The Annals of Applied Statistics
2013, Vol. 7, No. 1, 570–587
DOI: 10.1214/12-AOAS601

CANONICAL CORRELATION ANALYSIS BETWEEN TIME SERIES AND STATIC OUTCOMES, WITH APPLICATION TO THE SPECTRAL ANALYSIS OF HEART RATE VARIABILITY

BY ROBERT T. KRAFTY[1,2] AND MARTICA HALL[2]

University of Pittsburgh

Although many studies collect biomedical time series signals from multiple subjects, there is a dearth of models and methods for assessing the association between frequency domain properties of time series and other study outcomes. This article introduces the random Cramér representation as a joint model for collections of time series and static outcomes where power spectra are random functions that are correlated with the outcomes. A canonical correlation analysis between cepstral coefficients and static outcomes is developed to provide a flexible yet interpretable measure of association. Estimates of the canonical correlations and weight functions are obtained from a canonical correlation analysis between the static outcomes and maximum Whittle likelihood estimates of truncated cepstral coefficients. The proposed methodology is used to analyze the association between the spectrum of heart rate variability and measures of sleep duration and fragmentation in a study of older adults who serve as the primary caregiver for their ill spouse.

1. Introduction. Scientific and technological advances have led to an increase in the number of studies that collect and analyze biological time series signals from multiple subjects. In many instances, the frequency domain properties of the time series contain interpretable physiological information. Examples of such time series include electroencephalographic signals [Buysse et al. (2008), Qin and Wang (2008)], hormone concentration levels [Diggle and Al Wasel (1997), Gronfier and Brandenberger (1998)], and heart rate variability [Hall et al. (2007), Krafty, Hall and Guo (2011)]. The goal of many such studies is to quantify the association between power spectra and collections of correlated outcomes.

This article is motivated by a study whose objective is to better understand the association between stress and sleep in older adults who are the primary caregiver for their spouse. In this study, heart rate variability and multiple measures of sleep duration and fragmentation are collected from participants during a night of sleep. Heart rate variability is the measure of variability in the elapsed time between consecutive heart beats. Its power spectrum has been shown to be an indirect measure of autonomic nervous system activity and is used by researchers as a measure

Received December 2011; revised September 2012.
[1]Supported by NIH Grants P01-AG20677 and HL104607.
[2]Supported by NSF Grant DMS-01-04129.
Key words and phrases. Canonical correlation analysis, cepstral analysis, heart rate variability, sleep, spectral analysis, time series.

of stress [Task Force of the ESC/ASPE (1996), Hall et al. (2007)]. Measures of sleep duration and fragmentation have been shown to be associated with health and functioning when measured either subjectively through self-report sleep diaries or objectively through the collection of electrophysiological signals known as polysomnography (PSG) [McCall et al. (1995), Hall et al. (2008), Nock et al. (2009), Silva et al. (2007), Vgontzas et al. (2010)]. We desire an analysis of these data that can illuminate the relationship between stress and sleep by quantifying the association between the spectrum of heart rate variability and both self-reported and PSG derived measures of sleep.

The majority of methods for the spectral analysis of time series from multiple subjects where spectra depend on static covariates deal exclusively with covariates that take the form of qualitative grouping variables [Shumway (1971), Diggle and Al Wasel (1997), Brillinger (2001), Fokianos and Savvides (2008), Stoffer et al. (2010)]. These methods are not applicable when the covariates are quantitative variables such as measures of sleep duration and fragmentation. Krafty, Hall and Guo (2011) introduced the mixed effects Cramér representation as a model for time series data where subject-specific power spectra depend on covariates and can account for quantitative variables. The mixed effects Cramér representation has two characteristics which limit its effectiveness for modeling and analyzing time series and correlated static outcomes. First, it assumes a semiparametric model for log-spectra conditional on static outcomes. As is the case in our motivating study, a semiparametric form is often unknown and a nonparametric model is required. Second, it provides a measure of association between time series and a static outcome conditional on the other outcomes through a regression coefficient. When the outcomes are correlated, extracting scientifically meaningful information from the conditional associations provided by the multiple regression coefficients can be challenging. In our motivating study, clinically useful information concerning the relationship between the spectrum of heart rate variability and the multiple correlated measures of sleep duration and fragmentation requires parsimonious measures of association.

To offer a nonparametric model and interpretable measures of association between time series and sets of correlated static outcomes, we introduce the random Cramér representation and ensuing canonical correlation analysis (CCA). The random Cramér representation considered in this article is a nonparametric joint model for time series and sets of static outcomes where the transfer function of the time series is random and the subject-specific log-spectra are correlated with the static outcomes. Unlike the mixed effects Cramér representation of Krafty, Hall and Guo (2011), no conditional semiparametric form for the log-spectrum is assumed. The theoretical framework introduced by Eubank and Hsing (2008) is used to define a CCA between the cepstral coefficients, or the Fourier coefficients of the log-spectrum [Bogert, Healy and Tukey (1963)], and the static outcomes. Estimates of canonical correlations and weight functions are obtained through a procedure which first estimates the cepstral coefficients via Whittle likelihood regression, then performs a standard multivariate CCA between the estimated cepstral coefficients and static outcomes.

The rest of the article is organized as follows. Section 2 describes our motivating study: the AgeWise Study. Section 3 introduces the random Cramér representation and CCA. The estimation procedure is developed in Section 4. Section 5 presents the results of a simulation study and the proposed method is applied to data from the AgeWise Study in Section 6. A discussion is offered in Section 7.

2. The AgeWise study. The mental and emotional stress faced by older adults who are the primary caregiver for their ill spouse places them at an increased risk for the development of disturbed sleep which can effect their health and functioning [McCurry et al. (2007)]. A goal of the AgeWise Study conducted at the University of Pittsburgh is to gain a better understanding of the association between stress and sleep in older adults who are the primary caregiver for their ill spouse in order to inform the development of behavioral interventions to enhance their sleep.

The participants in this project are $N = 46$ men and women between 60–89 years of age. Each participant serves as the primary caregiver for their spouse who is suffering from a progressive dementing illness such as Alzheimer's or advanced Parkinson's disease. Participants were studied during a night of in-home sleep through ambulatory PSG. The recorded PSG signals were used to compute objective measures of total sleep time (TST) as the number of minutes spent asleep during the night, sleep latency (SL) as the number of minutes elapsed between attempted sleep and sleep onset, and wakefulness after sleep onset (WASO) as the number of minutes spent awake between sleep onset and the final morning awakening. Upon awakening, the participants completed a self-report sleep diary which was used to compute self-reported measures of TST, SL, and WASO.

The ambulatory PSG included a modified 2-lead electrode placement to collect the electrocardiogram signal continuously throughout the night at a 512 Hz sampling rate. The electrocardiogram signal was digitally stored for off-line processing, cleaned of artifacts, and used to identify the R-waves as the location of the upward deflection of the electrocardiogram signal associated with each heartbeat. Interbeat intervals were then computed as the number of milliseconds between each successive pair of R-waves to provide a measure of the elapsed time between consecutive heart beats. Visually scored sleep staging was temporally aligned with the interbeat intervals for each participant to allow for the isolation of the epochs of interbeat intervals during the first three minutes of uninterrupted stage 2 sleep. To ensure proper physiological interpretation in accordance with established guidelines for nonparametric spectral analysis, we analyze the heart rate variability series generated by sampling the cubic interpolation of the interbeat intervals versus the R-waves at 2 Hz, resulting in time series of length $T = 360$ [Task Force of the ESC/ASPE (1996)].

