



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Statistical Planning and Inference

journal homepage: [www.elsevier.com/locate/jspi](http://www.elsevier.com/locate/jspi)

## Smoothing spline ANOPOW

David S. Stoffer<sup>a,\*</sup>, Sangdae Han<sup>b,1</sup>, Li Qin<sup>c,2</sup>, Wensheng Guo<sup>d,3</sup><sup>a</sup> Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA<sup>b</sup> Samsung Marine and Fire Insurance, Seoul, Korea<sup>c</sup> Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center, Seattle, WA, USA<sup>d</sup> Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA

## ARTICLE INFO

## Keywords:

Analysis of power  
 Bootstrap simultaneous confidence intervals  
 Generalized linear model  
 Penalized least squares  
 Reproducing kernel Hilbert space  
 Spectral density  
 Time series

## ABSTRACT

This paper is motivated by the pioneering work of Emanuel Parzen wherein he advanced the estimation of (spectral) densities via kernel smoothing and established the role of reproducing kernel Hilbert spaces (RKHS) in field of time series analysis. Here, we consider analysis of power (ANOPOW) for replicated time series collected in an experimental design where the main goals are to estimate, and to detect differences among, group spectra. To accomplish these goals, we obtain smooth estimators of the group spectra by assuming that each spectral density is in some RKHS; we then apply penalized least squares in a smoothing spline ANOPOW. For inference, we obtain simultaneous confidence intervals for the estimated group spectra via bootstrapping.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

The seminal publications Parzen (1961a, 1961b) and Parzen (1962), wherein Professor Parzen introduced the role of reproducing kernel Hilbert spaces (RKHS) in the analysis of time series and advanced the estimation of (spectral) densities via kernel smoothing (see also Rosenblatt, 1956) are the motivation for this paper. Here, we consider analysis of power (ANOPOW) for replicated time series collected in experimental designs wherein the main goal is to detect differences among group spectra.

For one-way ANOPOW, we suppose we have  $n_\ell > 1$ , for  $\ell = 1, \dots, L$ , stationary time series that are observations in group  $\ell$ . We denote the observations as  $Y_{ij}(t)$  for  $\ell = 1, \dots, L, j = 1, \dots, n_\ell$  and  $t = 1, \dots, n$ ; in addition, we define the corresponding group indicators  $x_{i,tj} = 1$  if  $i = \ell$  and 0 otherwise. We do not assume the series are in phase, but we assume the time series in each group are replications from a process with spectral density  $f_\ell(\omega)$ . The main problem is to determine whether the group spectra are equal, and if they are not, determine the location of the differences. To this end, we assume that the spectral density of  $Y_{ij}(t)$  can be written as

$$f_\ell(\omega) = \exp[\beta_0(\omega) + \beta_1(\omega)x_{1,tj} + \beta_2(\omega)x_{2,tj} + \dots + \beta_{L-1}(\omega)x_{L-1,tj}] \quad (1)$$

\* Corresponding author.

E-mail addresses: [stoffer@pitt.edu](mailto:stoffer@pitt.edu) (D.S. Stoffer), [lqin@scharp.org](mailto:lqin@scharp.org) (L. Qin), [wguo@cceb.upenn.edu](mailto:wguo@cceb.upenn.edu) (W. Guo).<sup>1</sup> The work of these authors was supported, in part, by Grants DMS-0706723 and DMS-0805050 (both to Stoffer) from the National Science Foundation.<sup>2</sup> The work of the author was supported, in part, by a career development grant from the Fred Hutchinson Cancer Research Center.<sup>3</sup> The work of the author was supported, in part, by Grant CA84438 from the National Institutes of Health.

where  $\{\beta_0(\omega), \dots, \beta_{L-1}(\omega)\}$  is a collection of  $L$  group parameter functions. Note that the only restriction imposed by (1) on the group spectra is that  $f_\ell(\omega) > 0$ . In addition, the  $\beta$ -functions have a meaningful interpretation in that

$$\log \frac{f_{\ell_1}(\omega)}{f_{\ell_2}(\omega)} = \beta_{\ell_1}(\omega) - \beta_{\ell_2}(\omega) \quad (2)$$

for  $\ell_1, \ell_2 = 1, \dots, L$ , where we take  $\beta_L(\omega) \equiv 0$ . Thus, for example, if  $\beta_{\ell_1}(\omega) = \beta_{\ell_2}(\omega)$  for all  $\omega$ , then the spectral densities of those particular groups are equal. If there are frequencies for which  $\beta_{\ell_1}(\omega) > \beta_{\ell_2}(\omega)$ , then  $f_{\ell_1}(\omega) > f_{\ell_2}(\omega)$  over those frequencies and vice versa.

The problem of testing for simultaneous group equality becomes the problem of testing the null hypothesis  $H_0: \beta_1(\omega) = \dots = \beta_{L-1}(\omega) \equiv 0$ , and if this hypothesis is rejected, the problem becomes identifying the frequencies for which the individual  $\beta$ -functions are different. Rather than performing a test of hypothesis under the null assumption, we prefer to estimate the individual group spectra and determine whether or not there is evidence that at least one of the  $\beta$ -functions differ from zero over a range of frequencies. Hence, the essential part of the problem is estimating the group spectra under the stated assumptions and we address that in the next section. Finally, we mention that this problem may be easily extended to higher-way ANOPOW, e.g., block designs, in an obvious analogy to the analysis of variance (ANOVA).

