The Spectral Envelope and Its Applications

David S. Stoffer, David E. Tyler and David A. Wendt

Abstract. The concept of the spectral envelope was recently introduced as a statistical basis for the frequency domain analysis and scaling of qualitative-valued time series. In the process of developing the spectral envelope methodology, many other interesting extensions became evident. In this article we explain the basic concept and give numerous examples of the usefulness of the technology. These examples include analyses of DNA sequences, finding optimal transformations for the analysis of real-valued time series, residual analysis, detecting common signals in many time series, and the analysis of textures.

Key words and phrases: Spectral envelope, optimal scaling, Fourier analysis, latent roots and vectors, principal components, canonical correlation, signal detection, optimal transformations, coherency, random fields, categorical-valued time series, EEG sleep states, DNA, US GNP growth rate, residual analysis, long range dependence, matching sequences, functional magnetic resonance imaging (fMRI), pain perception, textures, image retrival.

The reader will note that the various statistics presented are immediate functions of the discrete Fourier transforms of the observed values of the time series.... . The discrete Fourier transform is given such prominence because it has important empirical and mathematical properties. Following the work of Cooley and Tukey (1965), it may be computed rapidly. The definitions, procedures, techniques, and statistics discussed are, in many cases, simple extensions of existing...multivariate analysis techniques. This pleasant state of affairs is indicative of the widely pervasive nature of the important statistical and data analytic procedures. David Brillinger (1975, page viii)

1. INTRODUCTION

The concept of spectral envelope for the *spectral analysis* and *scaling of categorical time series* was first introduced in Stoffer, Tyler and McDougall (1993). Since then the idea has been extended in various directions (not only restricted to categorical

David S. Stoffer is Professor, Department of Statistics, University of Pittsburgh, 2630 Cathedral of Learning, Pittsburgh, Pennsylvania 15260. David E. Tyler is Professor, Department of Statistics, Rutgers University, Hill Center, Busch Campus, New Brunswick, New Jersey 08903. David A. Wendt is Visiting Assistant Professor, Ohio State University, Columbus, Ohio 43210.

time series) and we will explore these problems as well. Most of the reference material is reserved for the discussion in Section 7. First, we give a brief introduction to spectral analysis and the concept of scaling.

1.1 Spectral Analysis

Briefly, spectral analysis has to do with partitioning the variance of a stationary time series, $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$, into components of oscillation indexed by frequency ω, and measured in cycles per unit of time, for $-1/2 \leq \omega \leq 1/2$. Given a numerical-valued time series sample, X_t, $t = 1, \ldots, n$, that has been centered by its sample mean, the sample spectral density (or *periodogram*) is defined in terms of frequency ω,

$$
I_n(\omega) = \left| n^{-1/2} \sum_{t=1}^{n} X_t \exp(-2\pi i \omega t) \right|^2
$$

$$
(1.1) \qquad = n^{-1} \left| \sum_{t=1}^{n} X_t \cos(2\pi \omega t) \right|^2
$$

$$
+ n^{-1} \left| \sum_{t=1}^{n} X_t \sin(2\pi \omega t) \right|^2.
$$

The periodogram is essentially the squared-correlation of the data with sines and cosines that oscillate at frequency ω.

The spectral density, $f(\omega)$, of a stationary time series can be defined as the limit ($n \to \infty$) of $E[I_n(\omega)]$, provided that the limit exists; details can be found

in many of the time series texts listed in Section 7. It is worthwhile to note that $f(\omega) \geq 0$, $f(\omega) = f(-\omega)$ and

$$(1.2) \quad \int_{-1/2}^{1/2} f(\omega)\,d\omega = 2\int_{0}^{1/2} f(\omega)\,d\omega = \sigma^2,$$

where $\sigma^2 = \operatorname{var}(X_t) < \infty$. Thus, the spectral density can be thought of as the variance density of a time series relative to frequency of oscillation. That is, for positive frequencies between 0 and 1/2, the proportion of the variance that can be attributed to oscillations in the data at frequency ω is roughly $2f(\omega)\,d\omega$. If the time series X_t is *white noise*, that is, $E(X_t)$ is independent of time t, and $\operatorname{cov}(X_s, X_t) = 0$ for all $s \neq t$, then $f(\omega) \equiv \sigma^2$. The designation white originates from the analogy with white light and indicates that all possible periodic oscillations are present with equal strength.

If n is a highly composite integer, the fast Fourier transform provides for extremely fast calculation of $I_n(j/n)$, for $j = 1, 2, \ldots, [n/2]$, where $[n/2]$ is the greatest integer less than or equal to $n/2$. The frequencies $\omega_j = j/n$ are called the fundamental (or Fourier) frequencies. The sample equivalent of the integral equation (1.2) is

$$(1.3) \quad 2\sum_{j=1}^{[(n-1)/2]} n^{-1} I_n(j/n) + n^{-1} I_n(1/2) = s^2,$$

where s^2 is the sample variance of the data; the last term is dropped if n is odd. One usually plots the periodogram, $I_n(\omega_j)$, versus the fundamental frequencies $\omega_j = j/n$, for $j = 1, 2, \ldots, [n/2]$, and searches the graph for peaks. As previously mentioned, large values of the periodogram at ω_j indicate that the data are highly correlated with the sinusoid that is oscillating at a frequency of j cycles in n observations.

As a simple example, Figure 1 shows a time plot of 128 observations generated by

$$(1.4) \quad X_t = \cos(2\pi[\omega_0 t + \phi]) + \varepsilon_t,$$
$$t = 1, \ldots, 128,$$

where $\omega_0 = 13/128$ is the frequency of oscillation, $\phi = 10/128$ is a phase shift, and $\varepsilon_t \sim$ iid N(0, 1). Figure 2 shows the standardized periodogram, $I_n(\omega)/s^2$, of the data shown in Figure 1. Note that there is a large value of the periodogram at $\omega = 13/128$ and small values elsewhere [if there were no noise in (1.4) then the periodogram would only be nonzero at $\omega = 13/128$].

The periodogram is not consistent for the spectral density. To overcome this problem, one typically smooths the periodogram. This problem has been thoroughly studied and we will only discuss this matter briefly. Interested readers can see Brillinger

(1975, Chapter 7), Hannan (1970, Chapter 5) or one of the other texts listed in Section 7. One technique for smoothing is to take a symmetric moving average of the periodogram, that is,

$$(1.5) \quad \widehat{f}(\omega_j) = \sum_{q=-m}^{m} h_q I_n(\omega_{j+q}),$$

where the weights are chosen so that $h_q = h_{-q}$ are positive and $\sum_{q=-m}^{m} h_q = 1$. A simple average corresponds to the case where $h_q = 1/(2m+1)$ for $q = -m, \ldots, 0, \ldots, m$. The number m is chosen to obtain a desired degree of smoothness. Larger values of m lead to smoother estimates, but one has to be careful not to smooth away significant peaks (this type of situation has been called leakage because the power from one frequency leaks into another frequency). For $\widehat{f}(\omega)$ to be consistent under general conditions, the weights must satisfy $\sum h_q^2 \to 0$ as $m \to \infty$, but $m/n \to 0$ as $n \to \infty$.

Another related approach is window spectral estimation. Specifically, consider a window function $H(\alpha)$, $-\infty < \alpha < \infty$, that is real-valued, even, of bounded variation, with $\int_{-\infty}^{\infty} H(\alpha)\,d\alpha = 1$, and $\int_{-\infty}^{\infty} |H(\alpha)|\,d\alpha < \infty$. The window spectral estimator is

$$(1.6) \quad \widehat{f}(\omega) = n^{-1} \sum_{q=1}^{n-1} H_n(\omega - q/n) I_n(q/n),$$

where $H_n(\alpha) = B_n^{-1} \sum_{j=-\infty}^{\infty} H(B_n^{-1}[\alpha + j])$ and B_n is a bounded sequence of nonnegative scale parameters such that $B_n \to 0$ and $nB_n \to \infty$ as $n \to \infty$.

Estimation of the spectral density requires special attention to the issues of leakage and of the variance-bias tradeoff typically associated with the estimation of density functions. Readers who are unfamiliar with this material can consult one of the many texts on the spectral domain analysis of time series listed in Section 7.

An analogous theory applies if one collects k numerical-valued time series, say $X_{1t}, \ldots, X_{kt}$, for $t = 1, \ldots, n$. In this case, write $\mathbf{X}_t = (X_{1t}, \ldots, X_{kt})'$ as the $k \times 1$ column vector of data. The periodogram is now a $k \times k$ complex matrix,

$$(1.7) \quad I_n(\omega) = \left[n^{-1/2} \sum_{t=1}^{n} \mathbf{X}_t \exp(-2\pi it\omega) \right]$$
$$\times \left[n^{-1/2} \sum_{t=1}^{n} \mathbf{X}_t \exp(-2\pi it\omega) \right]^*$$

where * means to transpose and conjugate. Smoothing the periodogram also proceeds analogously to the univariate case; for example, in (1.5) and (1.6) one would simply replace the univariate periodogram with the multivariate one given in (1.7). The population spectral density matrix, $f(\omega)$, can
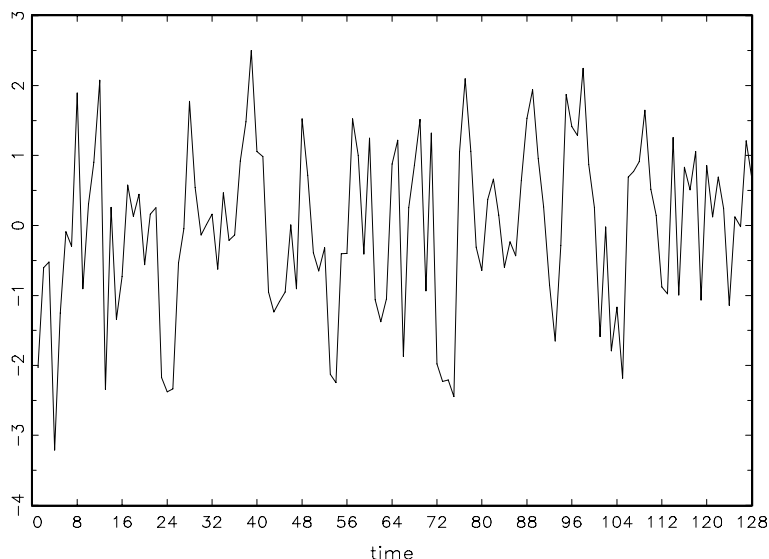
FIG. 1.   *Time series generated from* (1.4).

again be defined as the limit as n tends to infinity of $E[I_n(\omega)]$. The spectral matrix $f(\omega)$ is Hermitian $[f^*(\omega) = f(\omega)]$ and nonnegative definite. The diagonal elements of $f(\omega)$, say $f_{ii}(\omega)$, for $i = 1, \ldots, k$, are the individual spectra and the off-diagonal elements, say $f_{ij}(\omega)$, for $i \neq j = 1, \ldots, k$, are related to the pairwise dependence structure among the k sequences (these are called *cross-spectra*). A frequency-based measure of the cross-correlation

between the series X_{it} and X_{jt} is the *squared-coherency* given by

$$(1.8) \qquad \rho_{ij}^2(\omega) = \frac{|f_{ij}(\omega)|^2}{f_{ii}(\omega)f_{jj}(\omega)}.$$
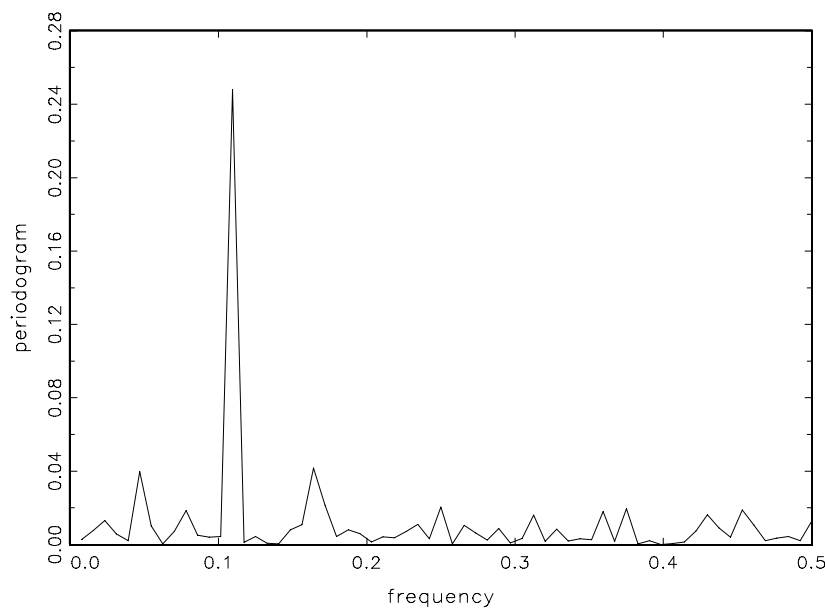

FIG. 2.   *Periodogram of data generated from* (1.4).

TABLE 1
Infant EEG sleep states (per minute) (read down and across)

ah	qt	qt	al	tr	qt	al	ah
ah	qt	qt	ah	tr	qt	al	ah
ah	qt	tr	ah	tr	qt	al	ah
ah	qt	al	ah	qh	qt	al	ah
ah	qt	al	ah	qh	qt	al	ah
ah	tr	al	ah	qt	qt	al	ah
ah	qt	al	ah	qt	qt	al	ah
ah	qt	al	ah	qt	qt	al	ah
tr	qt	tr	tr	qt	qt	al	tr
ah	qt	ah	tr	qt	tr	al	
tr	qt	al	ah	qt	al	al	
ah	qt	al	ah	qt	al	al	
ah	qt	al	ah	qt	al	al	
qh	qt	al	ah	qt	al	ah	

1.2 Scaling Categorical Time Series

Our work on the spectral envelope was motivated by collaborations with researchers who collected categorical-valued time series with an interest in the cyclic behavior of the data. For example, Table 1 shows the per minute sleep-state of an infant taken from a study on the effects of prenatal exposure to alcohol. Details can be found in Stoffer et al. (1988), but briefly, an electroencephalographic (EEG) sleep recording of approximately two hours is obtained on a full term infant 24 to 36 hours after birth, and the recording is scored by a pediatric neurologist for sleep state. Sleep state is categorized, per minute, into one of six possible states: qt: quiet sleep, trace alternant; qh: quiet sleep, high voltage; tr: transitional sleep; al: active sleep, low voltage; ah: active sleep, high voltage and aw: awake. This particular infant was never awake during the study.