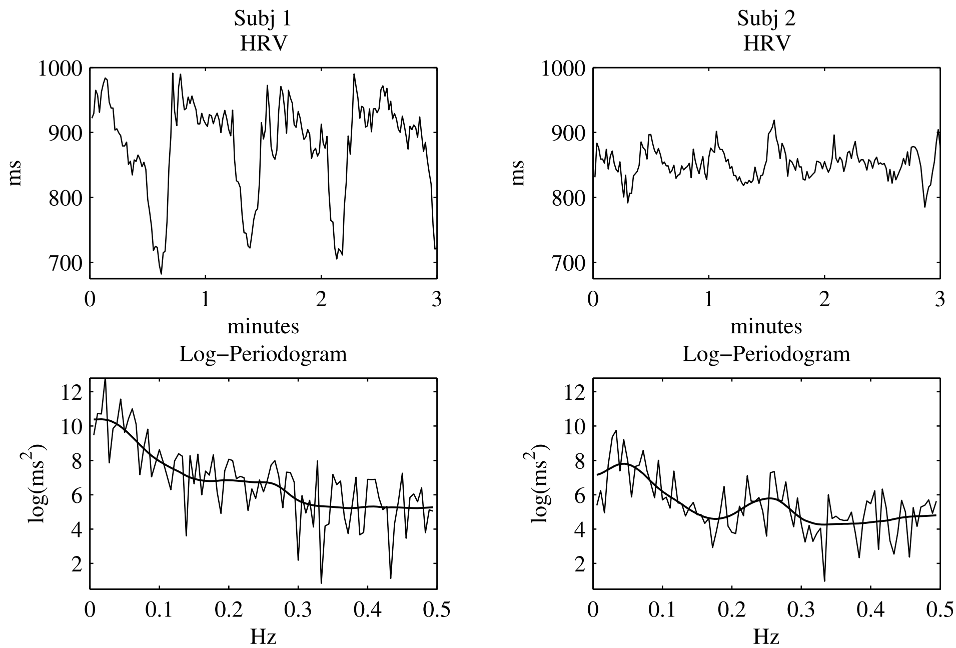

FIG. 1. *Heart rate variability, bias-adjusted log-periodograms, and estimated log-spectra for two subjects during the first three minutes of stage* 2 *sleep. The sleep outcomes for these two subjects are displayed in Table* 1.

The data for two subjects in the form of heart rate variability time series and six measures of sleep duration and fragmentation are displayed in the top panels of Figure 1 and in Table 1, respectively. The primary objective of our analysis is to illuminate the relationship between stress and sleep by obtaining low-dimensional and interpretable measures of the association between the spectrum of heart rate variability at the start of stage 2 sleep and self-reported and PSG derived measures of sleep duration and fragmentation.

3. Measuring association between log-spectra and static outcomes.

3.1. *Random Cramér representation.* This article is concerned with quantifying the association between the spectrum of a second order stationary time series of length T, $\{X_{j1}, \dots, X_{jT}\}$, and a P-dimensional vector of correlated outcomes, $\mathbf{Z}_j$, from $j = 1, \dots, N$ independent subjects. In our motivating sleep study, the time series of length $T = 360$ are three minute epochs of heart rate variability from $N = 46$ participants, while $\mathbf{Z}_j$ are $P = 6$ dimensional vectors of TST, WASO, and SL as measured by self-report sleep diary and by PSG.

The outcomes $\mathbf{Z}_j$ are assumed to be independent and identically distributed with $\boldsymbol{\mu}_Z = \mathrm{E}(\mathbf{Z}_j)$ and nonsingular covariance kernel

$$\boldsymbol{\Gamma}_Z = \mathrm{E}(\mathbf{Z}_j - \boldsymbol{\mu}_Z)(\mathbf{Z}_j - \boldsymbol{\mu}_Z)'.$$

TABLE 1
Time spent asleep (TST), sleep latency (SL), and wakefulness after sleep onset (WASO) as measured by polysomnography (PSG) and self-reported sleep diary (D) for the two subjects whose heart rate variability are displayed in Figure 1. All sleep outcomes are reported in minutes

	Subj. 1	**Subj. 2**
PSG-TST	394	373
PSG-SL	11	10
PSG-WASO	95	78
D-TST	383	403
D-SL	20	2
D-WASO	45	15

The time series are modeled through a random Cramér representation with a random mean u_j and a random transfer function Θ_j that is correlated with $\mathbf{Z}_j$. The random transfer functions Θ_j are independent and identically distributed complex-valued random functions over $\mathbb{R}$ that are Hermitian, square-integrable over [0, 1], and have period 1. Formally, the random Cramér representation for X_{jt} is

$$X_{jt} = u_j + \int_0^1 \Theta_j(\omega)e^{2\pi i\omega t}\, d\Lambda_j(\omega),$$

where Λ_j are mutually independent identically distributed mean-zero orthogonal increment processes over [0, 1] such that $\mathrm{E}|d\Lambda_j(\omega)|^2 = d\omega$ and Λ_j is independent of $\Theta_{j'}$, $\mathbf{Z}_{j'}$, and $u_{j'}$ for all j and j'. The time series $\{X_{jt} : t \in \mathbb{Z}\}$ exists with probability one and is second order stationary.

In many applications, such as our motivating study, scientific interest lies in the ratio of power at different frequencies. This is equivalent to looking at linear combinations of the log-spectrum. Consequently, we consider spectral properties on the log-scale and define the subject-specific log-spectrum for the jth subject as the random function

$$F_j(\omega) = \log\bigl|\Theta_j(\omega)\bigr|^2.$$

To assure that the first two moments of F_j exist and are bounded, it is assumed that $\sup_{\omega\in\mathbb{R}} E|\Theta_j(\omega)|^4 < \infty$ and $\inf_{\omega\in\mathbb{R}} E|\Theta_j(\omega)|^4 > 0$. We focus on log-spectra which possess square integrable first derivatives and define $\mathbb{F}$ as the space of even functions with period 1 whose first derivatives are square integrable.

When $F_j \in \mathbb{F}$ with probability 1, the subject-specific log-spectra possess the cosine expansion

$$F_j(\omega) = f_{j0} + \sum_{k=1}^{\infty} f_{jk}\sqrt{2}\cos(2\pi\omega k),$$

$$f_{j0} = \int_0^1 F_j(\omega)\, d\omega,$$

$$f_{jk} = \int_0^1 F_j(\omega)\sqrt{2}\cos(2\pi\omega k)\, d\omega, \qquad k = 1, 2, \ldots,$$

where $\mathbf{f}_j = (f_{jk} : k = 0, 1, \ldots)$ is the subject-specific cepstrum [Bogert, Healy and Tukey (1963)]. Our analysis will explore spectral properties of times series via the cepstral coefficients and will make use of the covariance and cross-covariance kernels

$$\Gamma_f(k, k') = \mathrm{E}[(f_{jk} - \mu_k)(f_{jk'} - \mu_{k'})], \qquad k, k' = 0, 1, \ldots,$$

$$\Gamma_{fZ}(k) = \mathrm{E}[(f_{jk} - \mu_k)(\mathbf{Z}_j - \boldsymbol{\mu}_Z)'], \qquad k = 0, 1, \ldots,$$

where $\mu_k = E(f_{jk})$.