The general estimation method is motivated by the following considerations and has been used by others (e.g., Wahba, 1980). Suppose we observe  $n$  observations from a stationary time series  $Y(t)$  with spectral density  $f(\omega)$ . Let  $I(\omega_k)$  denote the periodogram of  $Y(t)$ , i.e.,

$$I(\omega_k) = n^{-1} \left| \sum_{t=1}^n Y(t) \exp\left(-2\pi i t \frac{k}{n}\right) \right|^2 \quad (3)$$

at Fourier frequency  $\omega_k = k/n$ . Then, under general conditions (e.g., Shumway and Stoffer, 2006, Appendix C), the periodogram,  $I(\omega_k)$ , is distributed asymptotically as  $f(\omega_k)$  times a random variable with a Gamma distribution with shape and scale parameters both equal to 1; in addition,  $I(\omega_{k_1})$  and  $I(\omega_{k_2})$ , for  $\omega_{k_1} \neq \omega_{k_2}$ , are approximately independent for  $n$  sufficiently large. Based on these results we may write to a good approximation

$$I(\omega_k) = f(\omega_k) U_k$$

or

$$\log I(\omega_k) = \log f(\omega_k) + \eta_k \quad (4)$$

for  $k = 1, 2, \dots, [(n-1)/2]$ , where  $U_k \stackrel{iid}{\sim} \text{Gamma}(1,1)$ , the Gamma distribution with scale and shape parameters equal to unity. In this case,  $\eta_k = \log U_k$  so that  $E(-\eta_k) = \gamma \simeq 0.57721$  (Euler's constant) and  $\text{Var}(\eta_k) = \pi^2/6$ . Wahba (1980) used this model to obtain what she refers to as an optimally smoothed spline estimate of the logged spectral density.

In the case of ANOPOW, given a time series  $Y_{ij}(t)$  from group  $\ell$ , under assumption (1),

$$\log f_\ell(\omega_k) = \beta_0(\omega_k) + \beta_1(\omega_k) X_{1,\ell,j} + \dots + \beta_{L-1}(\omega_k) X_{L-1,\ell,j} \quad (5)$$

so that if  $I_{ij}(\omega_k)$  is the periodogram of  $Y_{ij}(t)$ , we have the generalized linear model (GLM)

$$\log I_{ij}(\omega_k) = \beta_0(\omega_k) + \beta_1(\omega_k) X_{1,\ell,j} + \dots + \beta_{L-1}(\omega_k) X_{L-1,\ell,j} + \eta_{ij,k} \quad (6)$$

where the  $\eta_{ij,k}$  can be considered as being independent and identically distributed (iid) as the log of Gamma(1,1) random variables.

An extensive discussion of some of the approaches to spectral regression and ANOPOW can be found in Shumway and Stoffer (2006, Chapter 7). In addition, Diggle and al Wasel (1997) present some approaches to the ANOPOW problem. The methods discussed in the aforementioned references fall short of being generally applicable in many ways. The typical shortcomings include the use of averaged periodograms, which implies the series are in phase or can be aligned (e.g., Shumway, 1971; Brillinger, 1980), the employment of frequency-by-frequency GLM analysis that ignores the fact that the spectral density is smooth whereas the periodogram is rough (e.g., Diggle and al Wasel, 1997), or the use of somewhat questionable parametric assumptions on the group spectra to ease the analysis (e.g., Diggle and al Wasel, 1997). With regard to the last shortcoming, Diggle and al Wasel (1997) proposed a model wherein the spectrum of group  $\ell$  has the form

$$\log f_\ell(\omega_k) = \sum_{j=1}^p d_{\ell kj} \beta_j \quad (7)$$

where  $p$  is arbitrary, the  $d$ 's are specified and the  $\beta$ 's are parameters to be estimated. While (7) looks similar to (5), the  $\beta_j$  in (7) do not depend on frequency or group, which are the two main interests in an ANOPOW. A model that is similar to ours for one-way ANOPOW can be found in Fokianos and Savvides (2008), wherein to induce smoothness, the  $\beta_j(\omega)$  are constrained to be functions of sinusoids in a similar fashion to the exponential model for the log-spectrum of a stationary process that was introduced by Bloomfield (1973) and later used in Parzen (1993) to propose a goodness-of-fit test.

To overcome these problems, we introduce the idea of analysis of power via penalized least squares in reproducing kernel Hilbert spaces.

