_{t=b_j}^{b_{j+1}-1} X_t \exp(i2\pi\omega_k t)|^2}{|b_{j+1} - b_j|}, \quad k = 1, \dots, K, \quad j = 1, \dots, J. \quad (13)$$

Guo et al. (2003) suggested that the block size should be in the order of $T^{1/2}$ and showed that the estimation is not very sensitive to the choices of sizes for the time and frequency grids. We investigate the impact of these choices in our simulations.

3.2 SS ANOVA Estimation

As Guo et al. (2003), we model logarithm of the time-dependent spectrum $g(w, u) = \log f(\omega, u)$ using the SS ANOVA model

$$g(\omega, u) = \beta_1 + \beta_2(u - .5) + s_1(\omega) + s_2(u) + s_3(\omega, u) + s_4(\omega, u), \quad (14)$$

where $\beta_2(u - .5)$ and $s_2(u)$ are linear and smooth main effects of time, $s_1(\omega)$ is the smooth main effect of frequency, and $s_3(\omega, u)$ and $s_4(\omega, u)$ are linear-smooth and smooth-smooth interactions between frequency and time. Let $\boldsymbol{\gamma} = (\omega, u)$ and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_i)_{i=1}^n$ be the selected time-frequency grid with corresponding log-local periodograms $\mathbf{y} = (y_1, \dots, y_n) = (\log(\hat{I}_{11}), \dots, \log(\hat{I}_{KJ}))$ where $n = KJ$. We estimate g as the minimizer of the penalized Whittle likelihood

$$\sum_{i=1}^n \{g_i + y_i \exp(-g_i)\} + n \sum_{r=1}^4 \lambda_r \|P_r g\|^2, \quad (15)$$

where $g_i = g(\boldsymbol{\gamma}_i)$, λ_r 's are smoothing parameters and P_r is the projection operator onto the subspace corresponding to s_r , $r = 1, \dots, 4$. Let $\lambda_r = \lambda/\theta_r$. The solution to (15) is (Gu 2002)

$$\hat{g}(\boldsymbol{\gamma}) = d_1 + d_2(u - .5) + \sum_{i=1}^n c_i \sum_{r=1}^4 \theta_r R_r(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}), \quad (16)$$

where $R_1(\omega_1, \omega_2)$ was defined in Section 2, $R_2(u_1, u_2) = B_2(u_1)B_2(u_2)/4 - B_4([\omega_1 - \omega_2])/24$, $R_3((\omega_1, u_1), (\omega_2, u_2)) = R_1(\omega_1, \omega_2)(u_1 - .5)(u_2 - .5)$, $R_4((\omega_1, u_1), (\omega_2, u_2)) = R_1(\omega_1, \omega_2)R_2(u_1, u_2)$, and $B_2(u) = (u - .5)^2 - 1/12$. Again, for fixed λ_r 's, coefficients d_1 , d_2 and c_i 's can be computed using the IRPLS procedure. Guo et al. (2003) selected λ_r 's at each iteration using the GML method. Again, the indirect approach may lead to non-convergence and inferior performance (Section 4).

3.3 Direct GML and GACV Methods

We simply introduce the direct GML and GACV criteria in this section. Derivations can be found in supplementary materials.

Consider the following prior for g

$$G(\boldsymbol{\gamma}) = \alpha_1 + \alpha_2(u - .5) + (n\lambda)^{-\frac{1}{2}} \sum_{r=1}^4 \theta_r^{\frac{1}{2}} W_r(\boldsymbol{\gamma}), \quad (17)$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)' \sim N(\mathbf{0}, aI)$, $W_r(\boldsymbol{\gamma})$'s are Gaussian processes independent of $\boldsymbol{\alpha}$ with $E\{W_r(\boldsymbol{\gamma})\} = 0$ and $E\{W_r(\boldsymbol{\gamma})W_r(\boldsymbol{\zeta})\} = R_r(\boldsymbol{\gamma}, \boldsymbol{\zeta})$.

Let $\mathbf{g} = (g_1, \dots, g_n)'$, $S = \{1, u_i - .5\}_{i=1}^n$, $\Sigma_\theta = \{\sum_{r=1}^4 \theta_r R_r(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j)\}_{i,j=1}^n$, $(Q_1 \ Q_2)(R', \mathbf{0})'$ be the QR decomposition of S , and UDU' be the spectral decomposition of $Q_2' \Sigma_\theta Q_2$ where $D = \text{diag}(\delta_1, \dots, \delta_{n-2})$ and δ_ν are eigenvalues. Let $\hat{g}_i = \hat{g}(\boldsymbol{\gamma}_i)$, $u_i = 1 - y_i \exp(-\hat{g}_i)$, $\hat{\mathbf{g}} = (\hat{g}_1, \dots, \hat{g}_n)'$, $\mathbf{u}_c = (u_1, \dots, u_n)'$, $\mathbf{y}_c = \hat{\mathbf{g}} - \mathbf{u}_c$, and $\mathbf{z} = (z_1, \dots, z_{n-2})' = U'Q_2'\mathbf{y}_c$. Ignoring some constants, an approximation to the negative log marginal likelihood of \mathbf{y} is

$$\text{GML}(\lambda, \boldsymbol{\theta}) = \sum_{i=1}^n \{\hat{g}_i + y_i \exp(-\hat{g}_i)\} - \frac{1}{2} \mathbf{u}_c' \mathbf{u}_c + \frac{1}{2} \sum_{\nu=1}^{n-2} \left\{ \ln(\delta_\nu/n\lambda + 1) + \frac{z_\nu^2}{\delta_\nu/n\lambda + 1} \right\}, \quad (18)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_4)'$. Note that δ_ν 's and z_ν 's depend on $\boldsymbol{\theta}$. The direct GML estimates of the smoothing parameters $(\lambda, \boldsymbol{\theta})$ are the minimizers of (18).

The comparative Kullback-Leibler criterion and its cross-validation estimate are defined similarly as those in Section 2.3. Then an approximation to the cross-validation estimate of the CKL criterion is

$$\text{GACV}(\lambda, \boldsymbol{\theta}) = \sum_{i=1}^n \{y_i \exp(-\hat{g}_i) + \hat{g}_i\} + \frac{\text{tr}H}{n - \text{tr}W_0^{1/2} H W_0^{1/2}} \sum_{i=1}^n y_i \exp(-\hat{g}_i) \{y_i - \exp(\hat{g}_i)\}, \quad (19)$$

where $H = (W + n\lambda\Omega)^{-1}V$, $W = \text{diag}(y_1 \exp(-\hat{g}_1)/2, \dots, y_n \exp(-\hat{g}_n)/2)$, $\Omega = Q_2(Q_2' \Sigma_\theta Q_2)^+ Q_2'$, $V = \text{diag}(\exp(-\hat{g}_1)/2, \dots, \exp(-\hat{g}_n)/2)$ and $W_0 = \text{diag}(\exp(\hat{g}_1), \dots, \exp(\hat{g}_n))$. The GACV estimates of the smoothing parameters $(\lambda, \boldsymbol{\theta})$ are the minimizer of (19).