It is not too difficult to notice a pattern in the data if one concentrates on active versus quiet sleep (that is, focus on the first letter). But it would be difficult to try to assess patterns in a longer sequence, or if there were more categories, without some graphical aid. One simple method would be to *scale* the data, that is, *assign numerical values to the categories* and then draw a time plot of the scales. Since the states have an order, one obvious scaling is

(1.9)
$$qt = 1 \quad qh = 2 \quad tr = 3 \quad al = 4$$
$$ah = 5 \quad aw = 6,$$

and Figure 3 shows the time plot using this scaling. Another interesting scaling might be to combine the quiet states and the active states,

(1.10)
$$qt = 1 \quad qh = 1 \quad tr = 2 \quad al = 3$$
$$ah = 3 \quad aw = 4.$$

The time plot using (1.10) would be similar to Figure 3 as far as the cyclic (in and out of quiet sleep) behavior of this infant's sleep pattern. Figure 4 shows the periodogram of the sleep data using the scaling in (1.9). Note that there is a large peak at the frequency corresponding to 1 cycle every 60 minutes. As one might imagine, the general appearance of the periodogram using the scaling (1.10) (not shown) is similar to Figure 4. Most of us would feel comfortable with this analysis even though we made an arbitrary and ad hoc choice about the particular scaling. It is evident from the data (without any scaling) that if the interest is in infant sleep cycling, this particular sleep study indicates that an infant cycles between active and quiet sleep at a rate of about one cycle per hour.

The intuition used in the previous example is lost when one considers a long DNA sequence. Briefly, a DNA strand can be viewed as a long string of linked nucleotides. Each nucleotide is composed of a nitrogenous base, a five-carbon sugar, and a phosphate group. There are four different bases that can be grouped by size: the pyrimidines, thymine (T) and cytosine (C), and the purines, adenine (A) and guanine (G). The nucleotides are linked together by a backbone of alternating sugar and phosphate groups with the 5′ carbon of one sugar linked to the 3′ carbon of the next, giving the string direction. DNA molecules occur naturally as a double helix composed of polynucleotide strands with the bases facing inwards. The two strands are complementary, so it is sufficient to represent a DNA molecule by a sequence of bases on a single strand. Thus, a strand of DNA can be represented as a sequence of letters, termed base pairs (*bp*), from the finite alphabet {A, C, G, T}. The order of the nucleotides contains the genetic information specific to the organism. Expression of information stored in these molecules is a complex multistage process. One important task is to translate the information stored in the protein-coding sequences (CDS) of the DNA. A common problem in analyzing long DNA sequence data is in identifying CDS that are dispersed throughout the sequence and separated by regions of noncoding (which makes up most of the DNA). Table 2 shows part of the Epstein–Barr virus (EBV) DNA sequence. The entire EBV DNA sequence consists of approximately 172,000 bp.

One could try scaling according to the pyrimidine-purine alphabet, that is, A = G = 0 and C = T = 1, but this is not necessarily of interest for every CDS of EBV. There are numerous possible alphabets of interest; for example, one might focus on the strong–weak hydrogen bonding alphabet C = G = 0 and A = T = 1. While model calculations as well as experimental data strongly agree that some kind of periodic signal exists in certain DNA sequences, there
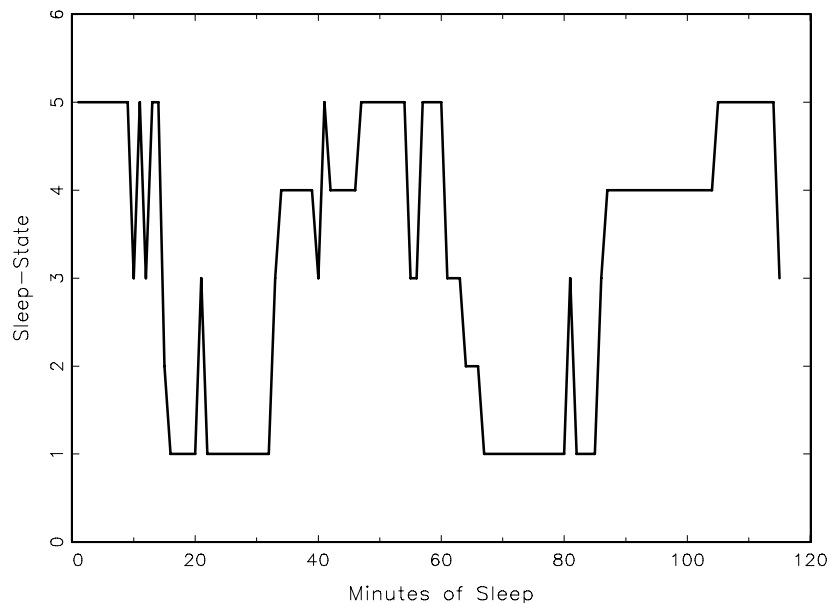
FIG. 3.   *Time plot of the EEG sleep state data in Table* 1 *using the scaling in* (1.9).

is a large disagreement about the exact type of periodicity. In addition, there is disagreement about which nucleotide alphabets are involved in the signals (e.g., compare Ioshikhes, Bolshoy and Trifonov, 1992 with Satchwell, Drew and Travers, 1986); more details are provided in the discussion in Section 7.

If we consider the naive approach of arbitrarily assigning numerical values (scales) to the categories and then proceeding with a spectral analysis, the result will depend on the particular assignment of numerical values. For example, consider the artificial sequence ACGTACGTACGT.... Then, setting A =
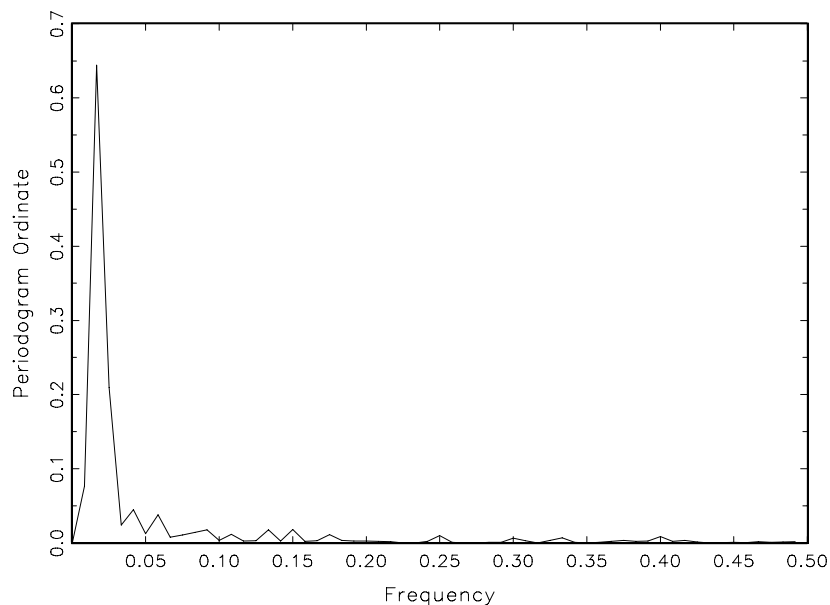

FIG. 4.   *Periodogram of the EEG sleep state data in Figure* 1 *based on the scaling in* (1.9). *The peak corresponds to a frequency of approximately one cycle every* 60 *minutes.*

TABLE 2
Part of the Epstein-Barr virus DNA sequence (read across and down)

AGAATTCGTC	TTGCTCTATT	CACCCTTACT	TTTCTTCTTG	CCCGTTCTCT	TTCTTAGTAT
GAATCCAGTA	TGCCTGCCTG	TAATTGTTGC	GCCCTACCTC	TTTTGGCTGG	CGGCTATTGC
CGCCTCGTGT	TTCACGGCCT	CAGTTAGTAC	CGTTGTGACC	GCCACCGGCT	TGGCCCTCTC
ACTTCTACTC	TTGGCAGCAG	TGGCCAGCTC	ATATGCCGCT	GCACAAAGGA	AACTGCTGAC
ACCGGTGACA	GTGCTTACTG	CGGTTGTCAC	TTGTGAGTAC	ACACGCACCA	TTTACAATGC
ATGATGTTCG	TGAGATTGAT	CTGTCTCTAA	CAGTTCACTT	CCTCTGCTTT	TCTCCTCAGT
CTTTGCAATT	TGCCTAACAT	GGAGGATTGA	GGACCCACCT	TTTAATTCTC	TTCTGTTTGC
ATTGCTGGCC	GCAGCTGGCG	GACTACAAGG	CATTTACGGT	TAGTGTGCCT	CTGTTATGAA
ATGCAGGTTT	GACTTCATAT	GTATGCCTTG	GCATGACGTC	AACTTTACTT	TTATTTCAGT
TCTGGTGATG	CTTGTGCTCC	TGATACTAGC	GTACAGAAGG	AGATGGCGCC	GTTTGACTGT
TTGTGGCGGC	ATCATGTTTT	TGGCATGTGT	ACTTGTCCTC	ATCGTCGACG	CTGTTTTGCA
GCTGAGTCCC	CTCCTTGGAG	CTGTAACTGT	GGTTTCCATG	ACGCTGCTGC	TACTGGCTTT
CGTCCTCTGG	CTCTCTTCGC	CAGGGGGCCT	AGGTACTCTT	GGTGCAGCCC	TTTTAACATT
GGCAGCAGGT	AAGCCACACG	TGTGACATTG	CTTGCCTTTT	TGCCACATGT	TTTCTGGACA
CAGGACTAAC	CATGCCATCT	CTGATTATAG	CTCTGGCACT	GCTAGCGTCA	CTGATTTTGG
GCACACTTAA	CTTGACTACA	ATGTTCCTTC	TCATGCTCCT	ATGGACACTT	GGTAAGTTTT
CCCTTCCTTT	AACTCATTAC	TTGTTCTTTT	GTAATCGCAG	CTCTAACTTG	GCATCTCTTT
TACAGTGGTT	CTCCTGATTT	GCTCTTCGTG	CTCTTCATGT	CCACTGAGCA	AGATCCTTCT
GGCACGACTG	TTCCTATATG	CTCTCGCACT	CTTGTTGCTA	GCCTCCGCGC	TAATCGCTGG
TGGCAGTATT	TTGCAAACAA	ACTTCAAGAG	TTTAAGCAGC	ACTGAATTTA	TACCCAGTGA

$G = 0$ and $C = T = 1$ yields the numerical sequence $010101010101\ldots$, or one cycle every two base pairs ($\omega = 1/2$). Another interesting scaling is $A = 1$, $C = 2$, $G = 3$ and $T = 4$, which results in the sequence $123412341234\ldots$, or one cycle every four bp ($\omega = 1/4$). In this example, both scalings, $\{A, C, G, T\} = \{0, 1, 0, 1\}$ and $\{A, C, G, T\} = \{1, 2, 3, 4\}$, of the nucleotides are interesting and bring out different properties of the sequence. It should be clear, then, that one does not want to focus on only one scaling. Instead, the focus should be on finding scalings that bring out all of the interesting features in the data. Rather than choose values arbitrarily, the spectral envelope approach selects scales that help emphasize any periodic feature that exists in a categorical time series of virtually any length in a quick and automated fashion. In addition, the technique can help in determining whether a sequence is merely a random assignment of categories.

2. THE SPECTRAL ENVELOPE FOR CATEGORICAL TIME SERIES

As a general description, the spectral envelope is a frequency-based, principal components technique applied to a multivariate time series. In this section we will focus on the basic concept and its use in the analysis of categorical time series. Technical details can be found in Stoffer, Tyler and McDougall (1993), and the relationship of the spectral envelope to other established techniques will be discussed in Section 7.

Briefly, in establishing the spectral envelope for categorical time series, we addressed the basic question of how to efficiently discover periodic components in categorical time series. This was accomplished via nonparametric spectral analysis as follows. Let X_t, $t = 0, \pm 1, \pm 2, \ldots$, be a categorical-valued time series with finite state-space $\mathscr{C} = \{c_1, c_2, \ldots, c_k\}$. Assume that X_t is stationary and $p_j = \text{pr}\{X_t = c_j\} > 0$ for $j = 1, 2, \ldots, k$. For $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)' \in \mathbf{R}^k$, denote by $X_t(\boldsymbol{\beta})$ the real-valued stationary time series corresponding to the scaling that assigns the category c_j the numerical value β_j, $j = 1, 2, \ldots, k$. Our goal was to find scalings $\boldsymbol{\beta}$ so that the spectral density is in some sense interesting and to summarize the spectral information by what we called the spectral envelope.

We chose $\boldsymbol{\beta}$ to maximize the power (variance) at each frequency ω, across frequencies $\omega \in (-1/2, 1/2]$, relative to the total power $\sigma^2(\boldsymbol{\beta}) = \text{var}\{X_t(\boldsymbol{\beta})\}$. That is, we chose $\boldsymbol{\beta}(\omega)$, at each ω of interest, so that

$$(2.1) \qquad \lambda(\omega) = \sup_{\boldsymbol{\beta}} \left\{ \frac{f(\omega; \boldsymbol{\beta})}{\sigma^2(\boldsymbol{\beta})} \right\},$$

over all $\boldsymbol{\beta}$ not proportional to $\mathbf{1}_k$, the $k \times 1$ vector of ones. Note that $\lambda(\omega)$ is not defined if $\boldsymbol{\beta} = a\mathbf{1}_k$ for $a \in \mathbf{R}$ because such a scaling corresponds to assigning each category the same value a; in this case $f(\omega; \boldsymbol{\beta}) \equiv 0$ and $\sigma^2(\boldsymbol{\beta}) = 0$. The optimality criterion $\lambda(\omega)$ possesses the desirable property of being invariant under location and scale changes of $\boldsymbol{\beta}$.

As in most scaling problems for categorical data, it was useful to represent the categories in terms of the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$, where $\mathbf{e}_j$ represents the $k \times 1$ vector with a one in the jth row and zeros elsewhere. We then defined a k-dimensional stationary time series $\mathbf{Y}_t$ by $\mathbf{Y}_t = \mathbf{e}_j$ when $X_t = c_j$. The time

series $X_t(\boldsymbol{\beta})$ can be obtained from the $\mathbf{Y}_t$ time series by the relationship $X_t(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_t$. Assume that the vector process $\mathbf{Y}_t$ has a continuous spectral density denoted by $f_Y(\omega)$. For each ω, $f_Y(\omega)$ is, of course, a $k \times k$ complex-valued Hermitian matrix. Note that the relationship $X_t(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_t$ implies that $f_Y(\omega; \boldsymbol{\beta}) = \boldsymbol{\beta}'f_Y(\omega)\boldsymbol{\beta} = \boldsymbol{\beta}'f_Y^{\mathrm{re}}(\omega)\boldsymbol{\beta}$, where $f_Y^{\mathrm{re}}(\omega)$ denotes the real part of $f_Y(\omega)$. The optimality criterion can thus be expressed as

$$(2.2) \qquad \lambda(\omega) = \sup_{\beta}\left\{\frac{\boldsymbol{\beta}'f_Y^{\mathrm{re}}(\omega)\boldsymbol{\beta}}{\boldsymbol{\beta}'V\boldsymbol{\beta}}\right\},$$

where V is the variance-covariance matrix of $\mathbf{Y}_t$. The resulting scaling $\boldsymbol{\beta}(\omega)$ is called the optimal scaling.