## 2. Penalized least squares in RKHS

We begin with the assumption that  $f$  is in some RKHS, that is, a Hilbert space,  $\mathcal{H}$ , of functions in which all the point evaluations are bounded; e.g., see Aronszajn (1950), Gu and Wahba (1993b), or Wahba (1990). Let  $\mathcal{H}$  be some RKHS of real-valued functions of  $\mathbf{t} = (t_1, t_2, \dots, t_d)$  and  $\mathbf{t} \in \mathcal{T} = \mathcal{T}^{(1)} \otimes \mathcal{T}^{(2)} \otimes \dots \otimes \mathcal{T}^{(d)}$ , where  $t_a$  is the  $a$ th variable in  $\mathcal{T}^{(a)}$  and  $\mathcal{T}^{(a)}$  is some measurable space. Now we construct a probability measure  $d\mu_a$  on  $\mathcal{T}^{(a)}$  for each  $a = 1, 2, \dots, d$  with the symbol  $(A_a f)(\mathbf{t})$ , defined by

$$(A_a f)(\mathbf{t}) = \int_{\mathcal{T}^{(a)}} f(t_1, \dots, t_d) d\mu_a(t_a) \tag{8}$$

is well defined and finite and for every  $f \in \mathcal{H}$  and  $\mathbf{t} \in \mathcal{T}$ .

We can consider  $A_a$  as an operator from  $\mathcal{H}$  to  $\mathcal{H}$  and the decomposition of the identity,  $I$ , as

$$\begin{aligned} I &= \prod_a [A_a + (I - A_a)] \\ &= \prod_a A_a + \sum_a (I - A_a) \prod_{b \neq a} A_b \\ &\quad + \sum_{a < b} (I - A_a)(I - A_b) \prod_{c \neq a, b} A_c + \dots + \prod_a (I - A_a) \end{aligned} \tag{9}$$

In general, as discussed in Gu and Wahba (1993b),  $f$  has an unique representation of the SS-ANOVA form

$$f(\mathbf{t}) = C + \sum_a f_a(t_a) + \sum_{a < b} f_{ab}(t_a, t_b) + \sum_{a < b < c} f_{abc}(t_a, t_b, t_c) + \dots \tag{10}$$

where  $C = (\prod_a A_a)f$  is the mean,  $f_a = [(I - A_a) \prod_{b \neq a} A_b]f$  are the main effects, and  $f_{ab} = [(I - A_a)(I - A_b) \prod_{c \neq a, b} A_c]f$  are the two-factor interactions in (9), and so on.

In this paper we work with the case  $d = 1$ , but we note that Guo et al. (2003) and Qin et al. (2009) have used the case  $d = 2$  to establish a time–frequency functional model. In our case, using Taylor’s theorem, if  $f$  is a real valued function on  $[0, 1]$  with  $f', f'', \dots, f^{(\alpha-1)}$  continuous derivatives and  $f^{(\alpha)} \in \mathcal{L}_2[0, 1]$ , then  $f$  can be expressed as

$$f(x) = \sum_{v=0}^{\alpha-1} \frac{x^v}{v!} f^{(v)}(0) + \int_0^1 \frac{(x-u)_+^{\alpha-1}}{(\alpha-1)!} f^{(\alpha)}(u) du \tag{11}$$

where  $(x)_+ = x$  if  $x \geq 0$  and  $(x)_+ = 0$  if  $x < 0$ .

Now, we define the Sobolev Hilbert space (e.g., see Adams, 1975)  $\mathcal{H}_\alpha$  given by  $\mathcal{H}_\alpha[0, 1] = \{f : f, f', f'', \dots, f^{(\alpha-1)} \text{ absolutely continuous and } f^{(\alpha)} \in \mathcal{L}_2\}$ . Then, each function  $f$  in  $\mathcal{H}_\alpha$  has a Taylor series expansion (11) to order  $\alpha$  and has a decomposition as  $f = f_0 + f_1$  with  $f_0 \in \mathcal{H}_0$  and  $f_1 \in \mathcal{H}_1$  given by the first and the second terms in (11). Moreover,  $\int_0^1 [(D^\alpha f_0)(u)]^2 du = 0$  and  $\sum_{v=0}^{\alpha-1} [(D^v f_1)(0)]^2 = 0$  where  $D^v$  denotes the  $v$ -th derivative of function; e.g., see Wahba (1990). Therefore, we can represent  $\mathcal{H}_\alpha$  as the direct sum,

$$\mathcal{H}_\alpha = \mathcal{H}_0 \oplus \mathcal{H}_1 \tag{12}$$

Consequently, the square norm of  $f \in \mathcal{H}_\alpha$  is given by

$$\|f\|^2 = \sum_{v=0}^{\alpha-1} [(D^v f)(0)]^2 + \int_0^1 [(D^\alpha f)(u)]^2 du \tag{13}$$

implying that the subspaces  $\mathcal{H}_0$  and  $\mathcal{H}_1$  of  $\mathcal{H}_\alpha$  are perpendicular.

The basis of our estimation technique is that of penalized least squares, which can be described for a single series as follows. In view of (4), let  $y_k = \log l(\omega_k)$  be the logged periodogram of the observed series and let  $g_k = \log f(\omega_k)$  be the corresponding logged spectral density. We establish a penalty functional  $\int_0^1 [g^\alpha(u)]^2 = \|P_1 g\|_{\mathcal{H}_\alpha}^2$  where  $P_1$  is the orthogonal projection of  $g$  onto  $\mathcal{H}_1$  and estimate  $g$  by penalized least squares, wherein we minimize

$$\frac{1}{m} \sum_{k=1}^m (y_k - g_k)^2 + \lambda \|P_1 g\|_{\mathcal{H}_\alpha}^2 \tag{14}$$

where  $m = [(n-1)/2]$  and  $\lambda$  is a smoothing parameter. In the next section we will use these ideas to form a smoothed estimate of group spectra specified in (1).