3.2. *Canonical correlation analysis*. To provide a parsimonious measure of association between time series and static outcomes following a random Cramér representation, we will utilize the definition of CCA between two sets of second order random variables introduced by Eubank and Hsing (2008). We want to find successive linear combinations of cepstral coefficients and static outcomes that are maximally correlated. The first canonical correlation ρ_1 is defined as

$$\rho_1^2 = \sup_{\alpha_k, \boldsymbol{\beta}} \mathrm{Cov}^2\left(\sum_{k=0}^{\infty} \alpha_k f_{jk}, \boldsymbol{\beta}'\mathbf{Z}_j\right)$$

over all $\alpha_k \in \mathbb{R}$, $k = 0, 1, \ldots,$ and $\boldsymbol{\beta} \in \mathbb{R}^P$ such that the random variables $\sum_{k=0}^{\infty} \alpha_k f_{jk}$ and $\boldsymbol{\beta}'\mathbf{Z}_j$ have unit variance. The series $\mathbf{a}_1 = (a_{1k} : k = 0, 1, \ldots)$ and vector $\mathbf{B}_1$ where this maximum occurs are referred to as first canonical weights for the cepstrum and static outcomes, while $\sum_{k=0}^{\infty} a_{1k} f_{jk}$ and $\mathbf{B}_1'\mathbf{Z}_j$ are first canonical variables. For $q = 2, \ldots, Q$ where Q is the minimum of P and the rank of Γ_f, the qth canonical correlation ρ_q is defined as

$$\rho_q^2 = \sup_{\alpha_k, \boldsymbol{\beta}} \mathrm{Cov}^2\left(\sum_{k=0}^{\infty} \alpha_k f_{jk}, \boldsymbol{\beta}'\mathbf{Z}_j\right)$$

over all $\alpha_k \in \mathbb{R}$, $k = 0, 1, \ldots,$ and $\boldsymbol{\beta} \in \mathbb{R}^P$ such that $\sum_{k=0}^{\infty} \alpha_k f_{jk}$ and $\boldsymbol{\beta}'\mathbf{Z}_j$ have unit variance and are pairwise uncorrelated with $\sum_{k=0}^{\infty} a_{q'k} f_{jk}$ and $\mathbf{B}_{q'}'\mathbf{Z}_j$ for $q' < q$. The series $\mathbf{a}_q = (a_{qk} : k = 0, 1, \ldots)$ and vector $\mathbf{B}_q$ where the maximum is achieved are referred to as qth weight functions for the cepstrum and static outcomes, while $\sum_{k=0}^{\infty} a_{qk} f_{jk}$ and $\mathbf{B}_q'\mathbf{Z}_j$ are the qth canonical variables. We will assume that $\boldsymbol{\Gamma}_{fZ}\boldsymbol{\Gamma}_Z^{-1}$ is well defined and Hilbert–Schmidt as an operator from $\mathbb{R}^P$ to the reproducing kernel Hilbert space with reproducing kernel $\boldsymbol{\Gamma}_f$. Under this regularity condition, Theorems 1 and 2 of Eubank and Hsing (2008) assure the existence of the canonical correlations, weight functions, and variables.

When $\sum_{k=0}^{\infty} |a_{qk}| < \infty$, the qth canonical variable can be represented as a linear function of the log-spectra F_j,

$$\sum_{k=0}^{\infty} a_{qk} f_{jk} = \int_0^1 A_q(\omega) F_j(\omega)\, d\omega,$$

where

$$A_q(\omega) = a_{q0} + \sum_{k=1}^{\infty} a_{qk} \sqrt{2} \cos(2\pi\omega k).$$

When it exists, we will refer to A_q as the qth weight function for the log-spectrum. In our application, the goal is to estimate and analyze canonical correlations and the canonical weight functions for the static outcomes and log-spectra. Although the canonical variables for the cepstra do not always possess forms as integral function of log-spectra, in Section 3.3 it is shown that they can always be approximated as such.

3.3. *Finite approximation.* The infinite-dimensional formulation of the CCA given above is not conducive to real data applications. A finite-dimensional approximation can be obtained by noting that when $F_j \in \mathbb{F}$, the cepstral coefficients decay such that $\sum_{k=1}^{\infty} k^2 f_{jk}^2 < \infty$ with probability 1. The decay of the cepstral coefficients was utilized by Bloomfield (1973) to offer a finite-dimensional model for the log-spectrum by truncating the cosine series at some $K < T$ such that $F_j(\omega) \approx f_{j0} + \sum_{k=1}^{K-1} f_{jk} \sqrt{2} \cos(2\pi\omega k)$. Under this approximation, the cepstrum can be represented as the K-vector $\tilde{\mathbf{f}}_j = (f_{j0}, \dots, f_{jK-1})'$, which has $K \times K$ covariance matrix $\widetilde{\boldsymbol{\Gamma}}_f$ and $K \times P$ cross-covariance with the static outcomes $\widetilde{\boldsymbol{\Gamma}}_{fZ}$.

The CCA between the truncated cepstrum and static outcomes is a standard multivariate CCA problem [Johnson and Wichern (2002), Chapter 10.2]. To find the canonical correlations $\widetilde{\rho}_q$ and variables $\widetilde{\mathbf{a}}_q' \widetilde{\mathbf{f}}_j, \widetilde{\mathbf{B}}_q' \mathbf{Z}_j$ between $\widetilde{\mathbf{f}}_j$ and $\mathbf{Z}_j$, define η_q to be the qth largest eigenvalue of $\boldsymbol{\Gamma}_Z^{-1/2} \widetilde{\boldsymbol{\Gamma}}_{fZ}' \widetilde{\boldsymbol{\Gamma}}_f^{-} \widetilde{\boldsymbol{\Gamma}}_{fZ} \boldsymbol{\Gamma}_Z^{-1/2}$ with associated eigenvector $\mathbf{v}_q$ where $\widetilde{\boldsymbol{\Gamma}}_f^{-}$ is the Moore–Penrose generalized inverse of $\widetilde{\boldsymbol{\Gamma}}_f$. The canonical correlations $\widetilde{\rho}_q$ and weight functions $\widetilde{\mathbf{a}}_q$ and $\widetilde{\mathbf{B}}_q$ can be computed as

$$\begin{aligned}
\widetilde{\rho}_q &= \sqrt{\eta_q}, \\
\widetilde{\mathbf{a}}_q &= \widetilde{\rho}_q^{-1} \widetilde{\boldsymbol{\Gamma}}_f^{-} \widetilde{\boldsymbol{\Gamma}}_{fZ} \boldsymbol{\Gamma}_Z^{-1/2} \mathbf{v}_q, \qquad (3.1)\\
\widetilde{\mathbf{B}}_q &= \boldsymbol{\Gamma}_Z^{-1/2} \mathbf{v}_q.
\end{aligned}$$

These canonical correlations and weight functions are approximations of the canonical correlations and weight functions defined in Section 3.2. A direct consequence of Lemma 5 of Eubank and Hsing (2008) is that, as $K \to \infty$, $\widetilde{\rho}_q \to \rho_q$, $(\widetilde{\mathbf{a}}_q', 0, \dots) \to \mathbf{a}_q$ in $L^2(\mathbf{f}_j)$, and $\widetilde{\mathbf{B}}_q \to \mathbf{B}_q$.