The $\mathbf{Y}_t$ process is a multivariate point process, and any particular component of $\mathbf{Y}_t$ is the individual point process for the corresponding state (for example, the first component of $\mathbf{Y}_t$ indicates whether or not the process is in state c_1 at time t). For any fixed t, $\mathbf{Y}_t$ represents a single observation from a simple multinomial sampling scheme. It readily follows that $V = D - \mathbf{p}\mathbf{p}'$, where $\mathbf{p} = (p_1, \ldots, p_k)'$, and D is the $k \times k$ diagonal matrix $D = \mathrm{diag}\{p_1, \ldots, p_k\}$. Since, by assumption, $p_j > 0$ for $j = 1, 2, \ldots, k$, it follows that $\mathrm{rank}(V) = k - 1$ with the null space of V being spanned by $\mathbf{1}_k$. For any $k \times (k-1)$ full rank matrix Q whose columns are linearly independent of $\mathbf{1}_k$, $Q'VQ$ is a $(k-1) \times (k-1)$ positive definite symmetric matrix.

With the matrix Q as previously defined, and for $-1/2 < \omega \leq 1/2$, define $\lambda(\omega)$ to be the largest eigenvalue of the determinantal equation

$$\left|Q'f_Y^{\mathrm{re}}(\omega)Q - \lambda Q'VQ\right| = 0,$$

and let $\mathbf{b}(\omega) \in \mathbf{R}^{k-1}$ be any corresponding eigenvector; that is,

$$Q'f_Y^{\mathrm{re}}(\omega)Q\mathbf{b}(\omega) = \lambda(\omega)Q'VQ\mathbf{b}(\omega).$$

The eigenvalue $\lambda(\omega) \geq 0$ does not depend on the choice of Q. Although the eigenvector $\mathbf{b}(\omega)$ depends on the particular choice of Q, the equivalence class of scalings associated with $\boldsymbol{\beta}(\omega) = Q\mathbf{b}(\omega)$ does not depend on Q. A convenient choice of Q is $Q = [I_{k-1} \mid \mathbf{0}]'$, where I_{k-1} is the $(k-1) \times (k-1)$ identity matrix and $\mathbf{0}$ is the $(k-1) \times 1$ vector of zeros. For this choice, $Q'f_Y^{\mathrm{re}}(\omega)Q$ and $Q'VQ$ are the upper $(k-1) \times (k-1)$ blocks of $f_Y^{\mathrm{re}}(\omega)$ and V, respectively. This choice corresponds to setting the last component of $\boldsymbol{\beta}(\omega)$ to zero.

The value $\lambda(\omega)$ itself has a useful interpretation; specifically, $\lambda(\omega)d\omega$ represents the largest proportion of the total power that can be attributed to the frequencies $\omega d\omega$ for any particular scaled process $X_t(\boldsymbol{\beta})$, with the maximum being achieved by the scaling $\boldsymbol{\beta}(\omega)$. Because of its central role, $\lambda(\omega)$ was

defined to be the *spectral envelope* of a stationary categorical time series.

The name "spectral envelope" is appropriate since $\lambda(\omega)$ envelopes the standardized spectrum of any scaled process. That is, given any $\boldsymbol{\beta}$ normalized so that $X_t(\boldsymbol{\beta})$ has total power one, $f(\omega; \boldsymbol{\beta}) \leq \lambda(\omega)$ with equality if and only if $\boldsymbol{\beta}$ is proportional to $\boldsymbol{\beta}(\omega)$.

Although the law of the process $X_t(\boldsymbol{\beta})$ for any one-to-one scaling $\boldsymbol{\beta}$ completely determines the law of the categorical process X_t, information is lost when one restricts attention to the spectrum of $X_t(\boldsymbol{\beta})$. Less information is lost when one considers the spectrum of $\mathbf{Y}_t$. Dealing directly with the spectral density $f_Y(\omega)$ itself is somewhat cumbersome since it is a function into the set of complex Hermitian matrices. Alternatively, one can view the spectral envelope as an easily understood, parsimonious tool for exploring the periodic nature of a categorical time series with a minimal loss of information.

If we observe a finite realization of the stationary categorical time series X_t, or equivalently, the multinomial point process $\mathbf{Y}_t$, $t = 1, \ldots, n$, the theory for estimating the spectral density of a multivariate, real-valued time series is well established (as discussed briefly in Section 1.1) and can be applied to estimating $f_Y(\omega)$, the spectral density of $\mathbf{Y}_t$. Given an estimate $\widehat{f}_Y(\omega)$ of $f_Y(\omega)$, estimates $\widehat{\lambda}(\omega)$ and $\widehat{\boldsymbol{\beta}}(\omega)$ of the spectral envelope, $\lambda(\omega)$, and the corresponding scalings, $\boldsymbol{\beta}(\omega)$, can then be obtained. Details on estimation and inference for the sample spectral envelope and the optimal scalings can be found in Stoffer, Tyler and McDougall (1993), but the main result of that paper is as follows: If $\widehat{f}_Y(\omega)$ is a consistent spectral estimator and if for each $j = 1, \ldots, J$, the largest root of $f_Y^{\mathrm{re}}(\omega_j)$ is distinct, then

$$(2.3) \qquad \begin{aligned} &\left\{\nu_n\left[\widehat{\lambda}(\omega_j) - \lambda(\omega_j)\right]/\lambda(\omega_j),\right. \\ &\left.\nu_n[\widehat{\boldsymbol{\beta}}(\omega_j) - \boldsymbol{\beta}(\omega_j)]; \quad j = 1, \ldots, J\right\} \end{aligned}$$

converges ($n \to \infty$) jointly in distribution to independent zero-mean normal distributions, the first of which is standard normal; the covariance structure of the asymptotic (normal) distribution of $\widehat{\boldsymbol{\beta}}(\omega_j)$ will be discussed in Section 5. The term ν_n in (2.3) depends on the type of estimator being used. For example, in the case of weighted averaging (we put $m \equiv m_n$ and take $m_n \to \infty$ but $m_n/n \to 0$ as $n \to \infty$), as in (1.5), then $\nu_n^{-2} = \sum_{q=-m}^{m} h_q^2$. If a simple average is used, that is, $h_q = 1/(2m + 1)$, then $\nu_n^2 = (2m + 1)$. Based on these results, asymptotic normal confidence intervals and tests for $\lambda(\omega)$ can be readily constructed. Similarly, for $\boldsymbol{\beta}(\omega)$, asymptotic confidence ellipsoids and chi-square tests can

be constructed; details can be found in Stoffer, Tyler and McDougall (1993, Theorems 3.1–3.3).

Peak searching for the smoothed spectral envelope estimate can be aided using the following approximations. Using a first-order Taylor expansion, we have

$$(2.4) \qquad \log \widehat{\lambda}(\omega) \approx \log \lambda(\omega) + \frac{\widehat{\lambda}(\omega) - \lambda(\omega)}{\lambda(\omega)},$$

so that $\nu_n[\log \widehat{\lambda}(\omega) - \log \lambda(\omega)]$ is approximately standard normal. It also follows that $E[\log \widehat{\lambda}(\omega)] \approx \log \lambda(\omega)$ and $\text{var}[\log \widehat{\lambda}(\omega)] \approx \nu_n^{-2}$. If there is no signal present in a sequence of length n, we expect $\lambda(j/n) \approx 2/n$ for $1 < j < n/2$, and hence approximately $(1 - \alpha) \times 100\%$ of the time, $\log \widehat{\lambda}(\omega)$ will be less than $\log(2/n) + (z_\alpha/\nu_n)$ where z_α is the $(1 - \alpha)$ upper tail cutoff of the standard normal distribution. Exponentiating, the α critical value for $\widehat{\lambda}(\omega)$ becomes $(2/n) \exp(z_\alpha/\nu_n)$. From our experience, thresholding at very small values of α relative to the sample size works well.

As a numerical example, we show how this technology can help detect heterogeneities and wrongly assigned gene segments in DNA sequences. A step-by-step approach for DNA sequences, using the nucleotide alphabet, is as follows:

• Let X_t denote the DNA sequence of interest. Holding the scale for T fixed at zero, form 3×1 vectors $\mathbf{Y}_t$,

$$\mathbf{Y}_t = (1, 0, 0)' \quad \text{if } X_t = \text{A};$$
$$\mathbf{Y}_t = (0, 1, 0)' \quad \text{if } X_t = \text{C};$$
$$\mathbf{Y}_t = (0, 0, 1)' \quad \text{if } X_t = \text{G};$$
$$\mathbf{Y}_t = (0, 0, 0)' \quad \text{if } X_t = \text{T}.$$

The scaling vector is $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$, and the scaled process is $X_t(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_t$.

• Calculate the discrete Fourier transform (DFT) of the data,

$$\mathbf{d}(j/n) = n^{-1/2} \sum_{t=1}^n \mathbf{Y}_t \exp(-2\pi itj/n).$$

Most statistical packages provide a subroutine to calculate the DFT of a time series using the fast Fourier transform. Note that $\mathbf{d}(j/n)$ is a 3×1 complex-valued vector. Calculate the periodogram, $I_n(j/n) = \mathbf{d}(j/n)\mathbf{d}^*(j/n)$, for $j = 1, \ldots, [n/2]$, and retain only the real part, say $I_n^{\text{re}}(j/n)$.

• Smooth the periodogram if desired, using (1.5) or (1.6). For example, using (1.5) with $m = 1$ and

triangular weighting, one would calculate

$$\widehat{f}^{\text{re}}(j/n) = \frac{1}{4}I_n^{\text{re}}\left(\frac{j-1}{n}\right) + \frac{1}{2}I_n^{\text{re}}\left(\frac{j}{n}\right)$$
$$+ \frac{1}{4}I_n^{\text{re}}\left(\frac{j+1}{n}\right).$$

• Calculate the 3×3 sample variance-covariance matrix, $S = n^{-1} \sum_{t=1}^n (\mathbf{Y}_t - \overline{\mathbf{Y}})(\mathbf{Y}_t - \overline{\mathbf{Y}})'$, where $\overline{\mathbf{Y}}$ is the sample mean of the data.

• For each $\omega_j = j/n$, $j = 1, \ldots, [n/2]$, determine the largest eigenvalue and the corresponding eigenvector of the matrix

$$2n^{-1}S^{-1/2}\widehat{f}^{\text{re}}(\omega_j)S^{-1/2}.$$

Note that $S^{1/2}$ is the unique square root matrix of S.

• The sample spectral envelope $\widehat{\lambda}(\omega_j)$ is the eigenvalue obtained in the previous step. If $\mathbf{b}(\omega_j)$ denotes the eigenvector obtained in the previous step, the optimal sample scaling is $\widehat{\boldsymbol{\beta}}(\omega_j) = S^{-1/2}\mathbf{b}(\omega_j)$; this will result in three values, the value corresponding to the fourth category (T in this case) being held fixed at zero.

In this example, we focus on a dynamic (or sliding-window) analysis of the gene labeled BNRF1 (bp 1736-5689) of Epstein–Barr. Figure 5 shows the spectral envelope [using (1.5) with $m = 5$: $h_0 = 6/36$, $h_1 = 5/36, \ldots, h_5 = 1/36$] of the entire coding sequence (3954 bp). The figure shows a strong signal at frequency 1/3; the corresponding optimal scaling was A = 0.04, C = 0.71, G = 0.70, T = 0, which indicates that the signal is in the strong–weak bonding alphabet, $S = \{\text{C}, \text{G}\}$ and $W = \{\text{A}, \text{T}\}$.