## 3. Estimation

As in Section 1, we focus on one-way ANOPOW; the analysis of higher-way models follows easily and directly from this case. Recall that we observe stationary time series  $Y_{tj}(t)$  for  $t=1, \dots, n$  where there are  $j = 1, \dots, n_\ell$  replications in  $\ell = 1, \dots, L$

groups. We assume that the corresponding group spectra,  $f_\ell(\omega)$ , are strictly positive and write the logged spectra as

$$g_\ell(\omega) \equiv \log f_\ell(\omega) = \beta_0(\omega) + \beta_\ell(\omega) \tag{15}$$

where we put  $\beta_L(\omega) = 0$ . Our goal is to estimate the  $\beta$ -functions from the replicated series.

Given the replicated time series,  $Y_{ij}(t)$ , let  $\log(I_{ij}(\omega_k))$  denote the corresponding logged periodogram [cf. (3)], and write

$$y_{ij}(\omega_k) = \log(I_{ij}(\omega_k)) + \gamma \tag{16}$$

where  $\gamma$  is Euler's constant for  $k=1, \dots, m = \lfloor (n-1)/2 \rfloor$ ; we avoid working with the frequencies  $\omega_k = 0, \frac{1}{2}$ .

Then, as discussion in Section 1, the  $y_{ij}(\omega_k)$  can be modeled approximately as

$$y_{ij}(\omega_k) = g_\ell(\omega_k) + \varepsilon_{ij,k} = \mathbf{x}_\ell' \boldsymbol{\beta}(\omega_k) + \varepsilon_{ij,k} \tag{17}$$

where  $\mathbf{x}_\ell = (x_{0,ij} = 1, x_{1,ij}, \dots, x_{L-1,ij})'$  is the  $L \times 1$  design vector (recall that  $x_{p,ij}$  is essentially a group indicator and  $x_{p,ij} = 1$  if  $p = \ell$  and 0 otherwise, for  $p=0, \dots, L-1, j = 1, \dots, n_\ell$  and  $\ell = 1, \dots, L-1$ ),  $\boldsymbol{\beta}(\omega_k) = (\beta_0(\omega_k), \beta_1(\omega_k), \dots, \beta_{L-1}(\omega_k))'$  is the  $L \times 1$  vector of parameters at Fourier frequency  $\omega_k = k/n$ , and the  $\varepsilon_{ij,k}$  are asymptotically independent with mean 0 and variance  $\pi^2/6$ . Note that the  $\boldsymbol{\beta}(\omega_k)$  are periodic functions with period 1, i.e.,  $\boldsymbol{\beta}(\omega_k) = \boldsymbol{\beta}(\omega_k + 1)$ , and are symmetric around  $\omega_k = 0.5$  i.e.,  $\boldsymbol{\beta}(\omega_k) = \boldsymbol{\beta}(1 - \omega_k)$  in the frequency domain. For convenience, we only use half of the frequency domain in (17) to be free of the periodic constraints.

We may write the model (17) in a more convenient form. First, we define

$$\mathbf{y}_{ij} = \begin{pmatrix} y_{ij}(\omega_1) \\ \vdots \\ y_{ij}(\omega_m) \end{pmatrix}, \quad \boldsymbol{\beta}_p(\Omega) = \begin{pmatrix} \beta_p(\omega_1) \\ \vdots \\ \beta_p(\omega_m) \end{pmatrix}$$

and  $D_{p\ell} = \text{diag}\{x_{p,\ell j}, \dots, x_{p,\ell j}\}$ , for  $p=0, 1, \dots, L-1$ . Then, the model can be written as

$$\mathbf{y}_{ij} = \sum_{p=0}^{L-1} D_{p\ell} \boldsymbol{\beta}_p(\Omega) + \varepsilon_{ij} \tag{18}$$

where  $\varepsilon_{ij} = (\varepsilon_{ij,1}, \dots, \varepsilon_{ij,m})'$ , for  $\ell = 1, \dots, L$  and  $j = 1, \dots, n_\ell$ .

We now use the penalized least squares (PLS) method to obtain smoothed estimates of the  $\beta$ -functions along the lines suggested in Section 2. To this end, we first stack the data vectors,  $\mathbf{y}_{ij}$ , and the corresponding design matrices,  $D_{p\ell}$ , as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_{1,1} \\ \vdots \\ \mathbf{y}_{1,n_1} \\ \vdots \\ \mathbf{y}_{L,1} \\ \vdots \\ \mathbf{y}_{L,n_L} \end{pmatrix}, \quad D_p = \begin{pmatrix} D_{p1} \\ \vdots \\ D_{p1} \\ \vdots \\ D_{pL} \\ \vdots \\ D_{pL} \end{pmatrix}$$

so that the model can be written in the simple form

$$\mathbf{Y} = \sum_{p=0}^{L-1} D_p \boldsymbol{\beta}_p(\Omega) + \boldsymbol{\varepsilon} \tag{19}$$

where  $\boldsymbol{\varepsilon}$  is the corresponding error vector.