Although the log-spectral weight function A_q does not necessarily exist, the log-spectral weight function from the finite approximation

$$\widetilde{A}_q(\omega) = \widetilde{a}_{q0} + \sum_{k=1}^{K-1} \widetilde{a}_{qk}\sqrt{2}\cos(2\pi\omega k)$$

is always well defined.

4. Estimation. Estimates of $\widetilde{\boldsymbol{\Gamma}}_f$, $\widetilde{\boldsymbol{\Gamma}}_{fZ}$, and $\widetilde{\boldsymbol{\Gamma}}_Z$ will be plugged into (3.1) to obtain estimates of the canonical correlations and weight functions. The covariance of $\mathbf{Z}_j$ is estimated with the standard estimator $\widehat{\boldsymbol{\Gamma}}_Z = (N-1)^{-1}\sum_{j=1}^{N}(\mathbf{Z}_j - \overline{\mathbf{Z}})(\mathbf{Z}_j - \overline{\mathbf{Z}})'$, where $\overline{\mathbf{Z}} = N^{-1}\sum_{j=1}^{N}\mathbf{Z}_j$.

To estimate $\widetilde{\boldsymbol{\Gamma}}_f$ and $\widetilde{\boldsymbol{\Gamma}}_{fZ}$, we consider the periodograms

$$Y_{j\ell} = T^{-1}\left|\sum_{t=1}^{T} X_{jt}e^{-2\pi i\ell t/T}\right|^2, \qquad j = 1,\dots,N, \ell = 1,\dots,\left\lfloor\frac{T-1}{2}\right\rfloor,$$

which are approximately independent and distributed as $e^{F_j(\ell/T)}\chi_2^2/2$ when T is large [Krafty, Hall and Guo (2011), Theorem 1]. This large sample distribution of the periodogram leads to a Whittle likelihood [Whittle (1953, 1954)] for truncated cepstral coefficients. The negative log-Whittle likelihood for the truncated cepstral coefficients of the jth subject is

$$\mathcal{L}_{jK}(f_0,\dots,f_{K-1}) = \sum_{\ell=1}^{\lfloor(T-1)/2\rfloor}\Bigg\{Y_{j\ell}e^{-[f_0+\sum_{k=1}^{K-1} f_k\sqrt{2}\cos(2\pi k\ell/T)]} + f_0 + \sum_{k=1}^{K-1} f_k\sqrt{2}\cos(2\pi k\ell/T)\Bigg\}.$$

We propose using Whittle likelihood regression to estimate $\widetilde{\boldsymbol{\Gamma}}_f$ and $\widetilde{\boldsymbol{\Gamma}}_{fZ}$ with

$$\widehat{\boldsymbol{\Gamma}}_f = (N-1)^{-1}\sum_{j=1}^{N}(\hat{\mathbf{f}}_j - \bar{\mathbf{f}})(\hat{\mathbf{f}}_j - \bar{\mathbf{f}})',$$

$$\widehat{\boldsymbol{\Gamma}}_{fZ} = (N-1)^{-1}\sum_{j=1}^{N}(\hat{\mathbf{f}}_j - \bar{\mathbf{f}})(\mathbf{Z}_j - \overline{\mathbf{Z}})',$$

where $\hat{\mathbf{f}}_j = (\hat{f}_{j0},\dots,\hat{f}_{jK-1})'$ minimizes $\mathcal{L}_{jK}$ and $\bar{\mathbf{f}} = N^{-1}\sum_{j=1}^{N}\hat{\mathbf{f}}_j$. A Fisher's scoring algorithm for computing $\hat{\mathbf{f}}_j$ is given in the Appendix.

Estimates $\widehat{\rho}_q$, $\widehat{\mathbf{a}}_q$, $\widehat{\mathbf{B}}_q$, and $\widehat{A}_q$ of the qth canonical correlation and weight functions for cepstra, static outcomes, and log-spectra are then defined for $q = 1,\dots,Q$

as

$$\widehat{\rho}_q = \sqrt{\widehat{\eta}_q},$$

$$\widehat{\mathbf{a}}_q = \widehat{\rho}_q^{-1}\widehat{\boldsymbol{\Gamma}}_f^{-}\widehat{\boldsymbol{\Gamma}}_{fZ}\widehat{\boldsymbol{\Gamma}}_Z^{-1/2}\widehat{\mathbf{v}}_q,$$

$$\widehat{\mathbf{B}}_q = \widehat{\boldsymbol{\Gamma}}_Z^{-1/2}\widehat{\mathbf{v}}_q,$$

$$\widehat{A}_q(\omega) = \widehat{a}_{q0} + \sum_{k=1}^{K-1}\widehat{a}_{qk}\sqrt{2}\cos(2\pi\omega k), \qquad \omega \in \mathbb{R},$$

where $\widehat{\eta}_q$ is the qth largest eigenvalue of the $P \times P$ matrix $\widehat{\boldsymbol{\Gamma}}_Z^{-1/2}\widehat{\boldsymbol{\Gamma}}'_{fZ}\widehat{\boldsymbol{\Gamma}}_f^{-}\widehat{\boldsymbol{\Gamma}}_{fZ}\widehat{\boldsymbol{\Gamma}}_Z^{-1/2}$ with associated eigenvector $\widehat{\mathbf{v}}_q$. The estimated truncated cepstral coefficients also provide estimates of the subject-specific log-spectra as $\widehat{F}_j(\omega) = \hat{f}_{j0} + \sum_{k=1}^{K-1}\hat{f}_{jk}\sqrt{2}\cos(2\pi\omega k)$.

These estimates depend on the number of nonzero cepstral coefficients K. Simulation studies have demonstrated favorable empirical performance of the AIC as a data driven procedure for selecting K by minimizing

$$\mathcal{C}(k) = \sum_{j=1}^{N}\mathcal{L}_{jk}(\hat{f}_{j0},\ldots,\hat{f}_{jk-1}) + 2Nk.$$

We provide Matlab code for implementing the proposed estimation procedure in the supplemental file Krafty and Hall (2013b).

5. Simulation study.

5.1. *Setting*. A simulation study was conducted to explore the empirical properties of the proposed estimation procedure and compare it to two alternatives. For each simulated data set, log-spectra

$$F_j(\omega) = 5 + \sqrt{2}\cos(2\pi\omega) + \xi_{j0} + \sum_{k=1}^{3}\xi_{jk}\sqrt{2}\cos(2\pi k\omega)$$

were simulated where ξ_{jk} are independent mean zero normal random variables with $\mathrm{Var}(\xi_{jk}) = 4$. Static outcomes $\mathbf{Z}_j$ of dimension $P = 3$ were drawn as mean zero normal random vectors with covariance matrices $\mathrm{diag}(4, 4, 4)$ such that the elements of $\mathbf{Z}_j$ are uncorrelated with ξ_{jk}, $k = 0,\ldots,3$, except $\mathrm{Corr}(\xi_{j2}, Z_{j1}) = 0.5$ and $\mathrm{Corr}(\xi_{j3}, Z_{j2}) = 0.25$. Under this setting, the canonical correlations are $\rho_1 = 0.5$, $\rho_2 = 0.25$, $\rho_3 = 0$, the weight functions for the log-spectra are $A_1(\omega) = \cos(4\pi\omega)/\sqrt{2}$, $A_2(\omega) = \cos(6\pi\omega)/\sqrt{2}$, and the weight functions for the static outcomes are $\mathbf{B}_1 = (0.5, 0, 0)'$, $\mathbf{B}_2 = (0, 0.5, 0)'$. After a replicate-specific log-spectrum was simulated, its square-root was calculated and used as the replicate-specific transfer function to simulate the conditionally Gaussian time series X_{jt} in accordance with Theorem 2 of Dai and Guo (2004). Five-hundred random samples were drawn for each of the six combinations of $N = 50, 100$ and

$T = 30, 50, 100$. Results from additional settings under varying levels of signal strength and smoothness are presented in the supplemental article Krafty and Hall (2013a).