3.4 Permutation Tests for Stationarity

It is often of interest to test if a locally stationary process is stationary:

$$H_0 : h(u) = 0 \text{ for all } u \quad \text{vs} \quad H_1 : h(u) \neq 0 \text{ for some } u,$$

where $h(u) = \beta_2(u - .5) + s_2(u) + s_3(\omega, u) + s_4(\omega, u)$. The model under H_1 is the full SS ANOVA model (14). Denote the model under H_0 , $g(\omega, u) = \beta_1 + s_1(\omega)$, as the reduced

model. Let \hat{g}^F and \hat{g}^R be estimates of g under the full and reduced models respectively. Let $D_F = \sum_{i=1}^n \{\hat{g}_i^F + y_i \exp(-\hat{g}_i^F) - \log y_i - 1\}$ and $D_R = \sum_{i=1}^n \{\hat{g}_i^R + y_i \exp(-\hat{g}_i^R) - \log y_i - 1\}$ be deviances under the full and reduced models. We construct two statistics

$$\begin{aligned} S_1 &= D_R - D_F, \\ S_2 &= \int_0^1 \int_0^1 \{\hat{g}^F(\omega, u) - \hat{g}^R(\omega, u)\}^2 d\omega du, \end{aligned}$$

where S_1 corresponds to the Chi-square statistics commonly used for generalized linear models, and S_2 computes the L_2 distance between \hat{g}^F and \hat{g}^R . We reject H_0 when these statistics are large. The null distributions of these statistics are unknown. Under H_0 , g does not depend on u . Therefore, we can use permutation to compute null distributions. Specifically, we generate permutation samples by shuffling time grid, compute three statistics for each permutation sample, and compute p-values as the proportion of statistics based on permutation samples which are larger than those based on the original data. Small scale simulations in the supplemental materials indicate that the permutation tests perform well.

4 Simulations

4.1 Simulations for Stationary Processes

We first conduct simulations to evaluate the performance of the direct GML and GACV methods for spectral density estimation of stationary processes. We simulate data from two processes used in Wahba (1980) and Pawitan and O'Sullivan (1994):

1. AR(3): $X_t = 1.4256X_{t-1} - .7344X_{t-2} + .1296X_{t-3} + \epsilon_t$,
2. MA(4): $X_t = \epsilon_t - .3\epsilon_{t-1} - .6\epsilon_{t-2} - .3\epsilon_{t-3} + .6\epsilon_{t-4}$,

where $\epsilon_t \stackrel{iid}{\sim} N(0, 1)$. We consider two sample sizes, $T = 128$ and $T = 256$, for each process. For each setting, we repeat simulation 1000 times.

To compare with the method in Wahba (1980), we fit cubic periodic splines to the logarithmic transformed periodograms $z_k = \log(y_k) + c_k$ where $c_k = .57721$ for $k \neq 0$ and $k \neq T/2$ and $c_k = .30135$ for $k = 0$ or $k = T/2$. We use the GML method to select the smoothing parameter on the logarithmic scale. Other methods such as GCV have also been tested and the comparative results remain the same. To compare with the method in Pawitan and O’Sullivan (1994), we fit a cubic periodic spline model using the penalized likelihood (5) with the smoothing parameter λ selected as the minimizer of

$$RE(\lambda) = \sum_{k=0}^{T-1} (\hat{g}_k - v_k)^2 + 2\text{tr}\{H(\lambda)\}, \quad (20)$$

where $v_k = \hat{g}_k + y_k \exp(-\hat{g}_k) - 1$ and $H(\lambda)$ is the smoother matrix at convergence of the IRPLS procedure with the fixed λ . The criterion (20) is a direct adaption of equation (12) in Pawitan and O’Sullivan (1994). v_k ’s need to be calculated with λ close to the optimal choice. Pawitan and O’Sullivan (1994) suggested a two-step procedure: first choose λ that minimizes $RE(\lambda)$ using $v_k = \hat{g}_k + y_k \exp(-\hat{g}_k) - 1$ and denote the resulting estimates as \hat{g}_{kp} , and then choose λ that minimizes $RE(\lambda)$ using $v_k = \hat{g}_{kp} + y_k \exp(-\hat{g}_{kp}) - 1$. We use the same two-step procedure with a uniform (on a logarithmic scale) grid search over the range $e^{-25} \leq \lambda \leq e^{-1}$. As Pawitan and O’Sullivan (1994), 5 values are used in the first step and 50 values are used in the second step.

For each simulation replication, we compute the mean squared error as

$$MSE_m = \frac{1}{T} \sum_{k=0}^{T-1} (\hat{g}_k - g_k)^2,$$

where m represents the method employed to estimate g . $m = \text{LS, IM, DM, DV}$ and PO respectively represent the method in Wahba (1980), the indirect GML method, the direct

GML method, the GACV method and the method in Pawitan and O’Sullivan (1994). The indirect GML method selects λ at each iteration of the IRPLS procedure using the GML criterion.

For the indirect GML method, several replications failed to converge (Table 1). The GACV method failed to converge in one replication. None of the other methods has this problem. The boxplots of MSEs are shown in Figures 2 and 3. As Pawitan and O’Sullivan (1994), we compute the relative efficiencies MSE_m/MSE_{DM} for $m = \text{LS, IM, DV}$ and PO . The medians of these relative efficiencies are listed in Table 1. From Figures 2 and 3, the MSEs based on the LS, IM, DV and PO methods have heavier tails than those from the DM method. Thus the mean relative efficiencies can be much bigger than medians (Table 1). We conclude that the direct GML is stable and has the best overall performance. Even though involving a somewhat ad hoc two-step procedure, the PO method performs well. One problem with the PO method is that it sometimes undersmooths the spectrum. The LS method is less efficient. The indirect GML method sometimes fails to converge and performs very badly for the MA4 process. The GACV method occasionally fails to converge due to numerical problems. Its performance is inferior to the direct GML when it converges. This is especially true in the case when n is small where GACV has many large MSEs due to undersmoothing, a phenomenon has been previously observed for the GCV method (Wahba and Wang 1993). We have conducted simulations with other stationary processes and sample sizes. The comparative results remain the same.

4.2 Simulations for Locally Stationary Processes

To investigate the performance of the direct GML and GACV methods for locally stationary processes, and compare them with the LS and indirect GML methods, we simulate locally

| | AR3 | | | | MA4 | | | |
|-----|------|----------|------|------|------|------------|----------|------|
| T | LS | IM | DV | PO | LS | IM | DV | PO |
| 128 | 1.51 | 1.22 (2) | 1.41 | 1.08 | 1.30 | 10.37 (7) | 1.77 | 1.03 |
| | 1.82 | 20.06 | 3.70 | 1.45 | 1.41 | 10.83 | 2.29 | 1.12 |
| 256 | 1.49 | 1.19 (4) | 1.19 | 1.06 | 1.33 | 11.80 (11) | 1.09 (1) | 0.99 |
| | 1.83 | 16.87 | 3.13 | 1.45 | 1.43 | 13.76 | 1.44 | 1.05 |

Table 1: Median (first row for each T) and mean (second row for each T) relative efficiencies. Numbers in the parentheses are the number of replications out of 1000 simulation replications that the indirect GML and GACV methods failed to converge.