Figure 6 shows the result of computing the spectral envelope over four 1000 bp windows (the fourth window is actually 954 bp long) across the CDS, namely, the first, second, third, fourth quarters of BNRF1. An approximate 0.001 significance threshold is 0.69%. The first three quarters contain the signal at the frequency 1/3 (Figure 7a–c); the corresponding sample optimal scalings for the first three windows were (a) A = 0.06, C = 0.69, G = 0.72, T = 0; (b) A = 0.09, C = 0.70, G = 0.71, T = 0; (c) A = 0.18, C = 0.59, G = 0.77, T = 0. The first two windows are strongly consistent with the overall analysis; the third section, however, shows some minor departure from the strong–weak bonding alphabet. The most interesting outcome is that the fourth window shows that no signal is present. This led to the conjecture that the fourth quarter of BNRF1 of Epstein–Barr was actually noncoding.
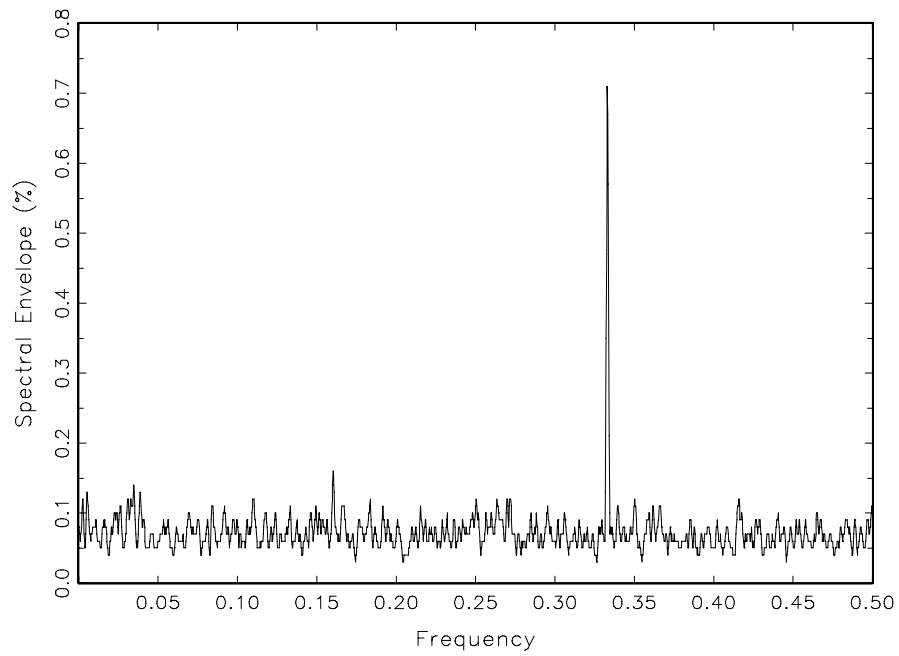
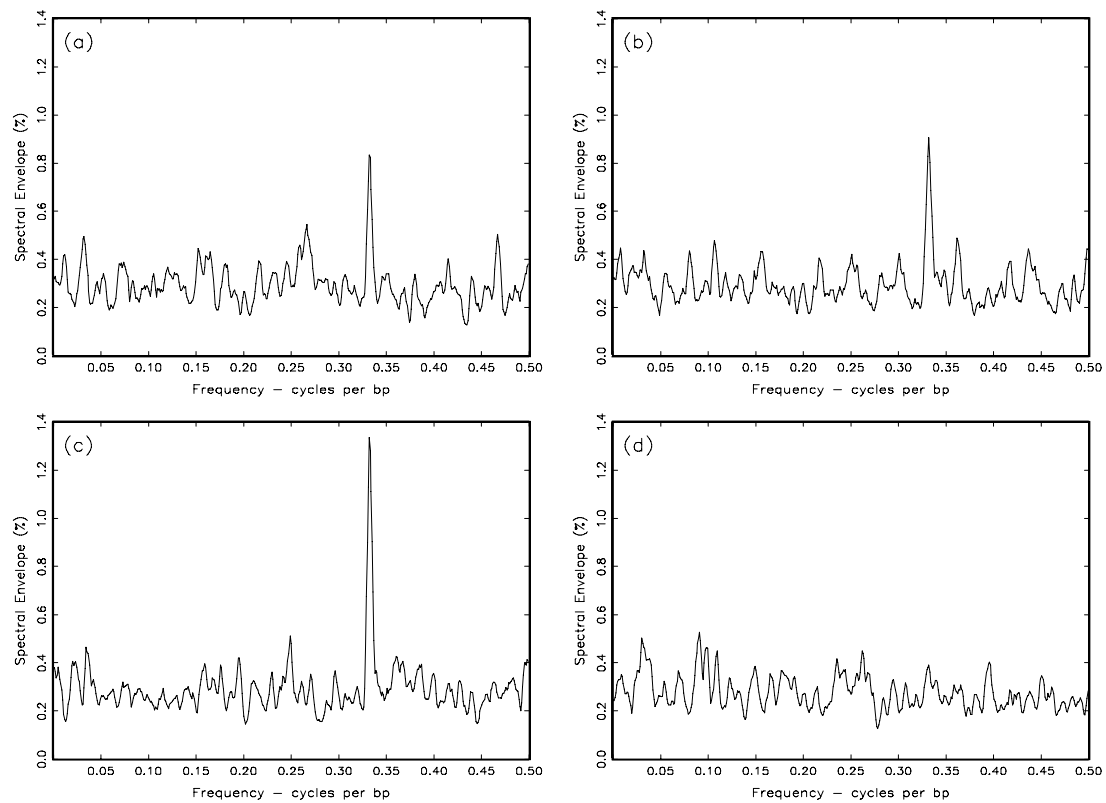FIG. 5.   *Smoothed sample spectral envelope of the* BNRF1 *gene from the Epstein–Barr virus.*

FIG. 6.   *Smoothed sample spectral envelope of the* BNRF1 *gene from the Epstein–Barr virus* (a) *first section* (1000 *bp*), (b) *second section* (1000 bp), (c) *third section* (1000 bp), (d) *fourth section* (954 bp).

3. THE SPECTRAL ENVELOPE FOR REAL-VALUED TIME SERIES

The concept of the spectral envelope presented in the previous section was extended to real-valued time series, $\{X_t; t = 0, \pm 1, \pm 2, \ldots, \}$ in McDougall, Stoffer and Tyler (1997). The process X_t can be vector-valued but here, we will concentrate on the univariate case. Further details can be found in McDougall, Stoffer and Tyler (1997). The technique is similar to projection pursuit (Friedman and Stuetzle, 1981). Let $\mathscr{G}$ denote a k-dimensional vector space of continuous real-valued transformations with $\{g_1, \ldots, g_k\}$ being a set of basis functions satisfying $E[g_i(X_t)] = 0$ and $E[g_i(X_t)^2] < \infty$, $i = 1, \ldots, k$. Analogous to the categorical time series case, define the scaled time series with respect to the set $\mathscr{G}$, to be the real-valued process

$$X_t(\boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{Y}_t = \beta_1 g_1(X_t) + \cdots + \beta_k g_k(X_t)$$

obtained from the vector process

$$\mathbf{Y}_t = (g_1(X_t), \ldots, g_k(X_t))',$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)' \in \mathbf{R}^k$. If the vector process, $\mathbf{Y}_t$, is assumed to have a continuous spectral density, say $f_Y(\omega)$, then $X_t(\boldsymbol{\beta})$ will have a continuous spectral density $f_X(\omega; \boldsymbol{\beta})$ for all $\boldsymbol{\beta} \neq \mathbf{0}$. Noting that $f_X(\omega; \boldsymbol{\beta}) = \boldsymbol{\beta}' f_Y(\omega) \boldsymbol{\beta} = \boldsymbol{\beta}' f_Y^{\mathrm{re}}(\omega) \boldsymbol{\beta}$, and $\sigma^2(\boldsymbol{\beta}) = \mathrm{var}[X_t(\boldsymbol{\beta})] = \boldsymbol{\beta}' V \boldsymbol{\beta}$, where $V = \mathrm{var}(\mathbf{Y}_t)$ is assumed to be positive definite, the optimality criterion,

$$(3.1) \qquad \lambda(\omega) = \sup_{\boldsymbol{\beta} \neq \mathbf{0}} \left\{ \frac{\boldsymbol{\beta}' f_Y^{\mathrm{re}}(\omega) \boldsymbol{\beta}}{\boldsymbol{\beta}' V \boldsymbol{\beta}} \right\},$$

is well defined and represents the largest proportion of the total power that can be attributed to the frequency ω for any particular scaled process $X_t(\boldsymbol{\beta})$. This interpretation of $\lambda(\omega)$ is consistent with the notion of the spectral envelope introduced in the previous section and provides the following working definition: *The spectral envelope of a time series with respect to the space $\mathscr{G}$ is defined to be $\lambda(\omega)$, for $-1/2 \leq \omega \leq 1/2$.*

As in the categorical case, $\lambda(\omega) = \lambda(-\omega)$ so it is suffices to consider $\lambda(\omega)$ for $0 \leq \omega \leq 1/2$. Similarly, the solution to this problem, as in the categorical case, is attained by finding the largest scalar $\lambda(\omega)$ such that

$$(3.2) \qquad f_Y^{\mathrm{re}}(\omega) \boldsymbol{\beta}(\omega) = \lambda(\omega) V \boldsymbol{\beta}(\omega)$$

for $\boldsymbol{\beta}(\omega) \neq \mathbf{0}$. That is, $\lambda(\omega)$ is the largest eigenvalue of $f_Y^{\mathrm{re}}(\omega)$ in the metric of V, and the optimal scaling, $\boldsymbol{\beta}(\omega)$, is the corresponding eigenvector.

If X_t is a categorical time series taking values in the finite state-space $\mathscr{S} = \{c_1, c_2, \ldots, c_k\}$, where

c_j represents a particular category, then an appropriate choice for $\mathscr{G}$ is the set of indicator functions $g_j(X_t) = I(X_t = c_j)$. Hence, this is a natural generalization of the categorical case. Note that in the categorical case, $\mathscr{G}$ does not consist of linearly independent g's, but it was easy to overcome this problem by reducing the dimension by one. In the vector-valued case, $\mathbf{X}_t = (X_{1t}, \ldots, X_{pt})'$, we consider $\mathscr{G}$ to be the class of transformations from $\mathbf{R}^p$ to $\mathbf{R}$ such that the spectral density of $g(\mathbf{X}_t)$ exists. One class of transformations of interest are linear combinations of $\mathbf{X}_t$. In Tiao, Tsay and Wong (1993), for example, linear transformations of this type are used in a time domain approach to investigate contemporaneous relationships among the components of multivariate time series. We will also discuss an interesting application of this in Section 5. Estimation and inference for the real-valued case are analogous to methods described in the previous section for the categorical case.

We focus on two examples here; numerous other examples can be found in McDougall, Stoffer and Tyler (1997). First, a relevant situation may be when X_t is the residual process obtained from some modeling procedure. If the fitted model is appropriate then the residuals should exhibit properties similar to an iid sequence. Departures of the data from the fitted model may suggest model misspecification, non-Gaussian data, or the existence of a nonlinear structure, and the spectral envelope would provide a simple diagnostic tool to aid in a residual analysis. The following is an example.

The series considered here is the quarterly US real GNP from the first quarter of 1947 to the first quarter of 1991. Tiao and Tsay (1994) analyzed this seasonally adjusted time series by focusing on growth rate, that is, $Z_t = \log(\mathrm{GNP}_t / \mathrm{GNP}_{t-1})$, and $n = 176$. Analysis of the ACF (autocorrelation function) and PACF (partial autocorrelation function) of Z_t suggested initially that an MA(2) or an AR(3) model would fit the data. Both models were fit to the data and were found to be similar; that is, the infinite AR representation of the fitted MA(2) model was nearly the fitted AR(3) model, and all parameter estimates were significant. Henceforth, we focus on the results from the MA(2) fit. The residuals from this fit are plotted in Figure 7. As Tiao and Tsay (1994) found, the residuals from the model fit appear to be uncorrelated; there appears to be one or two outliers but their magnitudes are not that extreme. The usual residual analyses showed no obvious structure among the residuals.

Although the MA(2) [or AR(3)] model appears to be appropriate, Tiao and Tsay (1994) investigated the possibility of nonlinearities in GNP growth rate.
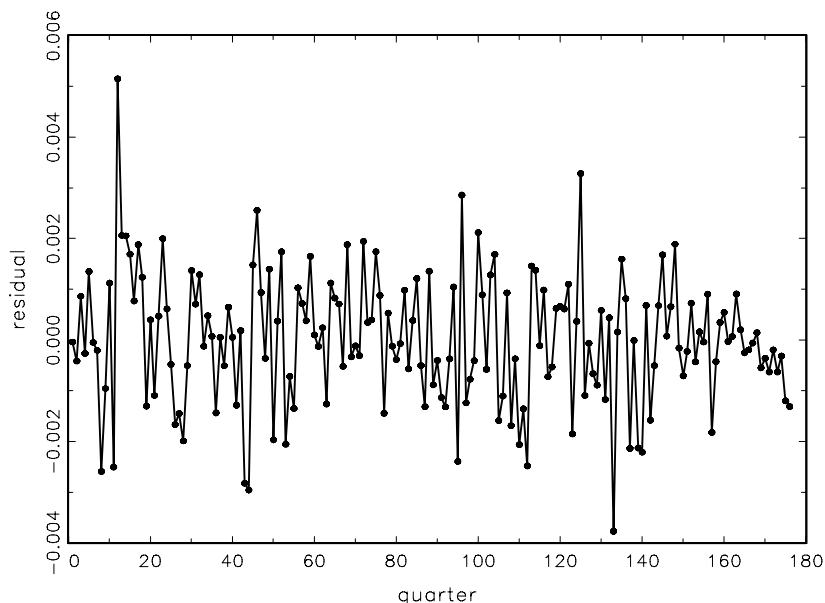
FIG. 7. *Residuals from an* MA(2) *fit to the US GNP growth rate data.*

Their overall conclusion was that there is (subtle) nonlinear behavior in the data, attributable to the fact that the process behaves differently during expansion periods than during recession periods.

The spectral envelope, used as a diagnostic tool on the residuals, clearly indicates that the MA(2) [or AR(3)] model is not adequate and that further analysis is warranted. Here, the generating set $\mathscr{G} = \{x, |x|, x^2\}$ was used to get a [smoothed using $m = 10$ and triangular weighting, $h_0 = 11/121$, $h_1 = 10/121, \ldots, h_{10} = 1/121$ in (1.5)] spectral envelope for the residuals from the MA(2) fit, and the result is plotted in Figure 8. It is clear that the residuals are not iid, and that there is considerable power at the low frequencies. The presence of spectral power at very low frequencies in detrended economic series has been frequently reported and is typically associated with long range dependence. In fact, our choice of $\mathscr{G}$ was partly influenced by the work of Ding, Granger and Engle (1993) who applied transformations of the form $|X_t|^d$, for $d \in (0, 3]$, to the S&P 500 stock market series. The estimated optimal transformation at the first nonzero frequency, $\omega = 0.006$ was $\hat{\boldsymbol{\beta}}(0.006) = (1, 20, -2916)'$, which leads to the transformation

$$(3.3) \qquad y = x + 20|x| - 2916x^2.$$

This transformation is plotted in Figure 9. The transformation, (3.3), is basically the absolute value (with some slight curvature and asymmetry) for most of the residual values, but note that the effect of extreme-valued residuals (outliers) is dampened.

In a second example, we consider a contrived data set where we know the optimal transformation, say g_0, and we determine whether or not the technology can find the transformation when g_0 is not in $\mathscr{G}$. The data, X_t, are generated by the nonlinear model

$$(3.4) \qquad X_t = \exp\{3\sin(2\pi t \omega_0) + \varepsilon_t\},$$
$$t = 1, \ldots, 512,$$

where $\omega_0 = 51/512$ and ε_t is white Gaussian noise with a variance of 16. This example is adapted from Breiman and Friedman (1985) where the ACE algorithm is introduced. The optimal transformation in this case is $g_0(X_t) = \ln(X_t)$ wherein the data are generated from a sinusoid plus noise as in (1.4). Of the 512 generated data, about 98% were less than 4000. Occasionally, the data values were extremely large (the data exceeded 100,000 about four times). The periodogram, in decibels ($10\log_{10} I(\omega_j)$), of the standardized and tapered data is shown in Figure 10 and provides no evidence of any dominant frequency, including ω_0. (The data were tapered by a cosine bell. That is, prior to analysis, the data X_t are replaced by $c_t X_t$, for $t = 1, \ldots, n$, where $c_t = \frac{1}{2}(1 + \cos(2\pi((t - \bar{t})/n)))$, and $\bar{t} = (n + 1)/2$.)