5.2. *Estimation procedures*. In addition to the proposed cepstral-based procedure, we also investigated two alternative estimation procedures adapted from the functional CCA literature. The CCA between two functional valued variables has been explored by many researches including Leurgans, Moyeed and Silverman (1993), He, Müller and Wang (2003, 2004), and Eubank and Hsing (2008). These methods can be adapted to our setting where one set of variables, F_j, is functional and observed with noise over a discrete grid through the periodograms, and the other, $\mathbf{Z}_j$, is multivariate.

The first alternative estimation procedure considered is an adaptation of the algorithm for functional CCA presented in Section 6 of Eubank and Hsing (2008). In this procedure, we used the penalized Whittle likelihood of Qin and Wang (2008) to obtain smoothing spline estimators of the subject-specific log-spectra F_j with smoothing parameters selected through direct generalized maximum likelihood. These estimated log-spectra were discretized to form vectors of estimated log-spectra at Fourier frequencies between 1 and $\lfloor (T-1)/2 \rfloor$; a canonical correlation analysis was performed between these vectors and $\mathbf{Z}_j$. The rank of the covariance kernel of the discretized log-spectra was selected through a cross-validation procedure which seeks to optimize the first canonical correlation by maximizing the function CV_1 discussed in Section 2.5 of He, Müller and Wang (2004).

The second alternative estimation procedure is an adaptation of the empirical basis approach for functional CCA advocated by He, Müller and Wang (2004). This procedure began by computing the singular value decomposition of the sample covariance of the bias-adjusted log-periodograms, $\log(Y_{j\ell}) + \gamma$ where $\gamma \approx 0.577$ is the Euler–Mascheroni constant. The eigenvectors were smoothed to obtain a functional basis. The bias-adjusted log-periodograms were then projected onto a finite number of these basis functions and a multivariate CCA was performed between the projections and the static outcomes. The number of basis functions was selected through cross-validation by maximizing CV_1 [He, Müller and Wang (2004), Section 2.5].

5.3. *Results*. We assessed performance through the square error of estimates of the canonical correlations, the square error of estimated weight functions for the static outcomes in the standard Euclidian norm, and the square error of the vector of estimated weight functions for the log-spectra evaluated at the Fourier frequencies in the standard Euclidian norm. The mean and standard deviation of the square errors are displayed in Table 2. The mean and variance of the errors of the proposed cepstrum-based estimator were smaller than those of the two alternatives for each parameter under every setting. The most drastic benefit in the cepstral-based procedure was found in the estimation of the canonical weight functions for the log-spectra. The modification of the function CCA algorithm of Eubank and Hs-

TABLE 2

Simulation results: The mean (standard deviation) of the square error $\times 10^2$ *of the estimators of the first two sets of weight functions and of the three canonical correlations. Three estimation procedures are implemented: Cep, the proposed cepstral-based procedure; FDA, adaptation of the functional CCA algorithm presented by* Eubank and Hsing (2008); *EB, adaptation of the empirical basis approach to functional CCA presented by He, Müller and Wang* (2004)

	$\widehat{A}_1$	$\widehat{A}_2$	$\widehat{\mathbf{B}}_1$	$\widehat{\mathbf{B}}_2$	$\widehat{\rho}_1$	$\widehat{\rho}_2$	$\widehat{\rho}_3$
$N = 100$, $T = 100$							
Cep	0.27 (0.75)	0.79 (2.47)	1.28 (2.19)	3.32 (3.67)	0.57 (0.66)	0.85 (1.07)	1.81 (1.62)
FDA	1.87 (3.48)	2.26 (4.37)	1.60 (2.37)	4.66 (4.18)	0.80 (0.94)	1.24 (1.58)	1.98 (2.57)
EB	1.17 (1.65)	2.17 (2.91)	1.73 (2.87)	4.64 (4.15)	1.09 (1.62)	1.79 (1.94)	2.87 (3.37)
$N = 100$, $T = 50$							
Cep	0.57 (1.11)	1.67 (3.50)	1.30 (2.08)	3.45 (3.67)	0.58 (0.68)	0.84 (1.02)	1.93 (1.68)
FDA	2.03 (3.50)	3.03 (4.86)	1.66 (2.46)	4.63 (4.49)	0.80 (0.91)	1.16 (1.44)	1.97 (2.41)
EB	1.74 (2.43)	3.34 (4.24)	1.61 (2.56)	4.48 (4.24)	1.15 (2.03)	1.72 (1.94)	2.79 (2.96)
$N = 100$, $T = 30$							
Cep	0.87 (0.83)	1.61 (2.38)	1.42 (2.21)	3.75 (4.03)	0.54 (0.64)	0.84 (1.09)	1.97 (1.67)
FDA	1.99 (3.18)	3.20 (4.52)	1.51 (2.06)	4.77 (4.65)	0.80 (1.00)	1.04 (1.19)	2.03 (2.75)
EB	1.65 (2.23)	3.27 (4.27)	1.62 (2.57)	4.44 (4.34)	1.19 (2.29)	1.44 (1.73)	2.54 (2.89)
$N = 50$, $T = 100$							
Cep	0.65 (2.47)	0.98 (2.89)	2.54 (3.32)	5.64 (4.94)	1.46 (1.68)	1.85 (2.19)	3.02 (2.74)
FDA	4.26 (7.16)	4.33 (7.41)	3.77 (3.93)	7.20 (5.14)	2.36 (2.47)	2.77 (3.07)	3.94 (5.69)
EB	1.53 (2.22)	2.07 (2.83)	3.44 (4.31)	7.24 (4.88)	2.62 (3.12)	3.94 (3.52)	4.82 (6.52)
$N = 50$, $T = 50$							
Cep	1.13 (2.48)	1.93 (3.71)	2.56 (3.38)	5.79 (4.92)	1.48 (1.70)	1.93 (2.18)	3.25 (2.89)
FDA	4.63 (7.05)	4.83 (6.75)	3.70 (4.14)	7.30 (5.17)	2.55 (2.57)	2.81 (3.20)	4.52 (5.63)
EB	2.64 (3.61)	3.32 (4.00)	3.60 (4.29)	6.86 (4.81)	2.63 (3.00)	3.79 (3.42)	4.95 (6.11)
$N = 50$, $T = 30$							
Cep	1.39 (1.71)	1.84 (2.38)	2.67 (3.33)	5.85 (4.95)	1.40 (1.73)	1.99 (2.20)	3.26 (2.94)
FDA	4.52 (6.39)	5.10 (6.65)	3.77 (4.05)	6.88 (5.05)	2.55 (2.63)	2.83 (3.41)	4.52 (5.90)
EB	3.12 (4.31)	4.04 (5.05)	3.50 (4.21)	6.95 (4.92)	2.54 (2.92)	3.50 (3.30)	4.55 (5.80)

ing (2008) had smaller error in estimating the canonical correlations as compared to the empirical basis approach, while the empirical basis approach demonstrated better performance in estimating the weight functions of the log-spectra.