The PLS method involves minimizing

$$\frac{1}{N} \left| \mathbf{Y} - \sum_{p=0}^{L-1} D_p \boldsymbol{\beta}_p(\Omega) \right|^2 + \sum_{p=0}^{L-1} \frac{\lambda}{\theta_p} |P_p \boldsymbol{\beta}_p|^2 \tag{20}$$

where  $N=Lm$ ,  $\lambda$  is the main smoothing parameter,  $\theta_p$ 's are subsidiary smoothing parameters, and  $P_p$  represents the orthogonal projections onto the RKHS in the smoothing spline ANOVA model.

Conditional on the smoothing parameters  $\lambda$  and  $\theta_p$ , Gu and Wahba (1993a) give the estimates that minimize (20) as follows:

$$\hat{\boldsymbol{\beta}}_p(\Omega) = U_p \mathbf{h}_p + Q_p \mathbf{c}_p \tag{21}$$

where  $U_p = (1, \omega_k)_{k=1}^m$ ,  $Q_p = \theta_p \mathbf{R}_p(\Omega, \Omega)$ , and  $\mathbf{R}_p(\Omega, \Omega) = \{R_p(\omega_k, \omega_{k'})\}_{k,k'=1}^m$ , where  $R_p(\omega_k, \omega_{k'}) = -k_4(|\omega_j - \omega_{j'}|)$ , where  $k_4(x) = B_4(x)/4!$  and  $B_4(x) = x^4 - 2x^3 + x^2 - 1/30$  is the fourth-order Bernoulli polynomial. The  $\mathbf{h}_p$  and  $\mathbf{c}_p$  vectors can be obtained as follows. Let  $D = \{D_0, \dots, D_{L-1}\}$ ,  $U = \text{diag}\{U_0, \dots, U_{L-1}\}$ ,  $H = \{h'_0, \dots, h'_{L-1}\}'$ ,  $Q = \text{diag}\{Q_0, \dots, Q_{L-1}\}$  and  $C = \{c'_0, \dots, c'_{L-1}\}'$ , then the PLS estimate of  $\boldsymbol{\beta}(\Omega)$  is given by

$$\hat{\boldsymbol{\beta}}(\Omega) = UH + QC \tag{22}$$

where

$$N^{-1}|\mathbf{Y}-DUH-DQC|^2 + \lambda C'QC \tag{23}$$

$$H = \{U'D'M^{-1}DU\}^{-1}U'D'M^{-1}\mathbf{Y} \tag{24}$$

and

$$C = DM^{-1}\{I-DU(U'D'M^{-1}DU)^{-1}U'D'M^{-1}\}\mathbf{Y} \tag{25}$$

where  $M = DQD' + N\lambda I$ .

When we only need an estimate for a specific  $\omega_0$ , we can also use the PLSE at any given frequency, say  $\omega_0$ , from Eq. (21):

$$\hat{\beta}_p(\omega_0) = U_{p0}h_p + Q_{p0}c_p \tag{26}$$

where  $U_{p0} = (1, \omega_0)$ ,  $Q_{p0} = \theta_p \mathbf{R}_p(\omega_0, \boldsymbol{\Omega})$  and  $\mathbf{R}_p(\omega_0, \boldsymbol{\Omega}) = \{R_p(\omega_0, \omega_k)\}_{k=1}^m$ , for  $p = 0, \dots, L-1$ ; see also Guo et al. (2003) and Wahba (1990). The consistency of PLSE for a single log-spectrum has been proved by Guo et al. (2003).

The smoothing parameters,  $\lambda$  and  $\theta_p$ , can be chosen by generalized cross validation or by generalized maximum likelihood (GML) as discussed in Wahba (1985). We focus on the GML criterion, which following Wahba (1985) using our notation, reduces to minimizing

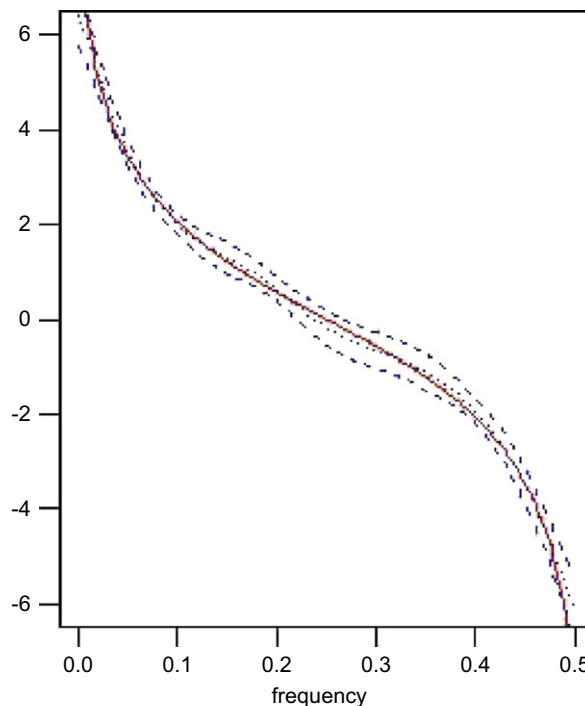
$$V(\lambda, \theta_p) = \frac{\mathbf{Y}'F_2(F_2'DQD'F_2 + N\lambda I)^{-1}F_2'\mathbf{Y}}{[\det(F_2'DQD'F_2 + N\lambda I)]^{1/(N-4)}} \tag{27}$$

where  $F_2$  is given by the QR decomposition of  $DU$ :  $DU = (F_1, F_2)\begin{pmatrix} G \\ 0 \end{pmatrix}$  where  $(F_1, F_2)$  is orthogonal and  $G$  is upper triangular; the dimension of  $F_1$  is  $N \times (2L-1)$  and the dimension of  $F_2$  is  $N \times 1$ .