In contrast, the sample spectral envelope (Figure 11) computed with respect to $\mathscr{G} = \{x, \sqrt{x}, \sqrt[3]{x}\}$ has no difficulty in isolating ω_0. Based on Stoffer, Tyler and McDougall (1993, Theorem 3.2), an approximate 0.0001 null significance threshold for the
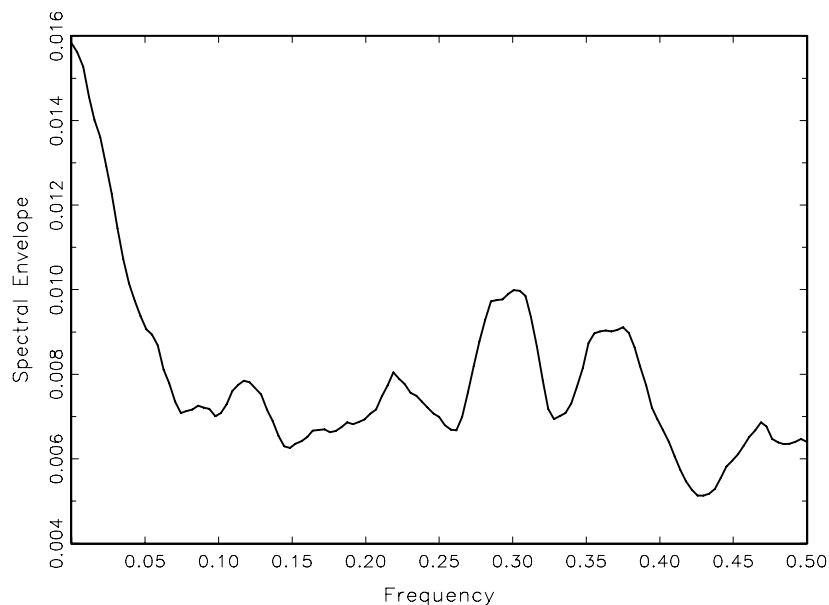
FIG. 8.   *Spectral envelope with respect to $\mathscr{I} = \{x, |x|, x^2\}$ of the residuals from an* MA(2) *fit to the US GNP growth rate data.*

spectral envelope is 4.84% (the null hypothesis being that X_t is iid).

Figure 11 compares the estimated optimal transformation with respect to $\mathscr{I}$ with the log transformation for values less than 4000. The estimated transformation at ω_0 is given by

(3.5) $y = -0.6 + 0.0003x - 0.3638\sqrt{x} + 1.9304\sqrt[3]{x};$

that is, $\widehat{\boldsymbol{\beta}}(\omega_0) = (0.0003, -0.3638, 1.9304)'$ after rescaling so that (3.5) can be compared directly to $y = \ln(x)$.

Finally, it is worth mentioning the result obtained when the rather inappropriate basis, $\{x, x^2, x^3\}$, was used. Surprisingly, the spectral envelope in this case (Figure 13) looks similar to that of Figure 11. Also, the resulting estimated optimal transformation at ω_0 is very close to the log transformation.
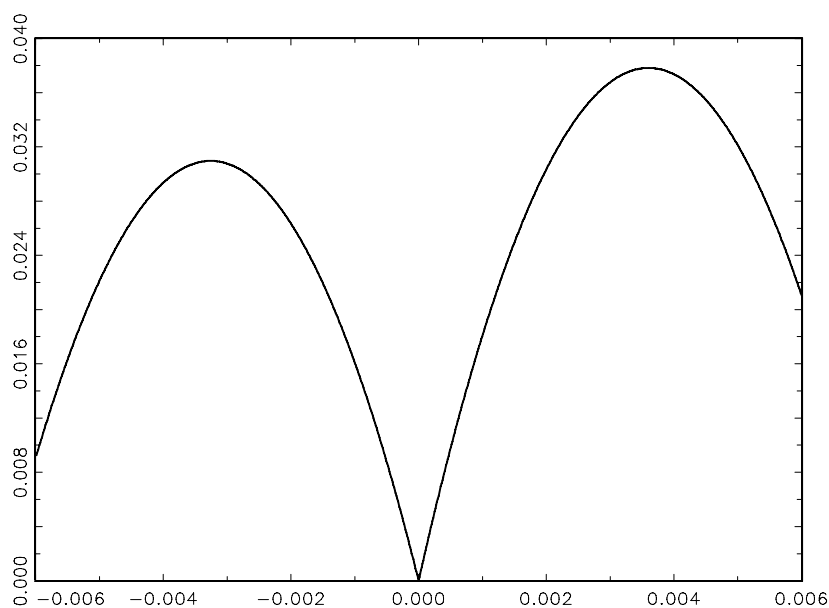

FIG. 9.   *Estimated optimal transformation,* (3.3), *for the GNP residuals at* $\omega = 0.006$.
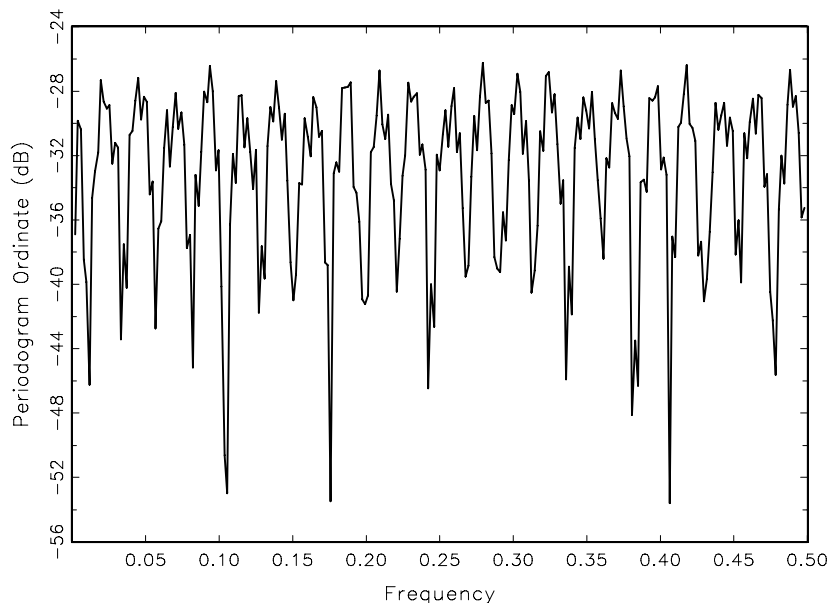
FIG. 10.   *Periodogram, in decibels, of the data generated from* (3.4) *after tapering by a cosine bell.*

In fact, as seen in Figure 12, it looks like what one would imagine as a linear approximation to $y = \ln(x)$ within the range of most of the data.

4. DETECTING COMMON SIGNALS IN MULTIPLE TIME SERIES

Frequently, $p > 1$ time series $\{Y_{jt}, t = 1, \ldots, n_j\}$ for $j = 1, \ldots, p$, are collected with the primary interest being whether any, and how many, have common cyclic components. The series need not be in phase and the sample lengths, n_j, need not be the same, but are of the same magnitude. In this case, a common sample length, n, that is highly composite is chosen and the data are padded or shortened accordingly. There are various methods to solve this type of problem (see MacNeill, 1977, for example), but here we look at an approach based on the spec-
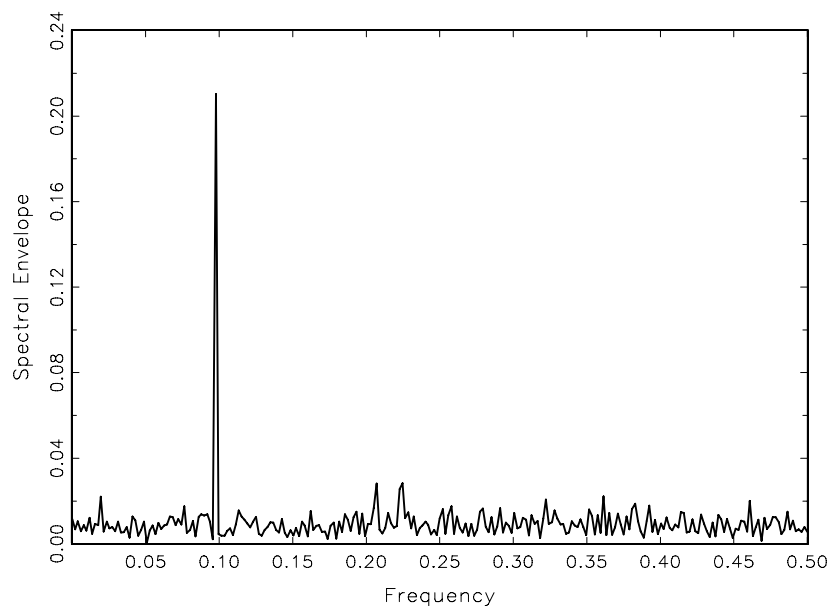

FIG. 11.   *Spectral envelope with respect to* $\mathscr{I} = \{x, \sqrt{x}, \sqrt[3]{x}\}$ *of data generated from* (3.4).
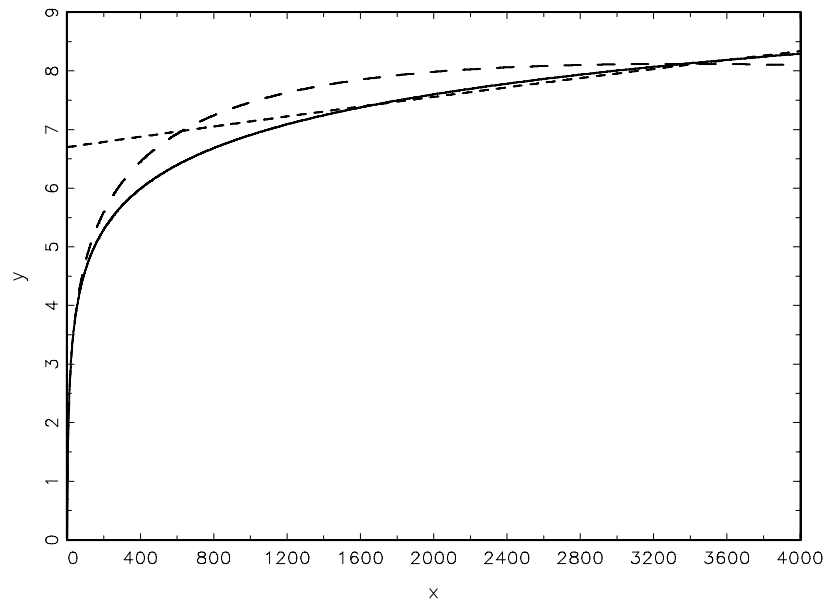
FIG. 12. *Log transformation, $y = \ln(x)$ (solid line), the estimated optimal transformation at ω_0 as given in (3.5) (dashed line), and the estimated optimal transformation at ω_0 using the inappropriate basis $\{x, x^2, x^3\}$ (short-dashed line).*

tral envelope that is developed more thoroughly in Stoffer (1999). Consider the basic model

$$(4.1) \qquad Y_{jt} = \mu_j + c_j S_{t-\tau_j} + \varepsilon_{jt}$$

where μ_j is the mean level of the jth series, S_t is a zero mean, unit variance, random signal with spectrum $f_S(\omega)$, c_j is the amplitude (which may be zero)

corresponding to the jth series and the ε_{jt} are independent white noise processes with variance σ_j^2 and independent of S_t. As described in Section 2, the spectral envelope, $\lambda(\omega)$, is a frequency dependent investigation in that its optimality criterion focuses on individual frequencies. That is, if $\lambda(\omega_0)$ is the spectral envelope at frequency ω_0, then the corresponding optimal scaling, $\boldsymbol{\beta}(\omega_0)$, is optimal at ω_0, but not necessarily anywhere else. Hence, al-
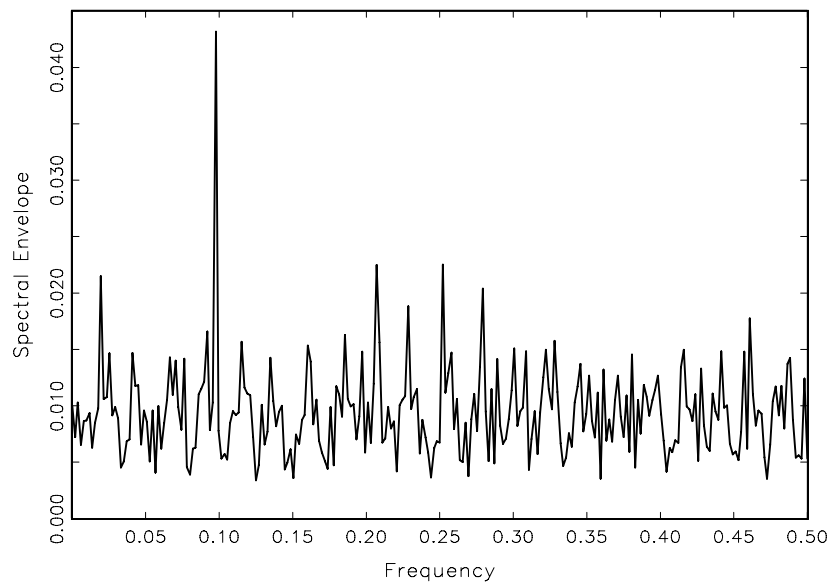

FIG. 13. *Spectral envelope with respect to $\mathscr{S} = \{x, x^2, x^3\}$ of data generated from (3.4).*

though the model (4.1) can be extended to the case of more than one signal as in (6.22), we can concentrate on the case of one signal with little loss of generality.

Denote the individual DFTs by $d_{Y_j}(\omega)$ for $j = 1, \ldots, p$; they are evaluated over ω in the set of positive fundamental frequencies $\{\omega = \ell/n;\ \ell = 1, 2, \ldots, [(n-1)/2]\}$. In terms of the model (4.1), these can be written as

$$(4.2) \qquad d_{Y_j}(\omega) = a_j\, d_S(\omega) + d_{\varepsilon_j}(\omega)$$

where $d_S(\omega)$ and $d_{\varepsilon_j}(\omega)$ represent the transforms of the signal and of the noise, respectively, and $a_j = c_j \exp(-2\pi i \tau_j \omega)$.