6. Analysis of data from the AgeWise study.

6.1. *Data analysis.* We analyzed the data from the project described in Section 2 that consist of time series of heart rate variability during the first three minutes of stage 2 sleep and $P = 6$ measures of sleep duration and fragmentation from $N = 46$ participants. The mean, standard deviation, and correlation matrix of the sleep variables are displayed in Table 3. To aid in the interpretation of the weight

TABLE 3
Correlation matrix, means, and standard deviations of the six sleep variables: time spent asleep (TST), sleep latency (SL), and wakefulness after sleep onset (WASO) as measured by polysomnography (PSG) and self-reported sleep diary (D)

	PSG-TST	**PSG-SL**	**PSG-WASO**	**D-TST**	**D-SL**	**D-WASO**
Mean (minutes)	377.4	18.3	81.8	383.2	21.1	36.9
Standard deviation (minutes)	66.4	19.9	39.3	89.9	31.9	52.8
Correlation						
PSG-TST	1.00	−0.12	−0.11	0.45	−0.08	0.17
PSG-SL	−0.12	1.00	−0.31	0.06	−0.04	−0.22
PSG-WASO	−0.11	−0.31	1.00	−0.19	0.13	0.38
D-TST	0.45	0.06	−0.19	1.00	−0.51	−0.51
D-SL	−0.08	−0.04	0.13	−0.51	1.00	0.40
D-WASO	0.17	−0.22	0.38	−0.51	0.40	1.00

functions, we standardized the six-dimensional vector of sleep outcomes. The proposed procedure estimated the first two canonical correlations as $\widehat{\rho}_1 = 0.52$ and $\widehat{\rho}_2 = 0.19$; the remaining higher order correlations were estimated to be less than 3%. Figure 2 displays the estimated weight functions $\widehat{A}_1, \widehat{A}_2$ of the log-spectra, while Table 4 displays the estimated weight functions $\widehat{\mathbf{B}}_1, \widehat{\mathbf{B}}_2$ of the standardized sleep variables.

The estimated first canonical weight function for the standardized sleep outcomes is negative for all sleep measures expect for PSG derived WASO, which is close to zero. Note that the sum of SL, TST, and WASO measures the total

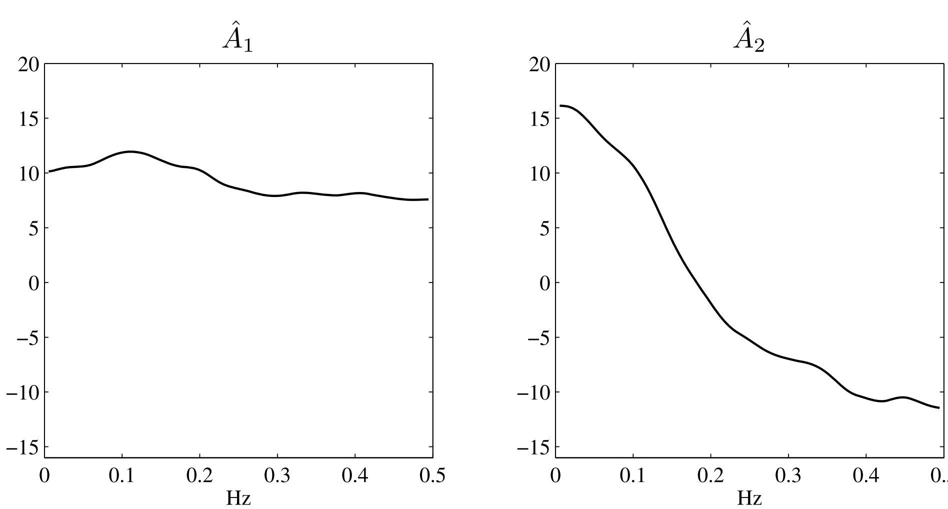

FIG. 2. *Estimated canonical weight functions for the log-spectrum of heart rate variability.*

TABLE 4

Estimated canonical weight functions of the standardized sleep variables: time spent asleep (TST), sleep latency (SL), and wakefulness after sleep onset (WASO) as measured by polysomnography (PSG) and self-reported sleep diary (D)

	$\widehat{\mathbf{B}}_1$	$\widehat{\mathbf{B}}_2$
PSG-TST	−0.42	−0.01
PSG-SL	−0.52	0.17
PSG-WASO	0.09	0.33
D-TST	−0.41	−0.75
D-SL	−0.55	0.42
D-WASO	−0.51	0.30

time in bed so that large values for the first canonical variable for the standardized sleep outcomes are associated with less time spent in bed. The weight function for the first canonical variable for the log-spectrum of heart rate variability is positive at all values. Consequently, the first canonical variable for the log-spectrum is a measure of total power or total variance.

The estimated second canonical weight function for the standardized sleep variables is a contrast between the amount of time spent awake during the night, as measured by both diary and PSG, and the amount of time asleep during the night, as measured primarily by self-report diary. The estimated second canonical weight function for the log-spectrum is positive for frequencies less than 0.17 Hz and negative for frequencies greater than 0.17 Hz. Recall that our analysis is on the log-scale so that the estimated second canonical variable is a ratio on the natural scale of power from low frequencies to power from high frequencies. The two subjects whose data are displayed in Figure 1 and Table 1 exemplify this association. Subject 1 displays a larger ratio of power between low and high frequencies as compared to subject 2 with an estimated second canonical variable of 4.13 as compared to −0.11. All sleep variables for subject 1 are larger than those for subject 2 aside from diary assessed TST. Consequently, the estimated second canonical variable for subject 1, 0.38, is larger than that for subject 2, −0.14.

The adaptation of the functional CCA method of Eubank and Hsing (2008) that was explored in the simulation study was also implemented. The subject-specific log-spectra were estimated using the smoothing spline of Qin and Wang (2008), while the rank of the log-spectral covariance matrix was selected through CV_1 [He, Müller and Wang (2004), Section 2.5.1]. This procedure estimated the first canonical correlation as 46% and all higher order canonical correlations as zero. The estimated first weight functions for both the log-spectra and sleep outcomes were similar to the estimates obtained through the proposed cepstral-based procedure. However, this procedure estimated the second canonical correlation as zero and consequently did not produce estimates of the second canonical weight functions.

6.2. *Results*. The autonomic nervous system is classically divided into two dynamically balanced branches: the parasympathetic branch and the sympathetic branch. The parasympathetic branch is responsible for the maintenance of the body at rest, while the sympathetic branch is associated with the fight-or-flight response. Increased modulation of the sympathetic nervous system is associated with increased power in the spectrum of heart rate variability at low frequencies, while increased modulation of the parasympathetic nervous system is associated with increases in power at both low and high frequencies [Task Force of the ESC/ASPE (1996)].