In the next section we discuss the construction of approximate simultaneous confidence intervals for the function estimates. Our preference is to accomplish this task via bootstrapping.

#### 4. Bootstrap simultaneous confidence intervals

In this section, we present a method to obtain approximate simultaneous confidence intervals that can be used to determine the frequencies where the spectra differ, if any. To make pairwise comparisons, we make use of



**Fig. 1.** Simulation example comparing  $L=2$  groups of four AR(1) processes ( $\phi = \pm 0.9$  and  $n=500$ ): True  $\beta_1(\omega) = A_{12}(\omega)$  (solid line in the middle), the estimated function,  $\hat{\beta}_1(\omega)$  (dotted line in the middle), and the approximate 95% upper and lower limits (dashed lines) for  $\beta_1(\omega)$  based on the bootstrap method with  $N_B = 20$  and  $R = 10$ .

the fact that

$$\Delta_{\ell_1, \ell_2}(\omega) = \beta_{\ell_1}(\omega) - \beta_{\ell_2}(\omega) = \log \frac{f_{\ell_1}(\omega)}{f_{\ell_2}(\omega)} = g_{\ell_1}(\omega) - g_{\ell_2}(\omega) \tag{28}$$

for  $\ell_1, \ell_2 = 1, \dots, L$ , where we take  $\beta_1(\omega) \equiv 0$ . Note that inferences for individual  $\beta$ -functions can be carried out by setting  $\ell_2 = L$  in (28); i.e.,  $\Delta_{\ell_1, L}(\omega) = \beta_{\ell_1}(\omega)$ . Our basic interest is in whether  $\Delta_{\ell_1, \ell_2}(\omega) \equiv 0$ , but the essential goal is to find the frequencies for which  $\Delta_{\ell_1, \ell_2}(\omega)$  differs from zero.

Our technique employs the bootstrap using the ideas presented in Olshen et al. (1989). To this end, suppose we obtain a random sample of size  $N_B$  time series with replacement from our original collection of time series. Compute from this bootstrap sample,  $N_B$  estimates of  $\Delta(\omega_k)$ , say  $\hat{\Delta}^{(i)}(\omega_k)$ , for  $i=1, \dots, N_B$ , and from those compute the average  $\bar{\Delta}_B(\omega_k)$  and variance  $\hat{\sigma}_{\Delta, B}^2(\omega_k)$ . For each positive  $c$ , let

$$\hat{F}_B(c) = \frac{1}{N_B} \# \left\{ \hat{\Delta}(\omega_k) : \max_{\omega_k} \left| \frac{\hat{\Delta}(\omega_k) - \bar{\Delta}_B(\omega_k)}{\hat{\sigma}_{\Delta, B}(\omega_k)} \right| \leq c \right\} \tag{29}$$

Here,  $\hat{\Delta}(\omega_k)$  refers to the estimate of  $\Delta(\omega_k)$  for an individual member of the original time series. Now, repeat the bootstrap sampling and computation of  $\hat{F}_B(c)$ , say  $R$  times and let  $\bar{F}_B(c)$  denote the average of the  $R \hat{F}_B(c)$ . For  $0 < \alpha < 1$ , we define  $c_\alpha$  by

$$c_\alpha = \min\{c : \bar{F}_B(c) \geq \alpha\} \tag{30}$$

so that  $c_\alpha$  is the approximate critical value used in our intervals of the form  $\bar{\Delta}(\omega_k) - c_\alpha \hat{\sigma}_{\Delta, B}(\omega_k)$  and  $\bar{\Delta}(\omega_k) + c_\alpha \hat{\sigma}_{\Delta, B}(\omega_k)$  for all  $\omega_k$ . Justification for this technique is provided in Olshen et al. (1989).

### 5. Simulation example

As a simple example, we simulated  $L=2$  groups of series, each with length  $n=500$ , and with four,  $n_1=n_2=4$ , series in each group. Each group of series was generated from a Gaussian AR(1) process,  $y(t) = \phi y(t-1) + w(t)$ , with  $\phi = 0.9$  for the first group and  $\phi = -0.9$  for the second group; in both cases  $w(t) \sim \text{iid } N(0,1)$ . In this case the model for the group log-spectrum is  $g_\ell(\omega) = \beta_0(\omega) + \beta_1(\omega)\delta_{\ell,1}$ , for  $\ell = 1, 2$ , where  $\delta$  is the Kronecker delta, so that  $\beta_1(\omega) = \Delta_{12}(\omega) = g_1(\omega) - g_2(\omega)$ . Fig. 1 displays the true  $\beta_1(\omega)$  (solid line in the middle),  $\hat{\beta}_1(\omega)$ 's (dotted line in the middle), and the approximate 95% upper and lower