Let $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{pt})'$ be the vector of observations, and let $\mathbf{d}_Y(\omega)$ denote the $p \times 1$ vector of transforms with jth component $d_{Y_j}(\omega)$. If component j does not contain the signal, then $a_j = 0$, and we arrange the vectors so that for $0 \leq q \leq p$, $a_{q+1} = \cdots = a_p = 0$, and none of the $\{a_1, \ldots, a_q\}$ are zero. Then, the model (4.2) can be written as

$$(4.3) \qquad \mathbf{d}_Y(\omega) = \mathbf{a} d_S(\omega) + \mathbf{d}_\varepsilon(\omega),$$

where $\mathbf{a}$ is the $p \times 1$ complex vector $\mathbf{a} = (a_1, \ldots, a_q, 0, \ldots, 0)'$, and $\mathbf{d}_\varepsilon(\omega)$ is the $p \times 1$ vector of transforms with jth component $d_{\varepsilon_j}(\omega)$.

Let $f_Y(\omega)$ be the $p \times p$ spectral matrix of $\mathbf{Y}_t$. Then, we have the following decomposition:

$$(4.4) \qquad f_Y(\omega) = \mathbf{a} f_S(\omega)\mathbf{a}^* + f_\varepsilon(\omega),$$

where $\mathbf{a} f_S(\omega)\mathbf{a}^*$ is a Hermitian block diagonal matrix, and $f_\varepsilon(\omega) = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_p^2\}$.

Rather than work with the original series, we prefer to work with the standardized series, $\mathbf{Z}_t = V^{-1/2}\mathbf{Y}_t$, where $V = \mathrm{diag}\{\sigma_{Y_1}^2, \ldots, \sigma_{Y_p}^2\}$; note that the diagonal elements of V get contributions from both the signal (if present) and the noise, that is, $\sigma_{Y_j}^2 = c_j^2 + \sigma_j^2$ where c_j could be zero. Writing (4.4) in terms of the standardized processes, $\mathbf{Z}_t$, we have

$$(4.5) \qquad f_Z(\omega) = \mathbf{a}(\omega)\mathbf{a}^*(\omega) + D(\omega);$$

note that (4.5) is a type of complex factor analytic representation for $f_Z(\omega) = V^{-1/2}f_Y(\omega)V^{-1/2}$. That is, $f_Z(\omega)$ is the sum of a rank 1, nonnegative definite Hermitian matrix, $\mathbf{a}(\omega)\mathbf{a}^*(\omega)$, and a diagonal positive definite matrix, $D(\omega)$, where $\mathbf{a}(\omega)$ is a $p \times 1$ complex-valued vector such that $\mathbf{a}(\omega)\mathbf{a}^*(\omega) = V^{-1/2}\mathbf{a} f_S(\omega)\mathbf{a}^* V^{-1/2}$ and the final $p - q$ elements of $\mathbf{a}(\omega)$ are zero. In addition, we are restricting attention to case where $D(\omega) = V^{-1/2}f_\varepsilon(\omega)V^{-1/2}$ is constant over ω, that is, $D(\omega) \equiv \mathrm{diag}\{\sigma_1^2/\sigma_{Y_1}^2, \ldots, \sigma_p^2/\sigma_{Y_p}^2\}$. The value q corresponds to the number of elements that contain the signal. Since $f_S(\omega)$ is real, if the series are in phase, that is, $\tau_1 = \cdots = \tau_q = 0$, then $f_Z(\omega)$ will also be real.

Let $\boldsymbol{\beta} \neq \mathbf{0}$ be a *complex-valued* $p \times 1$ vector of scalings and consider the standardized, scaled complex series, $X_t(\boldsymbol{\beta}) = \boldsymbol{\beta}^*\mathbf{Z}_t$. We may now write

$$(4.6) \qquad f_X(\omega, \boldsymbol{\beta}) = \boldsymbol{\beta}^* f_Z(\omega)\boldsymbol{\beta},$$

where $f_X(\omega, \boldsymbol{\beta})$ is the spectral density of $X_t(\boldsymbol{\beta})$. If we believe that the series are in phase ($\tau_1 = \cdots = \tau_q = 0$), then there is no particular advantage to considering complex-valued scales and we would restrict $\boldsymbol{\beta}$ to be a vector of real scalings. For generality we focus on the complex case, the real case follows as in Sections 2 and 3.

Our goal is to find $\boldsymbol{\beta}$, at each ω, to maximize (4.6), subject to the constraint $\boldsymbol{\beta}^*\boldsymbol{\beta} = 1$. We will denote the solution by $\boldsymbol{\beta}(\omega)$. Setting $\mathbf{b} = V^{-1/2}\boldsymbol{\beta}$ leads to the optimality criterion

$$(4.7) \qquad \lambda(\omega) = \sup_{\mathbf{b} \neq \mathbf{0}} \left\{ \frac{\mathbf{b}^* f_Y(\omega)\mathbf{b}}{\mathbf{b}^* V \mathbf{b}} \right\}.$$

The function $\lambda(\omega)$ is the spectral envelope because $f_X(\omega, \boldsymbol{\beta}) \leq \lambda(\omega)$, for any scaling $\boldsymbol{\beta}$, with equality when $\boldsymbol{\beta} = \boldsymbol{\beta}(\omega)$. The corresponding scaling $\boldsymbol{\beta}(\omega)$ is the optimal scaling. The idea in employing (4.7) is that the right (complex) linear combination of the p series, namely $X_t(\boldsymbol{\beta}(\omega))$, will enhance the signal and dampen the noise. Note that the role of $\boldsymbol{\beta}$ in this section is slightly different than in the previous section because we are working with the standardized series. If $V = \mathrm{diag}\{\sigma_{Y_1}^2, \ldots, \sigma_{Y_p}^2\}$ in (4.7) was replaced by $V_Y = \mathrm{var}\{\mathbf{Y}(t)\}$, then $\mathbf{b}$ in (4.7) is $\boldsymbol{\beta}$ of (2.2). The reason that V is preferred to V_Y in this problem is that using V leads to the simple decomposition of (4.5). In addition, simulation studies showed that using V was superior to V_Y in identifying signals.

The solution to (4.7) is obtained by finding, at each Fourier frequency ω, the largest eigenvalue, $\lambda(\omega)$, and corresponding eigenvector, $\boldsymbol{\beta}(\omega)$, of the matrix $f_Z(\omega) = V^{-1/2}f_Y(\omega)V^{-1/2}$, which, as we previously stated, is the sum of the block diagonal matrix (with nonzero values only in the upper $q \times q$ portion) $\mathbf{a}(\omega)\mathbf{a}^*(\omega) = V^{-1/2}\mathbf{a} f_S(\omega)\mathbf{a}^* V^{-1/2}$ and the diagonal real matrix $D(\omega) = V^{-1/2}f_\varepsilon(\omega)V^{-1/2}$ that is constant with respect to ω. From this it is seen that if there is a harmonic component near ω in q of the p series, then $\lambda(\omega)$ will be "large" and the final $p - q$ components of $\boldsymbol{\beta}(\omega)$ will be zero. In regions where there is no harmonic component, $f_S(\omega)$ will be negligible so that $\lambda(\omega)$ will be flat and the final $p - q$ components of $\boldsymbol{\beta}(\omega)$ will not be zero; in fact, the modulus of those $p - q$ elements may dominate the first q elements if the signal amplitudes are large. These facts will lead to reasonable conclusions about the nature of the signal and the number of elements q that contain the signal. We have also

found that it is worthwhile to look at the unstandardized results, that is, replace V with the identity matrix in (4.7); this aids in the identification of interesting frequencies. Estimation of the spectral envelope and optimal scalings is accomplished using the techniques of Section 2.

This idea can be extended to the case where some series contain signals that may be only approximately the same. Details are provided in Stoffer (1999), but briefly, to accomplish this, frequency is modeled as a random effect. We motivate the ideas with the following example.

Suppose that the time series of interest $Y_j(t)$ follow the model

$$Y_j(t) = \mu_j + c_j S_j(t) + \varepsilon_j(t),$$

(4.8)
$$j = 1, \ldots, p,$$

where μ_j is the level of the jth series, and $\varepsilon_j(t)$ are independent white noise processes with variance σ_j^2 that are independent of the signals. The amplitude of the jth signal is c_j (which may be zero), where the random signals are given by

(4.9) $\qquad S_j(t) = \cos(2\pi[\omega_0 + \delta_j]t + 2\pi\tau_j).$

In (4.9), τ_j is a fixed phase, ω_0 is the common frequency, and δ_j for $j = 1, \ldots, p$ are independent random uniform perturbations, with $\delta_j \sim U(-\Delta, \Delta)$, and $\Delta \geq 0$ is some small amount.

In this model, the frequency corresponding to each time series is a random effect so that the series are allowed to be only approximately the same. In addition, it is not necessary for ω_0 to be a Fourier frequency. The expected behavior of the signal is

$$\mathrm{E}\{S_j(t)\}$$

(4.10)
$$= \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} \cos\{2\pi(\omega_0 + \delta)t + 2\pi\tau_j\}\, d\delta$$

$$= \cos(2\pi[\omega_0 t + \tau_j])\, \mathrm{sinc}(2\pi\Delta t)$$

using the definition $\mathrm{sinc}(u) \equiv \sin(u)/u$, $u \neq 0$ and $\mathrm{sinc}(0) \equiv 1$. As general statement, (4.10) implies that the particular signals given in (4.9), satisfy the general property that

(4.11) $\qquad \mathrm{E}\{S_j(t)\} = S_0(t - \tau_j)\kappa(\Delta t)$

where $S_0(t - \tau_j)$ is a fixed harmonic signal (with arbitrary phase) that oscillates at frequency ω_0, and $\kappa(\Delta t)$ is a taper that is controlled by Δ; the larger the value of Δ the more dramatic the taper, and when $\Delta = 0$ the taper has no effect.

If we now consider the DFTs of the data, then we can write the model (4.8) as

$$d_{Y_j}(\omega) = a_j d_{S_j}(\omega) + d_{\varepsilon_j}(\omega),$$

(4.12)
$$j = 1, \ldots, p,$$

where $a_j = c_j \exp(-2\pi i \tau_j \omega)$, $d_{S_j}(\omega)$ is the DFT of the individual random signal term $\cos(2\pi[\omega_0 +$

$\delta_j]t)$, and $d_{\varepsilon_j}(\omega)$ is the DFT of the individual noise terms. Note that (4.12) is similar, but not equivalent, to (4.2).

To investigate (4.12), it is easier to work with the complex version of the signal; that is, let

$$S_j(t) = \exp\{2\pi i(\omega_0 + \delta_j)t\}.$$

Then, as in (4.10),

(4.13)
$$\mathrm{E}\{d_{S_j}(\omega)\} = n^{-1/2} \sum_{t=1}^{n} \exp\{-2\pi i(\omega - \omega_0)t\}$$

$$\times \mathrm{sinc}(2\pi\Delta t),$$

which is the tapered (by the sinc kernel) transform of $\exp(2\pi i \omega_0 t)$. The modulus of $\mathrm{E}\{d_{S_j}(\omega)\}$ will be extremely close to the modulus of the finite transform of $\exp(i\omega_0 t)$ provided that Δ is not too large relative to n.

The individual series, $Y_j(t)$, are coherent with each other when their respective signal amplitudes are not zero. If in (4.12), $a_j a_k = 0$ for any pair $j, k = 1, \ldots, p$, then

(4.14) $\quad \mathrm{E}\{d_{Y_j}(\omega)d_{Y_k}^*(\omega)\} = \sigma_j^2\, I_{\{0\}}(j - k),$

where σ_j^2 is the variance of $\varepsilon_j(t)$ and $I_A(\cdot)$ is the indicator of the set A. If $|a_j a_k| > 0$ when $j \neq k$, then

(4.15)
$$\mathrm{E}\{d_{Y_j}(\omega)d_{Y_k}^*(\omega)\}$$

$$= a_j a_k \mathrm{E}\{d_{S_j}(\omega)\}\mathrm{E}\{d_{S_k}^*(\omega)\}$$

where $\mathrm{E}\{d_{S_j}(\omega)\}$ was given in (4.13), and when $j = k$,

(4.16) $\quad \mathrm{E}\{|d_{Y_j}(\omega)|^2\} = |a_j|^2\, \mathrm{E}\{|d_{S_j}(\omega)|^2\} + \sigma_j^2,$

where

$$\mathrm{E}\{|d_{S_j}(\omega)|^2\} = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \exp\{-2\pi i(\omega - \omega_0)h\}$$

$$\times \mathrm{sinc}(2\pi\Delta h).$$

A key point made in this example is that, even in the situation of (4.8) and (4.9), we may carry out an analysis using the spectral envelope. That is, rather than focus on the p individual transforms, we can use the interrelationships inherent in the model to help identify ω_0. To do this, we make use of the cross-spectra among the p time series $Y_j(t)$ for $j = 1, \ldots, p$. This enhances the analysis because if a particular series contains the harmonic component of interest but with a low amplitude or a lot of noise, an individual analysis may miss this fact, but this particular series may match well with other series containing the harmonic. This particular effect will be seen in the example in Section 6.

As a general model we take the form of (4.8); that is,

$$Y_j(t) = \mu_j + c_j S_j(t) + \varepsilon_j(t),$$

(4.17)
$$j = 1, \ldots, p,$$

but where the signals $S_j(t)$ belong to the general class signals given by

(4.18) $E\{S_j(t)\} = S_0(t - \tau_j)\kappa(\Delta t),$

where $S_0(t)$ is a fixed harmonic signal that oscillates at frequency ω_0, and $\kappa(\Delta t)$ is a taper that does not annihilate $S_0(t)$. Of course we may extend (4.17) to the case of more than one signal in an obvious way.

As in the previous example, (4.12), the model can be written in terms of the DFTs

(4.19) $d_{Y_j}(\omega) = a_j d_{S_j}(\omega) + d_{\varepsilon_j}(\omega),$
$$j = 1, \ldots, p,$$

where $a_j = c_j \exp(-2\pi i \tau_j \omega)$, $d_{S_j}(\omega)$ is the DFT of the individual random signal term, and $d_{\varepsilon_j}(\omega)$ is the DFT of the individual noise terms. In particular, $d_{S_j}(\omega)$ is the DFT of $S_0(t)\kappa(\Delta t)$.

Let $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_p(t))'$ and let $\mathbf{d}_Y(\omega)$ denote the $p \times 1$ vector of finite transforms with j-th component $d_{Y_j}(\omega)$. As in Section 1, arrange the vector so that for $0 \leq q \leq p$, $a_{q+1} = \cdots = a_p = 0$, and none of the $\{a_1, \ldots, a_q\}$ are zero. Then the model (4.19) can be written as

(4.20) $\mathbf{d}_Y(\omega) = A\mathbf{d}_S(\omega) + \mathbf{d}_\varepsilon(\omega),$

where A is the $p \times p$ complex diagonal matrix $A = \text{diag}\{a_1, \ldots, a_q, 0, \ldots, 0\}$, and $\mathbf{d}_S(\omega)$ and $\mathbf{d}_\varepsilon(\omega)$ are the $p \times 1$ vectors of transforms with jth component $d_{S_j}(\omega)$ and $d_{\varepsilon_j}(\omega)$, respectively.