The estimated first canonical variables suggest that less time in bed is associated with increased modulation of the parasympathetic nervous system. Excessive time spent in bed has been shown to be associated with increased mortality and has led to the advocacy of sleep restriction in older adults [Youngstedt and Kripke (2004)]. The causal pathway through which excessive time in bed is associated with mortality is unknown and identifying possible confounders and causal intermediates of this relationship to inform future studies is a topic of interest [Patel et al. (2006)]. Diminished parasympathetic nervous system activity while at rest has also been linked to mortality [Ponikowski et al. (1997), Lanza et al. (1998)]. The estimated first canonical correlation suggests that future studies might be able to illuminate the pathway through which time in bed is connected to mortality by exploring the role played by the modulation of the parasympathetic nervous system.

The estimated second canonical variable for sleep is a contrast between the time spent initiating and maintaining sleep relative to the amount of perceived sleep and may be viewed as a measure of nocturnal wake–sleep balance. Negative values represent less wakefulness relative to perceived sleep; this profile is observed in healthy individuals without clinical sleep disturbances [Walsleben et al. (2004)]. In contrast, positive wake–sleep balance values represent more wakefulness relative to perceived sleep, as often observed in individuals with sleep disturbances such as insomnia [Carskadon et al. (1976)]. The estimated second canonical variable for the log-spectrum is a measure of the sympathovagal balance. Increased sympathovagal balance during sleep has been shown to be associated with symptoms of depression and perceived stress [Hall et al. (2004), Hall et al. (2007)]. Consequently, the second canonical correlation suggests this simple one-dimensional measure of the wake–sleep cycle might be useful in informing studies to develop and evaluate behavioral therapies for improving the sleep of older adults.

7. Discussion. This article considered an approach to analyzing the association between the second-order spectrum of a time series and a set of static outcomes. The random Cramér representation provided a formal model for these data, while the cepstrum-based CCA provided an interpretable means of quantifying the association. This approach was motivated by and used to analyze the association between heart rate variability during sleep and measures of sleep duration and fragmentation in a population of adults who are the primary caregiver for their ill spouse. The analysis suggested a connection between stress and sleep which can serve as a guide for designing behavioral interventions to enhance the lives of caregivers.

The work presented in this article represents one of the first approaches to analyzing a collection of time series whose power spectra depend on a set of correlated outcomes and is by no means exhaustive. This article only considered time series that are second order stationary. Many studies which collect heart rate variability are interested in the time-dependent spectral properties of long-term epochs which are nonstationary [Task Force of the ESC/ASPE (1996)]. A topic of future research will be the extension of the random Cramér representation to the locally stationary setting through the use of a time-varying stochastic transfer function and the development of a time dependent cepstral coefficient-based CCA.

The CCA considered in this paper was used as a tool for exploratory analysis. One might also be interested in inference on the canonical correlations and weight functions. Theorem 2 of Dai and Guo (2004) provides a method for simulating a time series with a given smooth spectral density function. Another topic of future research is the development of this sampling method to formulate a bootstrap procedure for performing inference on the canonical correlations and weight functions.

APPENDIX: FISHER SCORING ALGORITHM

The cepstral estimates $\hat{\mathbf{f}}_j$ that minimize $\mathcal{L}_{jK}$ can be computed through Fisher scoring. To formulate the algorithm, define the K-vectors

$$\mathbf{C}_\ell = \{1, \sqrt{2}\cos(2\pi\ell/T), \dots, \sqrt{2}\cos[2\pi\ell(K-1)/T]\}',$$
$$\mathbf{f} = (f_0, \dots, f_{K-1})'$$

so that the negative log-Whittle likelihood for the jth subject can be written as

$$\mathcal{L}_{jK}(\mathbf{f}) = \sum_{\ell=1}^{\lfloor (T-1)/2 \rfloor} (Y_{j\ell} e^{-\mathbf{C}_\ell' \mathbf{f}} + \mathbf{C}_\ell' \mathbf{f}).$$

The algorithm is defined iteratively where the estimated cepstral coefficients for the jth subject in the $(m+1)$st iteration are

$$\hat{\mathbf{f}}_j^{m+1} = \hat{\mathbf{f}}_j^m + \mathbf{H}^{-1}(\hat{\mathbf{f}}_j^m)\mathbf{U}(\hat{\mathbf{f}}_j^m)$$

for score function

$$\mathbf{U}(\hat{\mathbf{f}}_j^m) = \frac{d\mathcal{L}_{jK}}{d\mathbf{f}}\Big|_{\mathbf{f}=\hat{\mathbf{f}}_j^m} = \sum_{\ell=1}^{\lfloor (T-1)/2 \rfloor} (1 - Y_{j\ell} e^{-\mathbf{C}_\ell \hat{\mathbf{f}}_j^m})\mathbf{C}_\ell$$

and Fisher information matrix

$$\mathbf{H}(\hat{\mathbf{f}}_j^m) = -\mathrm{E}\left(\frac{d^2\mathcal{L}_{jK}}{d\mathbf{f}d\mathbf{f}'}\Big|_{\hat{\mathbf{f}}_j^m}\right) = -\sum_{\ell=1}^{\lfloor (T-1)/2 \rfloor} \mathbf{C}_\ell \mathbf{C}_\ell'.$$

The algorithm continues until the change in the minimized negative log-Whittle likelihood is below some preselected threshold. We initialize the algorithm with the log-periodogram least squares estimators

$$\mathbf{f}_j^0 = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{L}_j,$$

where $\mathbf{L}_j = [\log(Y_{j1}) + \gamma, \ldots, \log(Y_{j\lfloor (T-1)/2 \rfloor}) + \gamma]'$ and $\mathbf{C}$ is the $\lfloor (T-1)/2 \rfloor \times K$ matrix with ℓth row $\mathbf{C}'_\ell$.

Acknowledgments. The authors would like to thank Daniel Buysse, Anne Germain, Sati Mazumdar, Timothy Monk, and Eric Nofzinger for their contributions to the AgeWise study. They would also like to express their appreciation to the editor, Tilmann Gneiting, an anonymous Associate Editor, and two anonymous referees whose comments and suggestions greatly increased the clarity and depth of the article.

SUPPLEMENTARY MATERIAL

Supplement A: Additional simulation results (DOI: 10.1214/12-AOAS601SUPPA; .pdf). The pdf file contains the results from a more comprehensive simulation study.

Supplement B: Matlab Code (DOI: 10.1214/12-AOAS601SUPPB; .zip). The zip file contains Matlab code to run the proposed CCA and a file demonstrating its use.

REFERENCES

BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226. MR0323048

BOGERT, B. P., HEALY, M. J. R. and TUKEY, J. W. (1963). The quefrency analysis of time series for echoes: Cepstrum, pseudo autocovariance, cross-cepstrum and saphe cracking. In *Proceedings of the Symposium on Time Series Analysis* (M. Rosenblatt, ed.) 209–243. Wiley, New York.

BRILLINGER, D. R. (2001). *Time Series*: *Data Analysis and Theory*. *Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. Reprint of the 1981 edition. MR1853554

BUYSSE, D. J., GERMAIN, A., HALL, M., MOUL, D. E., NOFZINGER, E. A., BEGLEY, A., EHLER, C. L., THOMPSON, W. and KUPFER, D. J. (2008). EEG spectral analysis in primary insomnia: NREM period effects and sex differences. *Sleep* **31** 1673–1682.

CARSKADON, M. A., DEMENT, W. C., MITLER, M. M., GUILLEMINAULT, C., ZARCONE, V. P. and SPIEGEL, R. (1976). Self-reports versus sleep laboratory findings in 122 drug-free subjects with complaints of chronic insomnia. *Am. J. Psychiatry* **133** 1382–1388.