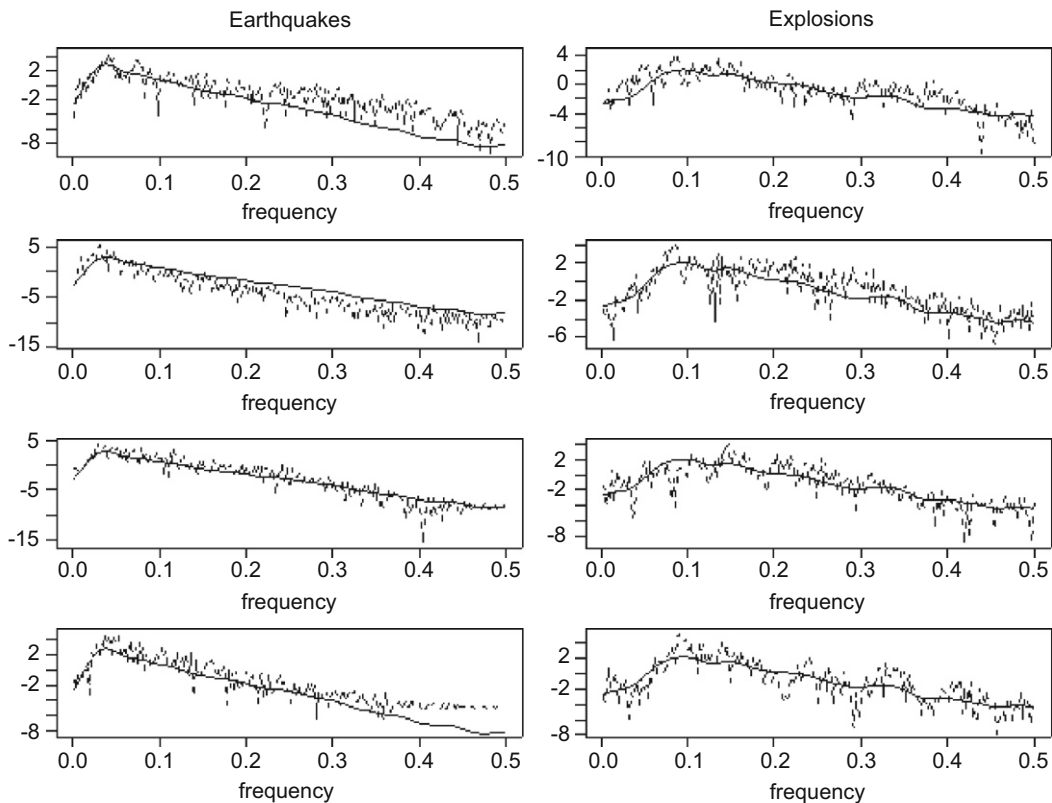
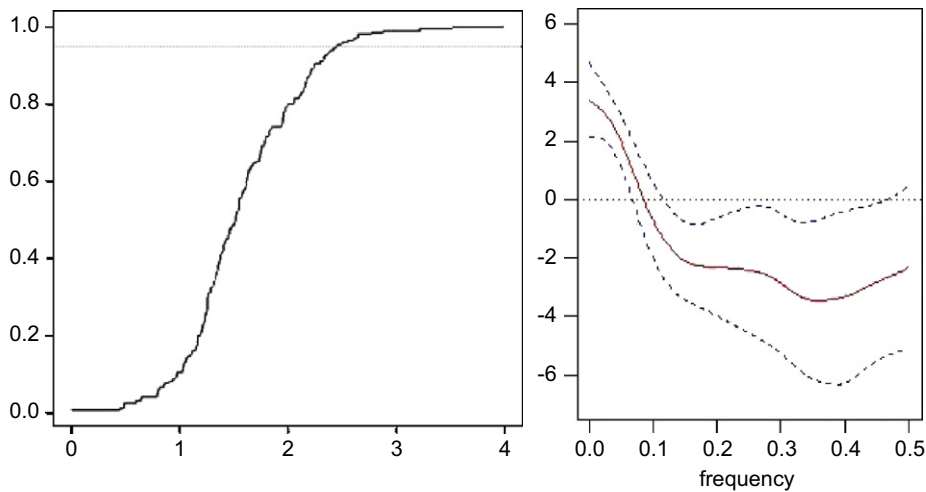


Fig. 2. The centered log-periodograms,  $y_{\ell, j}(\omega_k)$ , for the earthquake series (left) and the explosion series (right) as well as the corresponding smoothing spline estimates.



**Fig. 3.** Earthquake–explosion example: (left)  $\bar{F}_B(c)$  versus  $c$  with  $N_B = 20$  and  $R = 10$  indicating  $c_{0.95} = 2.45$ ; (right)  $\hat{\Delta}_{12}(\omega_k) = \hat{\beta}_1(\omega_k)$  along with the approximate 95% simultaneous confidence intervals. Note that  $\Delta_{12}(\omega) > 0$  implies that the spectral power an earthquake is greater than that of explosion at  $\omega$  and vice versa.

limits (dashed lines) for  $\beta_1(\omega)$  based on the bootstrap method with  $N_B = 20$  and  $R = 10$ . The empirical critical value in this case is  $c_{0.95} = 2.82$ .