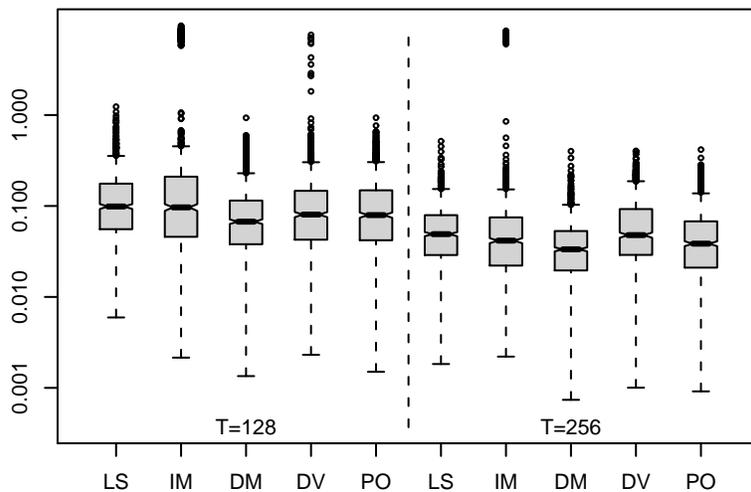


Figure 2: Boxplots of MSEs on logarithm scale for the AR3 process.

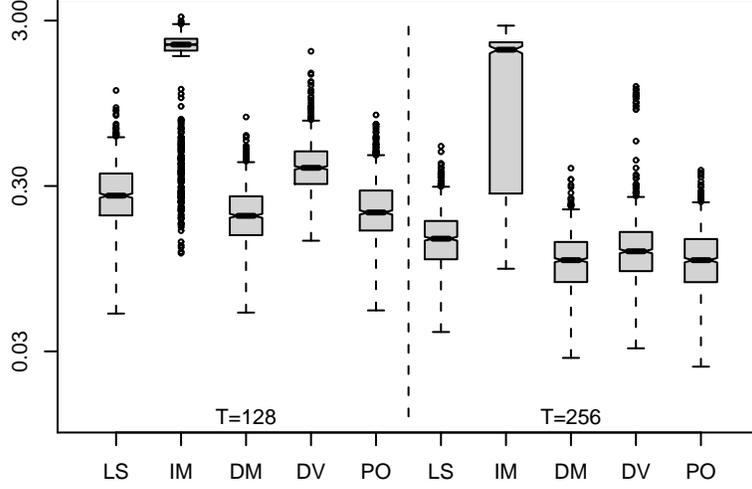


Figure 3: Boxplots of MSEs on logarithm scale for the MA4 process.

stationary time series from two time-varying spectra:

1. LS1: $g(\omega, u) = 4 + \sin(2\pi u) + \log\{1.25 - \cos(2\pi\omega)\}$,
2. LS2: $g(\omega, u) = 5 - 8(\omega - 0.5)^2 + \sin\{2\pi \exp(u)\} + 0.01(\omega - 0.5)^2 \sin\{2\pi \exp(u)\}$.

Similar to Guo and Dai (2006), the time series $X_t, t = 0, \dots, T - 1$, are simulated based on the following relationship:

$$X_t = \sum_{k=0}^{T-1} \exp\{g(k/T, t/T)\} \exp(2\pi kti/T) Z_k, \quad (21)$$

where $Z_k = Z(k/T)$ are mutually independent random variables distributed as complex Normal with mean zero and variance $1/T$. $Z_k = \overline{Z_{T-k}}$ when $k/T \neq 0, 0.5$, or 1 . When $k/T = 0, 0.5$, or 1 , $Z_k = Z(k/T)$ are mutually independent random variables distributed as real Normal with mean zero and variance $1/T$. We consider two sample sizes, $T = 1024$ and $T = 2048$, for each locally stationary process. To assess the impact of block sizes, we consider

three time-frequency grids: $(K, J) = (64, 16)$, $(K, J) = (32, 32)$ and $(K, J) = (16, 64)$. For each setting, we repeat simulation 100 times.

We compute the mean squared error as

$$MSE_m = \sum_{k=1}^K \sum_{j=1}^J (\hat{g}_{kj} - g_{kj})^2 / (KJ),$$

where $m = \text{LS, IM, DM and DV}$ which correspond to the method based on logarithmic transformation, the indirect GML method in Guo et al. (2003) and the direct GML method and the GACV method.

The boxplots of MSEs are shown in Figures 4 and 5. Table 2 lists median relative efficiencies and the number of replications that the indirect GML and GACV methods failed to converge. The comparative results are similar to those in the stationary case: the indirect GML and GACV methods may fail to converge and the direct GML is stable and has the best performance. When converged, the GACV has comparable performance to the direct GML method. However, the GACV method takes much longer to compute. Therefore, the direct GML method is recommended. The estimation is not very sensitive to the choices of block sizes. We have conducted simulations with other sample sizes and block sizes. The comparative results remain the same.

5 The IEEG analysis

Figure 1 shows the IEEGs of 5 minute pre seizure and baseline segments from two channels. These two channels recorded IEEGs from adjacent electrodes in the mesial temporal lobe of the brain believed to be most relevant to the seizure. Therefore, we expect to see close connections in the power and frequency activities between both channels. The important clinical questions are (1) does the characteristic frequencies change over time during the

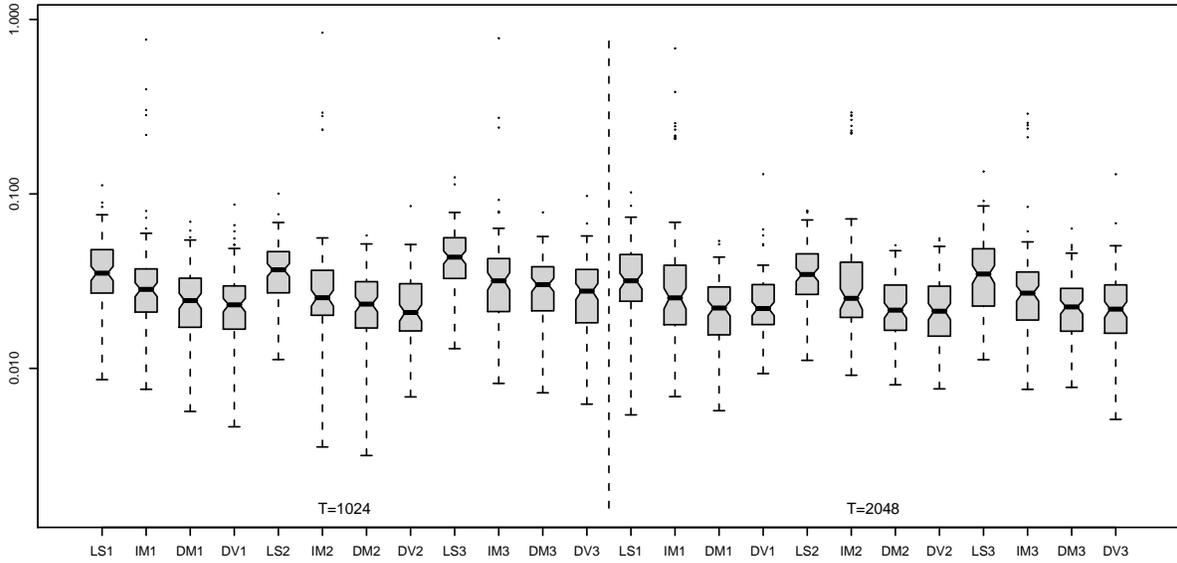


Figure 4: Boxplots of MSEs on logarithm scale for the LS1 process. Numbers in the x-axis labels represent three settings of time-frequency grids: 1 for (64,16), 2 for (32,32) and 3 for (16,64).

pre-ictal stage of the seizure as compared to the baseline in each channel? and (2) are such frequencies common across both channels? We answer these questions by analyzing the IEEGs using our estimation procedure.