DAI, M. and GUO, W. (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika* **91** 629–643. MR2090627

DIGGLE, P. J. and AL WASEL, I. (1997). Spectral analysis of replicated biomedical time series. *J. Roy. Statist. Soc. Ser. C* **46** 31–71. MR1452286

EUBANK, R. L. and HSING, T. (2008). Canonical correlation for stochastic processes. *Stochastic Process. Appl.* **118** 1634–1661. MR2442373

FOKIANOS, K. and SAVVIDES, A. (2008). On comparing several spectral densities. *Technometrics* **50** 317–331. MR2528655

GRONFIER, C. and BRANDENBERGER, G. (1998). Ultradian rhythms in pituitary and adrenal hormones: Their relations to sleep. *Sleep Med. Rev.* **2** 17–29.

HALL, M., VASKO, R., BUYSSE, D. J., OMBAO, H., CHEN, Q., CASHMERE, J. D., KUPFER, D. J. and THAYER, J. F. (2004). Acute stress affects heart rate variability during sleep. *Psychosomatic Medicine* **66** 56–62.

HALL, M., THAYER, J. F., GERMAIN, A., MOUL, D., VASKO, R., PUHL, M., MIEWALD, J. and BUYSSE, D. J. (2007). Psychological stress is associated with heightened physiological arousal during NREM sleep in primary insomnia. *Behav. Sleep Med.* **5** 178–193.

HALL, M. H., MULDOON, M. F., JENNINGS, J. R., BUYSSE, D. J., FLORY, J. D. and MANUCK, S. B. (2008). Self-reported sleep duration is associated with the metabolic syndrome in midlife adults. *Sleep* **31** 635–643.

HE, G., MÜLLER, H.-G. and WANG, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multivariate Anal.* **85** 54–77. MR1978177

HE, G., MÜLLER, H.-G. and WANG, J.-L. (2004). Methods of canonical analysis for functional data. *J. Statist. Plann. Inference* **122** 141–159. MR2057919

JOHNSON, R. A. and WICHERN, D. W. (2002). *Applied Multivaraite Statistical Analysis*. Prentice Hall, Upper Saddle River, NJ.

KRAFTY, R. T., HALL, M. and GUO, W. (2011). Functional mixed effects spectral analysis. *Biometrika* **98** 583–598. MR2836408

KRAFTY, R. T. and HALL, M. (2013a). Supplement to "Canonical correlation analysis between time series and static outcomes, with application to the spectral analysis of heart rate variability." DOI:10.1214/12-AOAS601SUPPA.

KRAFTY, R. T. and HALL, M. (2013b). Supplement to "Canonical correlation analysis between time series and static outcomes, with application to the spectral analysis of heart rate variability." DOI:10.1214/12-AOAS601SUPPB.

LANZA, G., GUIDO, V., GALEAZZI, M., MUSTILLI, M., NATALI, R., IERARDI, C., MILICI, C., BURZOTTA, F., PASCERI, V., TOMASSINI, F., LUPI, A. and MASERI, A. (1998). Prognostic role of heart rate variability in patients with recent acute myocardial infarction. *American Journal of Caridology* **82** 1323–1328.

LEURGANS, S. E., MOYEED, R. A. and SILVERMAN, B. W. (1993). Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc. Ser. B* **55** 725–740. MR1223939

MCCALL, W. V., TURPIN, E., REBOUSSIN, D., EDINGER, J. D. and HAPONIK, E. F. (1995). Subjective estimates of sleep differ from polysomnographic measurements in obstructive sleep apnea patients. *Sleep* **18** 646–650.

MCCURRY, S. M., LOGSDON, R. G., TERI, L. and VITIELLO, M. V. (2007). Sleep disturbances in caregivers of persons with dementia: Contributing factors and treatment implications. *Sleep Med. Rev.* **11** 143–153.

NOCK, N. L., LI, L., LARKIN, E. K., PATEL, S. R. and REDLINE, S. (2009). Empirical evidence for "syndrome Z": A hierarchical 5-factor model of the metabolic syndrome incorporating sleep disturbance measures. *Sleep* **32** 615–622.

PATEL, S., MALHOTRA, A., GOTTLIEB, D. J., WHITE, D. P. and HU, F. B. (2006). Correlates of long sleep duration. *Sleep* **29** 881–889.

PONIKOWSKI, P., ANKER, S. D., CHUA, T. P., SZELEMEJ, R., PIEPOLI, M., ADAMOPOULOS, S., WEBB-PEPLOE, K., HARRINGTON, D., BANASIAK, W., WRABEC, K. and COATS, A. J. (1997). Depressed heart rate variability as an independent predictor of death in chronic congestive heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. *Am. J. Cardiol.* **79** 1645–1650.

QIN, L. and WANG, Y. (2008). Nonparametric spectral analysis with applications to seizure characterization using EEG time series. *Ann. Appl. Stat.* **2** 1432–1451. MR2655666

SHUMWAY, R. H. (1971). On detecting a signal in N stationarily correlated noise series. *Technometrics* **13** 499–519.

SILVA, G. E., GOODWIN, J. L., SHERRILL, D. L., ARNOLD, J. L., BOOTZIN, R. R., SMITH, T., WALSLEBEN, J. A., BALDWIN, C. M. and QUAN, S. F. (2007). Relationship between reported and measured sleep times: The sleep heart health study (SHHS). *J. Clin. Sleep Med.* **3** 622–630.

STOFFER, D. S., HAN, S., QIN, L. and GUO, W. (2010). Smoothing spline ANOPOW. *J. Statist. Plann. Inference* **140** 3789–3796. MR2674166

TASK FORCE OF THE ESC/ASPE (1996). Heart rate variability—standards of measurement, physiological interpretation, and clinical use. *Circulation* **93** 1043–1065.

VGONTZAS, A. N., LIAO, D., PEJOVIC, S., CALHOUN, S., KARATARAKI, M., BASTA, M., FERNÁNDEZ-MENDOZA, J. and BIXLER, E. O. (2010). Insomnia with short sleep duration and mortality: The penn state cohort. *Sleep* **33** 1159–1164.

WALSLEBEN, J. A., KAPUR, V. K., NEWMAN, A. B., SHAHAR, E., BOOTZIN, R. R., ROSENBERG, C. E., O'CONNOR, G. and NIETO, F. J. (2004). Sleep and reported daytime sleepiness in normal subjects: The sleep heart health study. *Sleep* **27** 293–298.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434. MR0060797

WHITTLE, P. (1954). Some recent contributions to the theory of stationary processes. *A Study in the Analysis of Stationary Time Series* **2** 196–228.

YOUNGSTEDT, S. D. and KRIPKE, D. F. (2004). Long sleep and mortality: Rationale for sleep restriction. *Sleep Med. Rev.* **8** 159–174.

DEPARTMENT OF STATISTICS
UNIVERSITY OF PITTSBURGH
2702 CATHEDRAL OF LEARNING
PITTSBURGH, PENNSYLVANIA 15260
USA
E-MAIL: krafty@pitt.edu

DEPARTMENT OF PSYCHIATRY
UNIVERSITY OF PITTSBURGH
3811 O'HARA STREET
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: hallmh@upmc.edu