Let $f_Y(\omega) = E\{\mathbf{d}_Y(\omega)\mathbf{d}_Y^*(\omega)\}$ be the $p \times p$ complex matrix with elements given by the values in (4.14)–(4.16), with $\text{sinc}(2\pi\Delta t)$ replaced by the general form $\kappa(\Delta t)$. In an obvious way, based on (4.20), decompose the elements of $f_Y(\omega)$ as

(4.21) $f_Y(\omega) = A f_S(\omega)A^* + f_\varepsilon(\omega),$

where $A f_S(\omega)A^*$ is a Hermitian block diagonal matrix and $f_\varepsilon(\omega) = \text{diag}\{\sigma_1^2, \ldots, \sigma_p^2\}$. This situation is somewhat different than (4.4); however, (4.21) is a type of complex factor analytic representation for $f_Y(\omega)$; that is, $f_Y(\omega)$ is the sum of a rank $r \leq q \leq p$ nonnegative definite matrix $A f_S(\omega)A^*$ and a positive definite diagonal matrix $f_\varepsilon(\omega)$. The elements of the matrix $f_S(\omega)$ depend on n, κ, Δ and ω_0. In particular, if $S_0(t) = \cos(2\pi\omega_0 t)$, then from (4.13), (4.15) and (4.16), the off-diagonal elements of $f_S(\omega)$ are of the form (for ease we use the complex version of the signal)

(4.22) $f_{S,jk}(\omega)$
$$= n^{-1}\left|\sum_{t=1}^{n} \exp\{-2\pi i(\omega - \omega_0)t\}\kappa(\Delta t)\right|^2,$$
$$j \neq k = 1, \ldots, p,$$

while the diagonal elements are

(4.23) $f_{S,jj}(\omega) = \sum_{|h|<n}\left(1 - \frac{|h|}{n}\right)$
$$\times \exp\{-2\pi i(\omega - \omega_0)h\}\kappa(\Delta h),$$
$$j = 1, \ldots, p.$$

For ω far from ω_0, $f_S(\omega)$ is approximately the zero matrix (recall that we do not assume that ω_0 is a Fourier frequency) because its elements are convoluted Dirichlet kernels evaluated away from zero. The situation in (4.1)–(4.5) can be thought of as the case where $\Delta = 0$, in which case all of the elements of $f_S(\omega)$ are the same implying that $f_S(\omega)$ is rank one.

Again, our preference is to work with the standardized series $\mathbf{Z}(t) = V^{-1/2}\mathbf{Y}(t)$. Thus with $f_Z(\omega) = V^{-1/2}f_Y(\omega)V^{-1/2}$, we put

(4.24) $B(\omega)B^*(\omega) = AV^{-1/2}f_Y(\omega)V^{-1/2}A^*$

and

(4.25) $D(\omega) = V^{-1/2}f_\varepsilon(\omega)V^{-1/2}$
$$= \text{diag}\{\sigma_1^2/\sigma_{Y_1}^2, \ldots, \sigma_p^2/\sigma_{Y_p}^2\}$$

and write the model as

(4.26) $f_Z(\omega) = B(\omega)B^*(\omega) + D(\omega),$

noting that $B(\omega)$ is a $p \times q$ matrix such that all of the elements in the final $p - q$ rows of $B(\omega)$ are zero.

Estimation of the spectral envelope and optimal scalings is accomplished in the usual way so that cross-spectra can be estimated. We note that smoothing techniques to estimate the spectral envelope is reasonable even near ω_0 because one expects the spectra to be spread out around this frequency as seen in (4.22) and (4.23).

For an example, we discuss part of an analysis that was performed in Stoffer (1999). The analysis focuses on a study that used functional magnetic resonance imaging (fMRI) to examine pain perception in humans (Antognini, Buonocore, Disbrow and Carstens, 1998). Two types of stimuli were presented to *awake* subjects: electric *shock* (15 mA, 2 Hz) and nonpainful *brush*. The effects of general anesthesia on pain perception were evaluated by comparing results from alert volunteers with those from the same volunteers while anesthetized and paralyzed with 0.15 mg/kg Vecuronium. Data were collected under two anesthetic conditions, 0.7% (*low*), and 1.3% (*high*) Isoflurane. During the anesthetized conditions a *supramaximal shock* (50 mA, 100 Hz) stimulus was added. This stimulus was used to simulate surgical incision without inflicting tissue damage. The stimulus conditions were: [1] Awake-brush (5 subjects), [2] Awake-shock

(5 subjects), [3] Low-brush (3 subjects), [4] Low-supramaximal shock (4 subjects), [5] High-brush (4 subjects), and [6] High-supramaximal shock (5 subjects). The specific locations of the brain where the signal was measured were [L1] Cortex: Primary somatosensory, contralateral; [L2] Cortex: Primary somatosensory, ipsilateral; [L3] Cortex: Secondary somatosensory, contralateral; [L4] Cortex: secondary somatosensory, ipsilateral; [L5] Caudate; [L6] Thalamus: contralateral; [L7] Thalamus: ipsilateral; [L8] Cerebellum: contralateral; [L9] Cerebellum: ipsilateral.

The data consist of consecutive measures of blood oxygenation level dependent (BOLD) signal intensity; for details, see Ogawa, Lee, Nayak and Glynn (1990) and Ogawa and Lee (1990). Areas of activation were analyzed over time at the level of the voxel, or three dimensional pixel; the voxel with the highest activation was chosen from each brain location. The sampling rate was one observation every two seconds for 256 seconds ($n = 128$). Each stimulus was applied for 32 seconds and then stopped for 32 seconds; thus, the signal period is 64 seconds. For this analysis, we averaged over subjects in each stimulus condition (these were evoked responses and all subjects were in phase) producing data, Y_{jt}, for each stimulus condition [1]–[6]. Thus, for a given stimulus condition, Y_{jt} represents the average BOLD signal intensity at location j, for $j = 1, \ldots, 9$, at time point t, for $t = 1, \ldots, 128$.

The types of data encountered in this experiment are quite varied. For example, Figure 14 shows the nine time series, one for each location, of the stimulus condition Awake-brush. It is clear from the data, in this case, that most locations received the brush signal of approximately four cycles in 128 points. Figure 15, however, for the stimulus condition High-brush, is quite different than Figure 14, and it is not clear whether or not any location received any signal.

If we focus on real scales, then the results of Section 2 [in particular (2.3) with details found in Theorem 3.3 of Stoffer, Tyler and McDougall, 1993] apply. If the focus is on complex scales, then the first part of (2.3) remains valid, but $\widehat{\boldsymbol{\beta}}(\omega)$ will have a complex multivariate normal distribution. In the analyses that we present here, we focus on the complex scaling case; the case of real-valued scalings gives approximately the same results and leads to the same conclusions about the nature of pain perception. In this analysis we have $n = 128$, $m = 2$ with weights $\{h_0 = 3/9, \ h_{\pm 1} = 2/9, \ h_{\pm 2} = 1/9\}$, so that $\nu_n = \sqrt{81/19} = 2.065$, and using (2.4), the approximate 0.001 critical threshold for this example is $\exp\{3.09/2.065\}/64 = 7\%$. Inference about the opti-

TABLE 3
Scalings for the brush stimulus under conditions awake and high anesthesia

Location	Awake-brush $\left|\widehat{\boldsymbol{\beta}}\left(\frac{4}{128}\right)\right|$	High-brush $\left|\widehat{\boldsymbol{\beta}}\left(\frac{4}{128}\right)\right|$
L1	0.46	0.50
L2	0.40	0.28
L3	0.45	0.48
L4	0.40	0.43
L5	0.08*	0.34
L6	0.28	0.24
L7	0.15	0.16
L8	0.09*	0.14*
L9	0.39	0.18

*The value of zero is in an approximate 99% confidence region for this component.

mal (complex) scaling vector can be performed using Brillinger (1975, Theorems 9.4.3 and 9.4.4) wherein the asymptotic ($n, m \to \infty$) covariance matrix of the sample optimal scaling, $\widehat{\boldsymbol{\beta}}(\omega)$, say $\boldsymbol{\Sigma}_\beta(\omega)$, is given by

$$\boldsymbol{\Sigma}_\beta(\omega) = \nu_n^{-2}\lambda_1(\omega) \sum_{\ell=2}^{p} \lambda_\ell(\omega)\{\lambda_1(\omega) - \lambda_\ell(\omega)\}^{-2}$$

$$\times \boldsymbol{\beta}_\ell(\omega)\boldsymbol{\beta}_\ell^*(\omega),$$

where $\{\lambda_1(\omega) = \lambda(\omega), \lambda_2(\omega), \ldots, \lambda_p(\omega)\}$ are the latent roots of $V^{-1}f_Y(\omega)$ arranged in decreasing order, and $\{\boldsymbol{\beta}_1(\omega) = \boldsymbol{\beta}(\omega), \boldsymbol{\beta}_2(\omega), \ldots, \boldsymbol{\beta}_p(\omega)\}$ are the corresponding latent vectors.

We may use this result to form confidence regions for the components, $\widehat{\boldsymbol{\beta}}_{1,j}(\omega)$, $j = 1, \ldots, p$, of the optimal scaling vector by approximating the distribution of

$$(4.27) \qquad \frac{2\left|\widehat{\boldsymbol{\beta}}_{1,j}(\omega) - \boldsymbol{\beta}_{1,j}(\omega)\right|^2}{s_j^2(\omega)}$$

by a χ^2 distribution with 2 degrees of freedom. In (4.27), $s_j^2(\omega)$ is the jth diagonal element of $\widehat{\boldsymbol{\Sigma}}_\beta(\omega)$, the estimate of $\boldsymbol{\Sigma}_\beta(\omega)$. We can use (4.27) to check whether or not the value of zero is in the confidence region by comparing $2|\widehat{\boldsymbol{\beta}}_{1,j}(\omega)|^2/s_j^2(\omega)$ with $\chi_2^2(1 - \alpha)$, the $1 - \alpha$ upper tail cutoff of the χ_2^2 distribution.

First we describe the analysis of the stimulus condition Awake-brush. We consider 7% to be an approximate significance threshold for the sample spectral envelope. Figure 16 shows the sample spectral envelope for the Awake-brush condition. There is a clear peak in the sample spectral envelope, $\widehat{\lambda}(4/128) = 34\%$, corresponding to frequency of $\omega = 4$ cycles/128 points, or 4/256 Hz, which corresponds to the stimulation period of 64 seconds. The magnitudes of the optimal sample scalings for Awake-brush at this frequency are listed in Table 3. In addition, Table 3 indicates whether an approximate 99%
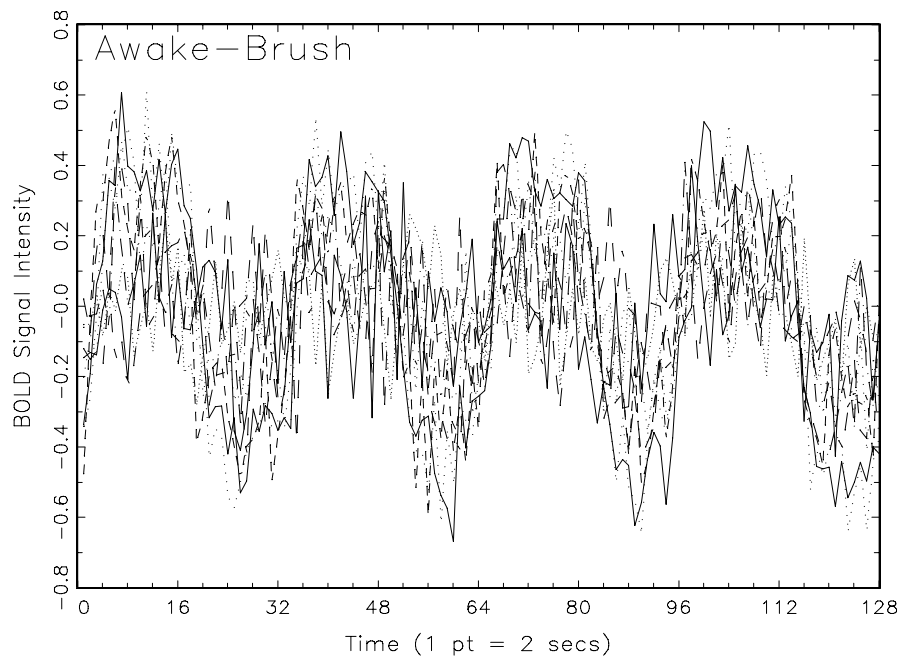
Fig. 14.   *The* **BOLD** *signal intensities at all nine locations for the stimulus condition Awake-brush.*
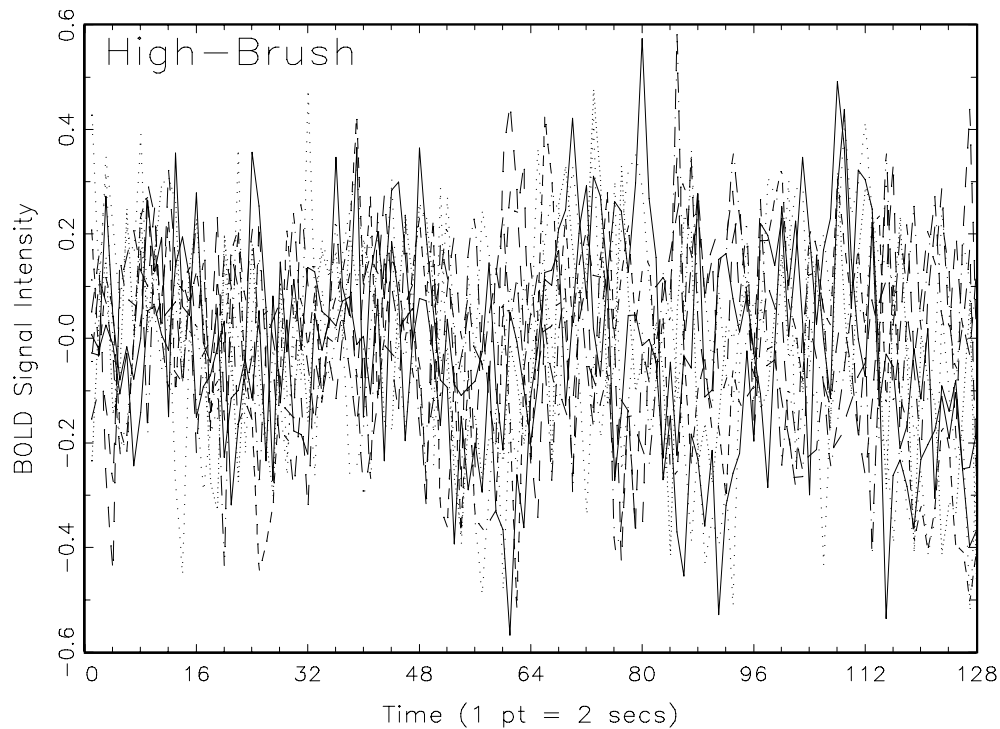