We assume that the IEEG segments are locally stationary. To remove artifacts, we first normalize the raw IEEG recordings by subtracting their across-channel means. Each segment is then partitioned into 64 time blocks. 32 equally-spaced frequency points are selected to calculate the local periodograms. Specifically, the time-frequency grids for the calculation are $(\omega_k, u_j) = (k/33, (938 * j - 468.5)/60000)$ for $k = 1, \dots, 32$, $j = 1, \dots, 63$ and $(\omega_k, u_{64}) = (k/33, 0.9925)$ for $k = 1, \dots, 32$. Figure 6 shows the SS ANOVA estimates of time-varying spectra for both the baseline and pre-seizure segments of the two channels.

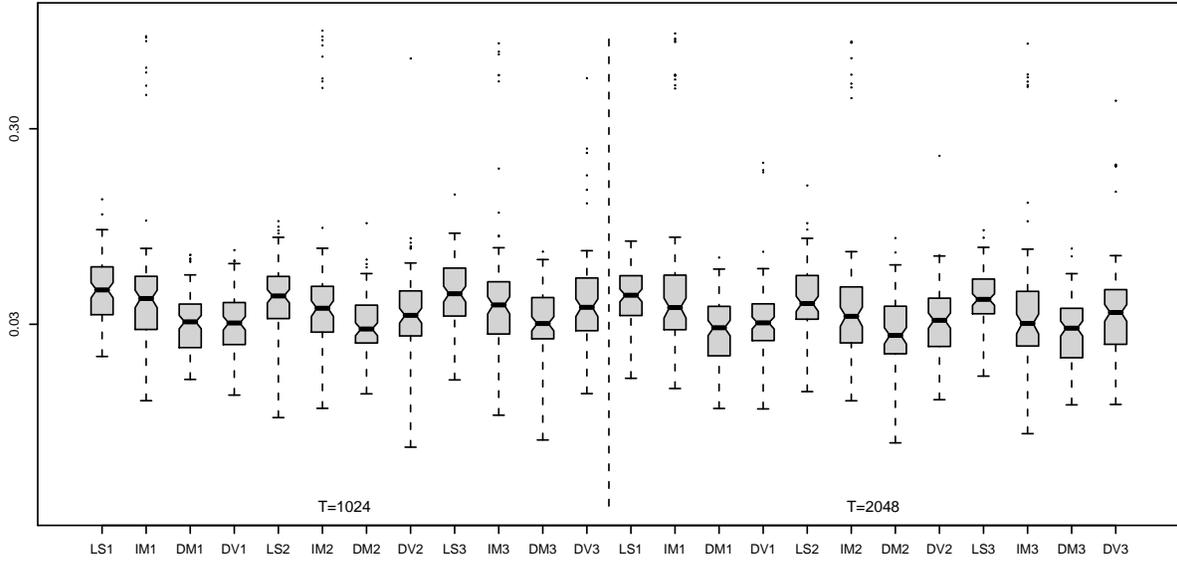


Figure 5: Boxplots of MSEs on logarithm scale for the LS2 process. Numbers in the x-axis labels represent three settings of time-frequency grids: 1 for (64,16), 2 for (32,32) and 3 for (16,64).

The direct GML method is used for all fits in this section. Because the sampling rate is 200 HZ and the spectrum is symmetric around 100 HZ, we can only assess power changes in frequency bands ([0 HZ-100 HZ]). It appears that the baseline spectrum does not vary much over time.

For channel 1, the p-values for testing stationarity based on two statistics are .81 and .73 for the baseline segment and .01 and .01 for the pre-seizure segment. For channel 2, the p-values are .31 and .16 for the baseline segment and 0 and .0067 for the pre-seizure segment. We conclude that for this data set, the processes far away from the seizure's clinical onset can be regarded as stationary while the processes close to the seizure's clinical onset is non-stationary. As expected the pre-ictal spectra have similar shapes, indicating the possible

| | | LS1 | | | LS2 | | |
|------|----------|------|-----------|----------|------|-----------|----------|
| T | (K, J) | LS | IM | DV | LS | IM | DV |
| 1024 | (64,16) | 1.49 | 1.14 (4) | 0.98 | 1.49 | 1.27 (13) | 1.02 (1) |
| | (32,32) | 1.55 | 1.10 (6) | 0.91 | 1.38 | 1.22 (13) | 1.06 (1) |
| | (16,64) | 1.47 | 1.14 (11) | .99 | 1.39 | 1.21 (10) | 1.18 |
| 2048 | (64,16) | 1.48 | 1.17 (6) | 1.05 | 1.50 | 1.33 (12) | 1.07 (1) |
| | (32,32) | 1.51 | 1.13 (5) | 1.12 (1) | 1.53 | 1.28 (8) | 1.18 (1) |
| | (16,64) | 1.56 | 1.19 (3) | 0.95 (1) | 1.53 | 1.16 (10) | 1.26 (1) |

Table 2: Median relative efficiencies. Numbers in the parentheses are the number of replications out of 100 simulation replications that the indirect GML and GACV methods failed to converge.

connectivity between the power evolutions of the two adjacent channels.

To find significant changes of the preseizure power spectra from those of the baseline segments, we compute 95% Bayesian confidence intervals for the estimated preseizure power spectra using the same method described in Wahba et al. (1995). We also compute the difference in estimated power spectra between preseizure and baseline IEEGs. At a particular point of time and frequency, the difference is deemed significant if the estimated baseline power spectrum is outside the 95% confidence interval. Figure 7 shows the contour plots of estimated power spectra differences for time-frequency regions where the differences are significant.