## 6. Analysis of earthquakes and explosions

In this section we analyze the S-waves of four earthquakes (group 1) and four explosions (group 2) taken from [Shumway and Stoffer \(2006\)](#). Each series consists of  $n=512$  points. [Fig. 2](#) shows the centered log-periodograms,  $y_{\ell,j}(\omega_k) = \log I_{\ell,j}(\omega_k) + \gamma$ , for  $n_\ell = 4$  and  $\ell = 1, 2$ , from each group as well as the corresponding smoothing spline estimates with smoothing parameters  $\lambda = 0.0001$  and  $\theta_1 = 1$ . As in the simulated example of the previous section, the model for the group log-spectrum is  $g_\ell(\omega) = \beta_0(\omega) + \beta_1(\omega)\delta_{\ell,1}$ , for  $\ell = 1, 2$ , where  $\delta$  is the Kronecker delta, so that  $\beta_1(\omega) = \Delta_{12}(\omega) = g_1(\omega) - g_2(\omega)$  is the parameter function of interest; in particular, the interest is in discovering the frequencies for which  $\beta_1(\omega)$  is non-zero.

[Fig. 3](#) (left) displays  $\bar{F}_B(c)$  with  $N_B = 20$  and  $R = 10$  for obtaining the critical value for the simultaneous approximate confidence intervals for  $\beta_1(\omega)$  and we note that  $c_{0.95} = 2.45$  in this case. [Fig. 3](#) (right) shows  $\hat{\beta}_1(\omega_k)$ 's along with the approximate 95% simultaneous confidence intervals. Note that  $\beta_1(\omega) > 0$  implies that the spectral power an earthquake is greater than that of explosion at  $\omega$  and vice versa. Based on our results, we conclude that the two groups differ over almost the entire frequency range with the spectral power of the earthquakes being greater than that of explosions at the lower frequencies, say  $\omega \in (0, 0.15)$ , and the spectral power of the explosion being greater than that of the earthquakes at the midrange frequencies, say  $\omega \in (0.15, 0.40)$ .

## References

- Adams, R., 1975. Sobolev Space. Academic Press, New York.
- Aronszajn, N., 1950. Theory of reproducing kernels. *Transactions of the American Mathematical Society* 68, 337–404.
- Bloomfield, P., 1973. An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–226.
- Brillinger, D.R., 1980. Analysis of variance problems under time series models. In: Krishnaiah, P.R. (Ed.), *Handbook of Statistics*, vol. 1. North-Holland, Amsterdam, pp. 237–278.
- Diggle, P., al Wasel, I., 1997. Spectral analysis of replicated biomedical time series. *Applied Statistics* 46, 31–71.
- Fokianos, K., Savvides, A., 2008. On comparing several spectral densities. *Technometrics* 50, 317–331.
- Gu, C., Wahba, G., 1993a. Semiparametric analysis of variance with tensor product thin plate spline. *Journal of the Royal Statistical Society, Series B* 55, 353–368.
- Gu, C., Wahba, G., 1993b. Smoothing spline ANOVA with component-wise Bayesian confidence intervals. *Journal of the Computational and Graphical Statistics* 2, 97–117.
- Guo, W., Dai, M., Ombao, H., von Sachs, R., 2003. Smoothing spline ANOVA for time-dependent spectral analysis. *Journal of the American Statistical Association* 98, 643–652.
- Olshen, R., Biden, E., Wyatt, M., Sutherland, D., 1989. Gait analysis and the bootstrap. *The Annals of Statistics* 17 (4), 1419–1440.
- Parzen, E., 1961a. Mathematical considerations in the estimation of spectra. *Technometrics* 3, 167–190.
- Parzen, E., 1961b. An approach to time series analysis. *Annals of Mathematical Statistics* 32, 951–989.
- Parzen, E., 1962. On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076.
- Parzen, E., 1993. Stationary time series analysis using information and spectral analysis. In: Rao, T.S. (Ed.), *Developments in Time Series Analysis*. In Honour of M. B. Priestley. Chapman & Hall, London, pp. 139–148.

- Qin, L., Guo, W., Litt, B., 2009. A time-frequency functional model for locally stationary time series data. *Journal of Computational and Graphical Statistics* 18 (3), 675–693.
- Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density functions. *Annals of Mathematical Statistics* 27, 642–669.
- Shumway, R.H., 1971. On detecting a signal in N stationarily correlated noise series. *Technometrics* 13, 499–519.
- Shumway, R.H., Stoffer, D.S., 2006. *Time Series Analysis and its Applications: With R Examples*, second ed. Springer-Verlag, New York.
- Wahba, G., 1980. Automatic smoothing of the log periodogram. *Journal of the American Statistical Association* 75, 122–132.
- Wahba, G., 1985. A comparison of GCV and GML for choosing the smoothing parameters in the generalized spline smoothing problem. *The Annals of Statistics* 13, 1378–1402.
- Wahba, G., 1990. *Spline Models for Observational Data*. SIAM, Philadelphia.