In channel 1, a high frequency power discharge (at [75HZ - 100HZ]) was recorded during approximately -300 to -280 seconds before the seizure; then a power build-up was captured for the next 85 seconds at [10HZ - 40HZ], followed by another significant power decrease at

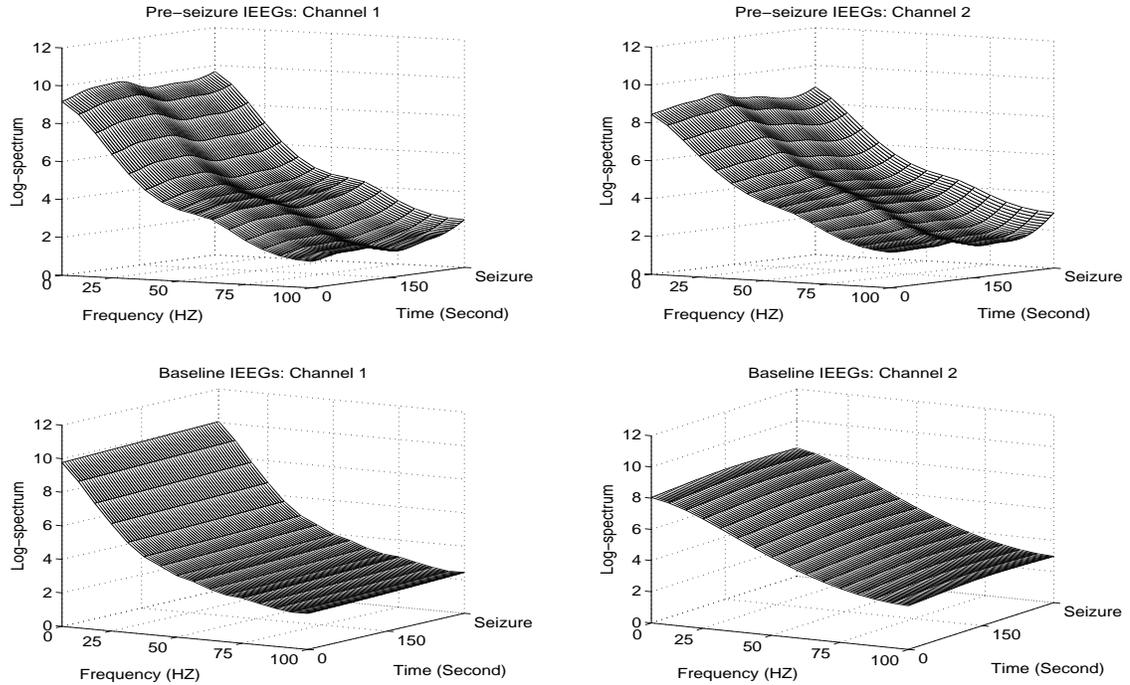


Figure 6: Time-varying spectra estimates (in log scale) for the pre-seizure segments (upper panels) and the baseline segments (lower panels) of the 5 minute IEEG segments from channel 1 (left) and channel 2 (right).

both high and low frequency ends ($[75\text{HZ} - 100\text{HZ}]$ and $< 5\text{HZ}$). However, the channel 2 IEEGs recorded significant power discharges as a broad band activity ($[75\text{HZ} - 100\text{HZ}]$ and $[20\text{HZ} - 40\text{HZ}]$) 5 minutes before the seizure. Interestingly, in this channel power increased around the same time as channel 1 (-270 to -150 seconds), but at lower frequencies ($< 10\text{HZ}$). This phenomenon implies that the short term power build up may be regarded as a warning signal of the coming seizure.

For both channels, the common frequency band for decreased power is within $[75\text{HZ} - 100\text{HZ}]$. While there are no common frequencies at which the power increase occurred, the across-frequency interaction between the two channels appeared during -270 to -150 seconds.

By further studying the within-frequency (at [75HZ - 100HZ]) and across-frequency (between [10HZ - 40HZ] and [0HZ - 10HZ]) signal coupling, the seizure origination and propagation path may be inferred, which would eventually allow preventive action or surgical resection to take place at the right location to prevent a future seizure. Our results conform recent findings in the literature. Specifically, high-frequency oscillations are usual suspects of the electrical activities for the epileptic brain, which may be important markers for epileptic network function ([80HZ - 500HZ]) (Worrell, Parish, Cranstoun, Jonas, Baltuch and Litt 2004), and γ -band ([25HZ - 60HZ]) activity may be associated with a reliable warning of the seizure (Aksenova, Volkovych and Villa 2007).

6 Discussion

We have developed methods and software for smoothing spline and SS ANOVA spectrum estimation of stationary and locally stationary processes. These methods are recommended since they are stable and perform better than other existing methods. The risk estimation method by Pawitan and O'Sullivan (1994) may also be extended to the SS ANOVA spectrum estimation of locally stationary processes. It is unlikely that such an extension will perform better than the direct GML method. For a stationary process, similar permutation tests can be constructed to test if its spectrum is a constant which amounts to a white noise series (Fan and Zhang 2004).

Our analyses have successfully picked up the important time-varying behaviors of the power and frequency components of the IIEEG channels, which may be indicative of the seizure onset and propagation patterns (Worrell et al. 2004, Mormann et al. 2003, Mormann et al. 2000, Schiller, Cascino, Busacker and Sharbrough 1998b). Future analyses would

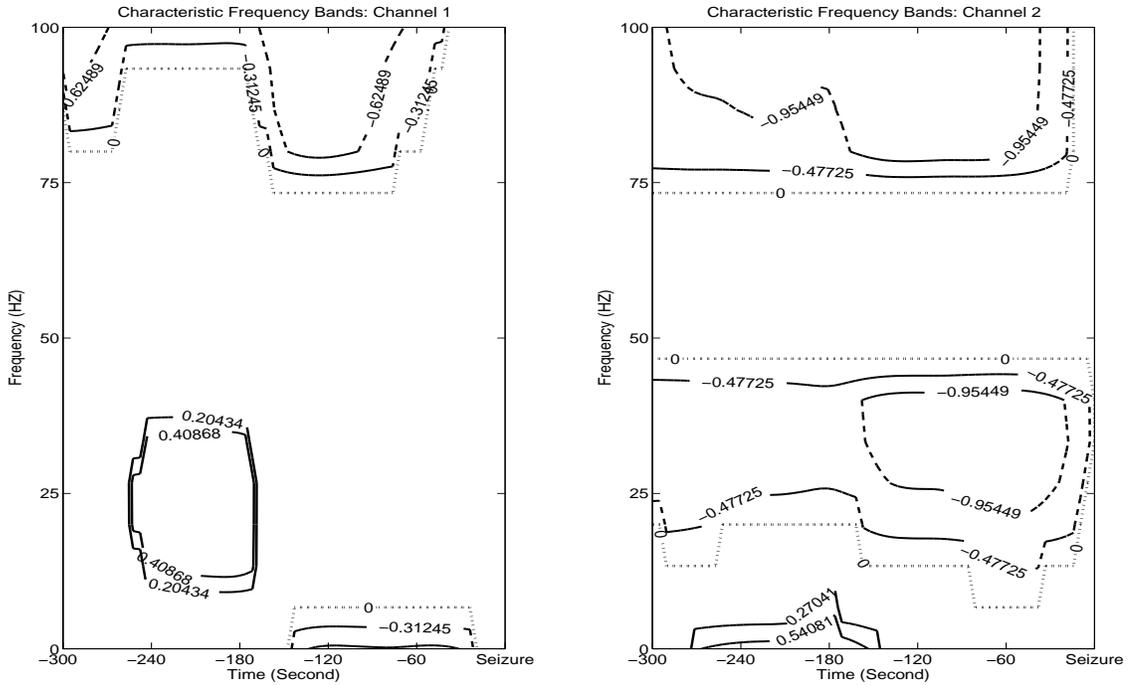


Figure 7: Contour plots for the significant differences in power spectra between the pre-seizure IEEGs and baseline IEEGs (defined by subtracting the spectra estimates of the baseline segments from those of the pre-seizure segments). The positive differences (solid lines) and negative differences (dashed lines) indicate significant increases and significant decreases in power at 5% significance level. For both channels, the common frequency band for power discharge is within [75HZ - 100HZ], occurred as early as 5 minutes before its onset, while the simultaneous power build-up between -270 and -150 seconds happened at different frequencies ([10HZ - 40HZ] for channel 1 and < 10 HZ for channel 2).

include adopting the technique for multiple channels and multiple seizures. With the aid of clinical judgements from neurologists and neurosurgeons, seizure propagation patterns may eventually be uncovered.

Even though we emphasize the application to the spectral analysis of EEG time series,

our methods are general with a wide range of applications including neurological cognition studies. And they may be applied to understand the periodic secretion patterns of hormone time series and assess the spectral properties of the magnetoencephalography (MEG) time series.

Acknowledgments

The authors thank the Editor, the Associate Editor and two references for constructive comments that substantially improved an earlier draft. Qin's research was supported by the Career Development Funds from the Fred Hutchinson Cancer Research Center, and Wang's research was supported by a grant from the National Science Foundation.

References

- Aksenova, T., Volkovych, V. and Villa, A. (2007). Detection of spectral instability in EEG recordings during the preictal period, *Journal of Neural Engineering* **4**: 173–178.
- Brillinger, D. (1981). *Time Series: Data Analysis and Theory*, Holden Day.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes, **25**: 1–37.
- D'Alessandro, M., Vachtsevanos, G., Esteller, R., Echauz, J. and Litt, B. (2001). A generic approach to selecting the optimal feature for epileptic seizure prediction, *IEEE international meeting of the Engineering in Medicine and Biology Society*.
- Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation, *Scand. J. Statist.* **25**(2): 359–369.

- Fan, J. and Zhang, W. (2004). Generalised likelihood ratio tests for spectral density, *Biometrika* **91**(1): 195–209.
- Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum, *Journal of Time Series Analysis* **18**(3): 231–251.
- Gu, C. (1992). Penalized likelihood regression: A Bayesian analysis, *Statistica Sinica* **2**: 255–264.
- Gu, C. (2002). *Smoothing Spline ANOVA Models*, Springer-Verlag, New York.
- Guo, W. and Dai, M. (2006). Multivariate time-dependent spectral analysis using cholesky decomposition, *Statistica Sinica* **16**: 825–845.
- Guo, W., Dai, M., Ombao, H. C. and von Sachs, R. (2003). Smoothing spline ANOVA for time-dependent spectral analysis, *J. Amer. Statist. Assoc.* **98**(463): 643–652.
- Hannig, J. and Lee, T. (2004). Kernel smoothing of periodograms under kllback-leibler discrepancy, *Signal Processing* **84**: 1255–1266.
- Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995). Logspline estimation of a possibly mixed spectral distribution, *J. Time Ser. Anal.* **16**(4): 359–388.
- Lee, T. (1997). A simple span selector for periodogram smoothing, *Biometrika* **84**: 965–969.
- Lin, X., Wahba, G., Xiang, D., Gao, F., Klein, R. and Klein, B. (2000). Smoothing spline ANOVA models for large data sets with bernoulli observations and the randomized gacv, *Annals of Statistics* **28**: 1570–1600.
- Liu, A., Meiring, W. and Wang, Y. (2005). Testing generalized linear models using smoothing spline methods, *Statistica Sinica* **15**: 235–256.

- Liu, A., Tong, T. and Wang, Y. (2006). Smoothing spline estimation of variance functions, *Journal of Computational and Graphical Statistics*.
- Lopes da Silva, F. (1978). Analysis of EEG non-stationarities, *Contemporary Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol*, Cobb W.A., and van Dujn H., Eds. **34 (suppl)**: 165–179.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall, London.
- Mormann, F., Andrzejak, R., Kreutz, T., Rieke, C., David, P., Elger, C. and Lehnertz, K. (2003). Automated detection of a pre-seizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients, *Physical Review E* **67**: 021912 1–10.
- Mormann, F., Lehnertz, K., David, P. and Elger, C. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients, *Physica D* **144**: 358–369.
- Ombao, H. C., Raz, J. A., Strawderman, R. L. and von Sachs, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing, *Biometrika* **88**(4): 1186–1192.
- Ombao, H., Raz, J., von Sachs, R. and Guo, W. (2002). The SLEX model of a nonstationary time series, *Annals of the Institute of Statistical Mathematics*. **54**: 171 – 200.
- Pawitan, Y. and O’Sullivan, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood, *J. Amer. Statist. Assoc.* **89**(426): 600–610.

- Schiller, Y., Cascino, G. D. and Sharbrough, F. W. (1998a). Chronic intracranial EEG monitoring for localizing the epileptogenic zone: An electroclinical correlation, *Epilepsia* **39(12)**: 1302–1308.
- Schiller, Y., Cascino, G. D., Busacker, N. E. and Sharbrough, F. W. (1998b). Characterization and comparison of local onset and remote propagated electrographic seizures recorded with intracranial electrodes, *Epilepsia* **39(4)**: 380–388.
- Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*, New York: Springer-Verlag.
- Wahba, G. (1980). Automatic smoothing of the log periodogram, *J. Amer. Statist. Assoc.* **75**: 122–132.
- Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.
- Wahba, G. and Wang, Y. (1993). Behavior near zero of the distribution of GCV smoothing parameter estimates for splines, *Statistics and Probability Letters* **25**: 105–111.
- Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy, *Annals of Statistics* **23**: 1865–1895.
- Winterhalder, M., Taiwald, T., Voss, H. U., Aschenbrenner-Scheibe, R., Timmer, J. and Schulze-Bonhage, A. (2003). The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods, *Epilepsy and Behavior* **4**: 318–325.

Worrell, G. A., Parish, L., Cranstoun, S. D., Jonas, R., Baltuch, G. and Litt, B. (2004). High-frequency oscillations and seizure generation in neocortical epilepsy, *Brain* **127**: 1496–1506.

Xiang, D. and Wahba, G. (1996). A generalized approximate cross validation for smoothing splines with non-Gaussian data, *Statistica Sinica* **6**: 675–692.

Supplemental Materials

1. Derivation of the direct GML criterion

We provide a sketchy derivation of the direct GML criterion (18) for the SS ANOVA models only. Let $g_i = g(\gamma_i)$ where $\gamma_i = (\omega_i, u_i)$ and $\mathbf{g} = (g_1, \dots, g_n)'$. Let $q(\mathbf{g})$ be the prior density of \mathbf{g} based on (17) as $a \rightarrow \infty$. It can be shown that

$$q(\mathbf{g}) \propto b^{-\frac{n-2}{2}} |\Sigma_\theta|^{-\frac{1}{2}} |S' \Sigma_\theta^{-1} S|^{-\frac{1}{2}} \exp\left(-\frac{1}{2b} \mathbf{g}' B \mathbf{g}\right),$$

where $S = \{1, u_i - .5\}_{i=1}^n$, $\Sigma_\theta = \{\sum_{r=1}^4 \theta_r R_r(\gamma_i, \gamma_j)\}_{i,j=1}^n$, and $B = \Sigma_\theta^{-1} - \Sigma_\theta^{-1} S (S' \Sigma_\theta^{-1} S)^{-1} S' \Sigma_\theta^{-1}$. The marginal density of \mathbf{y} , $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{g})q(\mathbf{g})d\mathbf{g}$, does not have a closed form. The mode of $p(\mathbf{y}|\mathbf{g})q(\mathbf{g})$ approximately equals $\hat{\mathbf{g}} = (\hat{g}_1, \dots, \hat{g}_n)'$ (Gu, 1992). Let $u_k = 1 - y_k \exp(-\hat{g}_k)$, $\mathbf{u}_c = (u_1, \dots, u_n)'$ and $\mathbf{y}_c = \hat{\mathbf{g}} - \mathbf{u}_c$. Expanding $-\log p(\mathbf{y}|\mathbf{g})$ around $\hat{\mathbf{g}}$, we have

$$-\log p(\mathbf{y}|\mathbf{g}) = \frac{1}{2}(\mathbf{g} - \mathbf{y}_c)'(\mathbf{g} - \mathbf{y}_c) + C,$$

where C is a constant. Let $(Q_1 \ Q_2)(R' \ \mathbf{0}')$ be the QR decomposition of S and UDU' be the spectral decomposition of $Q_2' \Sigma_\theta Q_2$ where $D = \text{diag}(\delta_1, \dots, \delta_{n-2})$ and δ_ν are eigenvalues. Let $\mathbf{z} = (z_1, \dots, z_{n-2})' = U' Q_2' \mathbf{y}_c$. Following similar arguments as in Liu, Meiring and Wang (2005), ignoring some constants, the negative log marginal likelihood $-\log p(\mathbf{y})$ can be approximated by the direct GML criterion

$$\text{GML}(\lambda, \boldsymbol{\theta}) = \sum_{i=1}^n \{\hat{g}_i + y_i \exp(-\hat{g}_i)\} - \frac{1}{2} \mathbf{u}_c' \mathbf{u}_c + \frac{1}{2} \sum_{\nu=1}^{n-2} \left\{ \ln(\delta_\nu / n\lambda + 1) + \frac{z_\nu^2}{\delta_\nu / n\lambda + 1} \right\}.$$

2. Derivation of the GACV criterion

We derive the GACV (19) for the SS ANOVA models only. Note that (Guo et al., 2003)

$$y_i \approx \exp(g_i) \chi_{i,d_i}^2 / d_i,$$

where χ_{i,d_i}^2 are independent Chi-square random variables with degree of freedom $d_i = 1$ when $\omega_i = 0$ and $\omega_i = 1/2$ (if T is even) and $d_i = 2$ otherwise. As Pawitan and O'Sullivan (1994) and Guo et al. (2003), we will ignore the slight difference in degrees of freedom. That is, all degrees of freedom are set to two. The approximate Whittle log-likelihood

$$l_i(g_i) = -g_i - y_i \exp(-g_i) + C_1,$$

where C_1 is a constant independent of g_i . Let $\hat{g}(\gamma)$ be the minimizer of the penalized Whittle likelihood (15) and $\hat{g}_i = \hat{g}(\gamma_i)$. The Kullback-Leibler distance

$$\text{KL}(g, \hat{g}) \triangleq \frac{1}{n} \sum_{i=1}^n \text{E}_{g_i} \{l_i(g_i) - l_i(\hat{g}_i)\} = \frac{2}{n} \sum_{i=1}^n \{\exp(g_i - \hat{g}_i) + \hat{g}_i\} + C_2,$$

where C_2 is a constant independent of \hat{g} . Ignoring this constant, the comparative Kullback-Leibler criterion is defined as

$$\text{CKL}(g, \hat{g}) = \frac{2}{n} \sum_{i=1}^n \{\exp(g_i - \hat{g}_i) + \hat{g}_i\}.$$

The CKL criterion cannot be used directly to select the smoothing parameters since it depends on the true log-spectrum. The GACV score in (19) provides a proxy for the CKL criterion.

Let $\hat{g}^{(-i)}$ be the estimate of g without the i th observation y_i . That is, $\hat{g}^{(-i)}$ is the minimizer of the penalized Whittle likelihood

$$\sum_{k \neq i} \{g_k + y_k \exp(-g_k)\} + n \sum_{r=1}^4 \lambda_r \|P_r g\|^2.$$

Let $\hat{g}_k^{(-i)} = \hat{g}^{(-i)}(\gamma_k)$. The leaving-out-one cross-validation estimate of the CKL criterion is

$$\text{CV}(\lambda, \boldsymbol{\theta}) = \frac{2}{n} \sum_{i=1}^n \{y_i \exp(-\hat{g}_i^{(-i)}) + \hat{g}_i\}.$$

The function $\text{CV}(\lambda, \boldsymbol{\theta})$ may be used directly to select the smoothing parameters. However, the computation is usually expensive. GACV provides an approximation of $\text{CV}(\lambda, \boldsymbol{\theta})$. Let $S = (Q_1 \ Q_2)(R' \ \mathbf{0}')$ be the QR decomposition of S . Note that $\mathbf{g} = S\mathbf{d} + \Sigma_\theta \mathbf{c}$ where $\mathbf{d} = (d_1, d_2)'$ and $\mathbf{c} = (c_1, \dots, c_n)'$. It can be shown that $\sum_{r=1}^4 \lambda_r \|P_r \mathbf{g}\| = \lambda \sum_{r=1}^4 \theta_r^{-1} \|P_r \mathbf{g}\| = \lambda \mathbf{c}' \Sigma_\theta \mathbf{c} = \lambda \mathbf{g}' \Omega \mathbf{g}$ where $\Omega = Q_2(Q_2' \Sigma_\theta Q_2)^+ Q_2'$ and $+$ represents the Moore-Penrose generalized inverse. Then the penalized Whittle likelihood (15) can be rewritten as

$$J(\mathbf{g}, \mathbf{y}) = \sum_{i=1}^n \{g_i + y_i \exp(-g_i)\} + n\lambda \mathbf{g}' \Omega \mathbf{g}.$$

Let $W = \text{diag}(y_1 \exp(-\hat{g}_1)/2, \dots, y_n \exp(-\hat{g}_n)/2)$, $V = \text{diag}(\exp(-\hat{g}_1)/2, \dots, \exp(-\hat{g}_n)/2)$, and $H = (W + n\lambda\Omega)^{-1}V$. As in Xiang and Wahba (1996), it can be shown that for any fixed i , $\hat{g}^{(-i)}$ is the minimizer of

$$J(\mathbf{g}, \mathbf{y}^{(-i)}) = \sum_{k \neq i} \{g_k + y_k \exp(-g_k)\} + \{g_i + y_i \exp(-\hat{g}_i^{(-i)})\} + n\lambda \mathbf{g}' \Omega \mathbf{g},$$

where $\mathbf{y}^{(-i)} = (y_1, \dots, y_{i-1}, \exp(\hat{g}_i^{(-i)}), y_{i+1}, \dots, y_n)^T$. Let $\hat{\mathbf{g}}^{(-i)} = (\hat{g}_1^{(-i)}, \dots, \hat{g}_n^{(-i)})^T$. Above result indicates that $\hat{\mathbf{g}}^{(-i)}$ is a minimizer of $J(\mathbf{g}, \mathbf{y}^{(-i)})$. Note that $\hat{\mathbf{g}}$ is a minimizer of $J(\mathbf{g}, \mathbf{y})$.

Therefore,

$$\frac{\partial J}{\partial \mathbf{g}}(\hat{\mathbf{g}}, \mathbf{y}) = \frac{\partial J}{\partial \mathbf{g}}(\hat{\mathbf{g}}^{(-i)}, \mathbf{y}^{(-i)}) = 0.$$

It is easy to check that $\partial^2 J / \partial \mathbf{g} \partial \mathbf{g}' = W + n\lambda\Omega$ and $\partial^2 J / \partial \mathbf{y} \partial \mathbf{g}' = -V$. Expanding $(\partial J / \partial \mathbf{g})(\hat{\mathbf{g}}^{(-i)}, \mathbf{y}^{(-i)})$ at the point $(\hat{\mathbf{g}}, \mathbf{y})$, we have

$$\begin{aligned} 0 &= \frac{\partial J}{\partial \mathbf{g}}(\hat{\mathbf{g}}^{(-i)}, \mathbf{y}^{(-i)}) \\ &= \frac{\partial J}{\partial \mathbf{g}}(\hat{\mathbf{g}}, \mathbf{y}) + \frac{\partial^2 J}{\partial \mathbf{g} \partial \mathbf{g}'}(\hat{\mathbf{g}}^*, \mathbf{y}^*)(\hat{\mathbf{g}}^{(-i)} - \hat{\mathbf{g}}) + \frac{\partial^2 J}{\partial \mathbf{y} \partial \mathbf{g}'}(\hat{\mathbf{g}}^*, \mathbf{y}^*)(\mathbf{y}^{(-i)} - \mathbf{y}). \end{aligned}$$

Therefore,

$$\hat{\mathbf{g}}^{(-i)} - \hat{\mathbf{g}} \approx (W + n\lambda\Omega)^{-1}V(\mathbf{y}^{(-i)} - \mathbf{y}) = H(\mathbf{y}^{(-i)} - \mathbf{y}).$$

Note that all elements of $\mathbf{y}^{(-i)} - \mathbf{y}$ are zero except the i th element which equals $\exp(\hat{g}_i^{(-i)}) - y_i$.

Thus we have

$$\frac{\hat{g}_i - \hat{g}_i^{(-i)}}{y_i - \exp(\hat{g}_i^{(-i)})} \approx h_{ii},$$

where h_{ii} is the i th diagonal element of H . Let $L(\lambda, \boldsymbol{\theta}) = \sum_{i=1}^n \{y_i \exp(-\hat{g}_i) + \hat{g}_i\}$. Then

$$\begin{aligned} CV(\lambda, \boldsymbol{\theta}) &= L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n y_i \left\{ \exp(-\hat{g}_i^{(-i)}) - \exp(-\hat{g}_i) \right\} \\ &\approx L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n y_i \exp(-\hat{g}_i) (\hat{g}_i - \hat{g}_i^{(-i)}) \\ &= L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n y_i \exp(\hat{g}_i) \frac{\hat{g}_i - \hat{g}_i^{(-i)}}{y_i - \exp(\hat{g}_i^{(-i)})} \frac{y_i - \exp(\hat{g}_i)}{1 - \frac{\exp(\hat{g}_i) - \exp(\hat{g}_i^{(-i)})}{y_i - \exp(\hat{g}_i^{(-i)})}} \\ &\approx L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n y_i \exp(-\hat{g}_i) \frac{\hat{g}_i - \hat{g}_i^{(-i)}}{y_i - \exp(\hat{g}_i^{(-i)})} \frac{y_i - \exp(\hat{g}_i)}{1 - \exp(\hat{g}_i) \frac{\hat{g}_i - \hat{g}_i^{(-i)}}{y_i - \exp(\hat{g}_i^{(-i)})}} \\ &= L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n \frac{y_i \exp(-\hat{g}_i) (y_i - \exp(\hat{g}_i))}{\frac{y_i - \exp(\hat{g}_i^{(-i)})}{\hat{g}_i - \hat{g}_i^{(-i)}} - \exp(\hat{g}_i)} \\ &\approx L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n \frac{y_i \exp(-\hat{g}_i) (y_i - \exp(\hat{g}_i))}{\frac{1}{h_{ii}} - \exp(\hat{g}_i)} \\ &\approx L(\lambda, \boldsymbol{\theta}) + \sum_{i=1}^n \frac{h_{ii} y_i \exp(-\hat{g}_i) (y_i - \exp(\hat{g}_i))}{1 - h_{ii} \exp(\hat{g}_i)}. \end{aligned}$$

Replacing $h_{ii} \exp(\hat{g}_i)$ in the denominator by $\text{tr}(W_0^{1/2} H W_0^{1/2})/n$ and h_{ii} in the numerator by $\text{tr}(H)/n$ where $W_0 = \text{diag}(\exp(\hat{g}_1), \dots, \exp(\hat{g}_n))$, we obtain the GACV criterion in (19).

3. Simulations for Permutation Tests of Stationarity

To evaluate the performance of the permutation tests proposed in Section 3.4, we simulate locally stationary time series from the following spectra

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

with $\beta = 0, .25, .5, .75, 1$ and 1.25 . To save computational time, we compute local periodograms on time-frequency grid $(K, J) = (16, 16)$. For each simulation sample, we compute three statistics and their corresponding null distributions based on 100 permutations. We repeat simulation 500 times for each setting. The powers at level .05 are listed in Table 3. It is clear that all three tests hold their levels correctly. The power increases quickly as the magnitude of the time component increases. The statistic S_2 is slightly more powerful.

| | $\beta = 0$ | $\beta = .25$ | $\beta = .5$ | $\beta = .75$ | $\beta = 1$ | $\beta = 1.25$ |
|-------|-------------|---------------|--------------|---------------|-------------|----------------|
| S_1 | .042 | .242 | .590 | .832 | .928 | .958 |
| S_2 | .034 | .270 | .620 | .870 | .966 | .996 |

Table 3: Powers of the permutation tests.

4. R code

We have developed a user friendly R function, *spest.r*, for spectrum estimation. It uses a periodic spline to estimate the spectrum of a stationary process and an SS ANOVA model to estimate the time-varying spectrum of a locally stationary process. A typical call to the function is

```
spest(x, method, process, freqs, times, control)
```

where `x` inputs the time series, `method` specifies a method for estimation of the smoothing parameter(s), `process` specifies if the time series is stationary or locally stationary, `freqs` and `times` specify time-frequency grid for the calculation of local periodograms for a locally stationary process (which handles none-equal-length time intervals), and `control` provides options for various parameters. For example, the estimate in the left panel of Figure 6 is computed by

```
spest(preseizure, method="DGML", process="LS",  
      freqs=1:32/33, times=(seq(from=469.5, by=938, length=64))/60000),
```

where “preseizure” corresponds to the 5 minute preseizure segment.