Fig. 15.   *The* **BOLD** *signal intensities at all nine locations for the stimulus condition High-brush.*

confidence region of an element of the sample opti-mal scaling includes zero; these regions are based on (4.27). Table 3 indicates that locations L5 (caudate) and L8 (cerebellum: contralateral) are probably not receiving the signal.
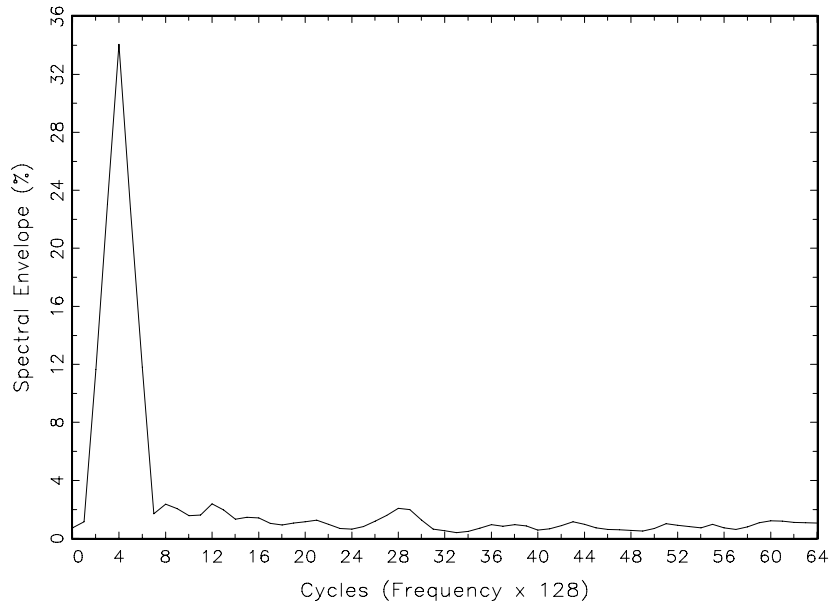
FIG. 16. *The sample spectral envelope,* $\widehat{\lambda}(\omega)$, *for the Awake-brush stimulus condition. The peak in the spectral envelope occurs at the frequency* $\omega = 4/128$.
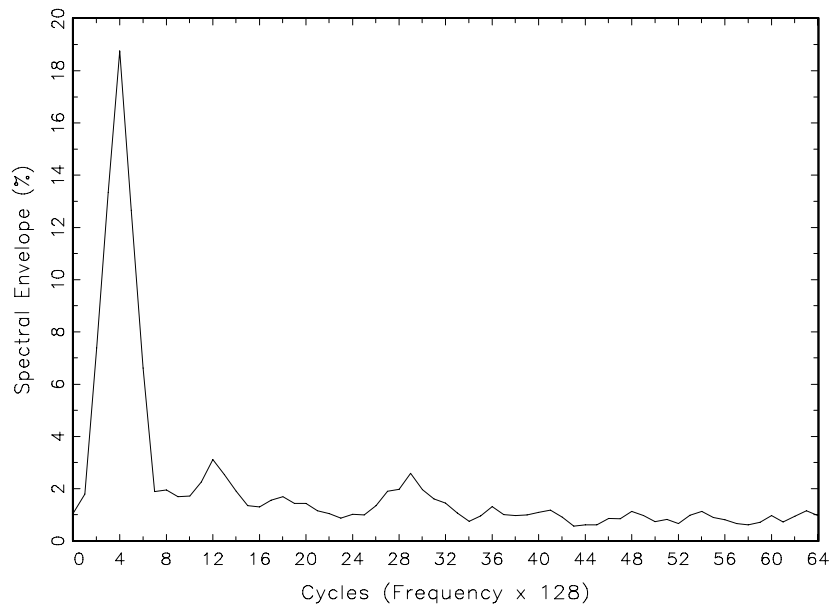

FIG. 17. *The sample spectral envelope,* $\widehat{\lambda}(\omega)$, *for the High-brush stimulus condition. The peak in the spectral envelope occurs at the frequency* $\omega = 4/128$.

Figure 17 refers to the stimulus condition High-brush. In this case, the spectral envelope shows a peak, $\widehat{\lambda}(4/128) = 18.8\%$ at $\omega = 4/128$, and Table 3 shows that at this frequency, only location L8 (cerebellum: contralateral) is probably not receiving the brush signal. Note that for the anesthetic condition, the sample spectral envelope (and hence the signal power) is about half that of the awake conditions; this is consistent across all stimulus conditions. In summary, the analyses provided evidence that nociceptive inputs play a role in activating areas of the cortex that are traditionally considered to be in-

volved primarily in tactile discrimination. Furthermore, the results contributed to a growing body of evidence suggesting that the basal ganglia are involved in the perception of pain.

5. QUALITATIVE RANDOM FIELDS

Another extension of optimal scaling and the spectral envelope that is currently under exploration is the analysis of qualitative spatial data. Here, we let $\{X_{\mathbf{t}}; \mathbf{t} = (t_1, t_2); t_1, t_2 = 0, \pm 1, \pm 2, \ldots\}$ be a homogeneous, categorical-valued, two-dimensional random field taking values in the finite set $\mathscr{C} = \{c_1, c_2, \ldots, c_{k+1}\}$. For example, $\mathscr{C}$ could be an eight bit grayscale or a 256 color palette ($k + 1 = 256$). Analogous to Section 2, for $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)' \in \mathbf{R}^k$, let $X_{\mathbf{t}}(\boldsymbol{\beta})$ denote the real-valued stationary two-dimensional random field corresponding to the scaling that assigns the category c_j the numerical value β_j, $j = 1, 2, \ldots, k$ and c_{k+1} the value 0. The goal is to find scalings $\boldsymbol{\beta}$ so that the spectral density (or wave number spectrum) of $X_{\mathbf{t}}(\boldsymbol{\beta})$, say $f_X(\boldsymbol{\omega}, \boldsymbol{\beta})$, with $\boldsymbol{\omega} = (\omega_1, \omega_2) \in [-1/2, 1/2]^2$, is in some sense interesting. As in the one-dimensional case, we choose $\boldsymbol{\beta}(\boldsymbol{\omega})$ at each fundamental frequency (or wave number) $\boldsymbol{\omega}$ so that

$$(5.1) \qquad \lambda(\boldsymbol{\omega}) = \sup_{\boldsymbol{\beta} \neq \mathbf{0}} \left\{ \frac{f_X(\boldsymbol{\omega}; \boldsymbol{\beta})}{\sigma^2(\boldsymbol{\beta})} \right\},$$

where $\sigma^2(\boldsymbol{\beta}) = \mathrm{var}\{X_{\mathbf{t}}(\boldsymbol{\beta})\}$.

Building on the ideas of Section 2, identify the categorical random field, $X_{\mathbf{t}}$, with the $k \times 1$ multiple indicator process $\mathbf{Y}_{\mathbf{t}}$ defined by $\mathbf{Y}_{\mathbf{t}} = \mathbf{e}_j$ if $X_{\mathbf{t}} = c_j$, $j = 1, \ldots, k$, and $\mathbf{Y}_{\mathbf{t}} = \mathbf{0}$ if $X_{\mathbf{t}} = c_{k+1}$, where $\mathbf{e}_j$ is a $k \times 1$ vector with a one in the jth position and zeros elsewhere, and $\mathbf{0}$ is a $k \times 1$ vector of zeros. Assume the existence of the $k \times k$ nonsingular spectral density $f_Y(\boldsymbol{\omega})$. The scaled series can be written as $X_{\mathbf{t}}(\boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{Y}_{\mathbf{t}}$ and hence the wave spectrum of $X_{\mathbf{t}}$ can be written as $f_X(\boldsymbol{\omega}) = \boldsymbol{\beta}' f_Y^{\mathrm{re}}(\boldsymbol{\omega})\boldsymbol{\beta}$. The optimality criterion is

$$(5.2) \qquad \lambda(\boldsymbol{\omega}) = \sup_{\boldsymbol{\beta} \neq \mathbf{0}} \left\{ \frac{\boldsymbol{\beta}' f_Y^{\mathrm{re}}(\boldsymbol{\omega})\boldsymbol{\beta}}{\boldsymbol{\beta}' V \boldsymbol{\beta}} \right\},$$

where V is the variance-covariance matrix of $\mathbf{Y}_{\mathbf{t}}$. Again the problem is reduced to an eigenvalue problem. The optimality criterion $\lambda(\boldsymbol{\omega})$ and the resulting optimal scaling $\boldsymbol{\beta}(\boldsymbol{\omega})$ have interpretations analogous to the one-dimensional case described in Section 2. In addition, inference and estimation of $\lambda(\boldsymbol{\omega})$ and $\boldsymbol{\beta}(\boldsymbol{\omega})$ follow from Section 2 results.

The applications that motivate this are automated image retrieval and pattern recognition with potential use in computer vision. Efforts of digitizing massive archives of image, film and video have created a demand for automated retrieval systems. Such systems would save the time and effort needed to browse entire databases, for example, in medical image query; see Liu and Picard (1996) for a recent discussion. In particular, the spectral envelope appears to have potential in the analysis of textures.

Textures are homogeneous patterns or spatial arrangements of pixels that regional intensity or color alone does not sufficiently describe. Textures may consist of structured and/or random placement of elements, but also may be without fundamental subunits. Texture is an important element to human vision and has been found to provide cues to scene depth and surface orientation. Due to the diversity of textures appearing in natural images it is difficult to narrowly define texture.

Most research on texture is conducted on the Brodatz (1966) texture collection which provides a set of mostly homogeneous texture images. It has been hard to adequately model texture and the substantial body of work on texture has not yet produced any clear solutions for the problems of texture analysis, classification and synthesis. Several recent content-based image retrieval systems utilize texture feature sets to aid in the retrieval of images. Within the applications of satellite image retrieval, recent systems by Li and Turek (1996), for example, use texture to retrieve images based upon the detection of various features of the earth's terrain. The IBM QBIC system (Niblack et al., 1993) uses several texture features for the retrieval of photographic images. Both the whole images and manually identified regions are indexed by texture in the QBIC system.

Consider an $n_1 \times n_2$ image, say $\{X_{\mathbf{t}}; \mathbf{t} = (t_1, t_2); t_i = 1, \ldots, n_i; i = 1, 2\}$, using an 8-bit gray scale. That is, at each pixel, $\mathbf{t}$, $X_{\mathbf{t}}$ takes on one of $2^8 = 256$ values from the vector of gray levels, $\mathbf{g} = (0, 1, \ldots, 255)'$. Currently, there are many compression techniques that are based on transforms of the image. The current standard is JPEG, which uses the discrete cosine transform (see Rao and Yip, 1990, e.g.) applied to 8×8 pixel blocks of $\{X_{\mathbf{t}}\}$. Recently, there has been a considerable amount of interest in using wavelet transforms to obtain a more sparse decomposition of an image, thus requiring less storage space; for an introduction to wavelet analysis of images, see Bruce and Gao (1996).

Current compression methods consider the gray levels of an image as fixed, quantitative values that will not be altered. We feel that there may be some advantage to investigating the possibility of reducing the number of gray levels contained in an image

before any compression takes place. For example, in larger images, one may bin the 256 gray levels by a factor of four (leaving $2^6 = 64$ gray levels) as an initial reduction. In addition to using the spectral envelope as a compression method, we will also consider the benefits of using the methodology to reduce the number of gray levels needed to reproduce an image. This problem is examined in more detail in Wendt (1999).

For our example, we use the herringbone weave texture. Figure 18 shows the original 256×256 texture using an 8-bit grayscale. In Figure 19a we reduced the image to 128×128 pixels; in addition, the 8-bit gray scale was binned by a factor of 4, resulting in $2^6 = 64$, rather than $2^8 = 256$ gray levels. Although these reductions were performed primarily to reduce processing time, the reductions combined give the image its noisy appearance. Figure 20 shows the smoothed (using a Gaussian smoother) sample spectral envelope of the reduced image, where, for clarity, any value of the sample spectral envelope that is below an approximate $\alpha = 10^{-14}$ threshold is zeroed out. Due to folding, that is, $\lambda(\omega_x, \omega_y) = \lambda(-\omega_x, -\omega_y)$, we only display frequencies in $(-1/2, 1/2) \times (0, 1/2)$.

Optimal reconstruction of the image using all of the pertinent information in the spectral envelope and the corresponding palettes (scalings) is currently under investigation. Let $\{\mathbf{Y_t}; \mathbf{t} = (t_1, t_2); t_1 = 1, \ldots, n_1; t_2 = 1, \ldots, n_2\}$ denote an image (or part of an image) such as the reduced herringbone image in Figure 19a. As discussed above (5.2), $\mathbf{Y_t}$ is the indicator vector process associated with the image $X_\mathbf{t}$. The image, in terms of $\mathbf{Y_t}$, is decomposed as

$$(5.3) \qquad \mathbf{Y_t} = \sum_\mathbf{j} \mathbf{d(j)}\psi_\mathbf{t}(\mathbf{j}),$$

where $\{\psi_\mathbf{t}(\mathbf{j})\}$ forms an orthonormal basis over the two-dimensional lattice, and $\mathbf{j} = (j_1, j_2)$ for $j_i = 1, \ldots, n_i$, and $i = 1, 2$. Smaller values of $|\mathbf{d(j)}|$ are thresholded to the zero vector, and the image is approximated by

$$(5.4) \qquad \widehat{\mathbf{Y}}_\mathbf{t} = \sum_{\mathbf{j} \in \mathscr{I}} \mathbf{d(j)}\psi_\mathbf{t}(\mathbf{j}),$$

where $\mathscr{I}$ is a set containing a small number of indices $\mathbf{j}$.

We focus on Fourier analysis here, but the theory does not preclude the use of any other basis such as a wavelet basis or Walsh functions (e.g., Ferryanto, 1995 extends the work of Stoffer, 1987 to estimating the Walsh spectral density of two-dimensional random fields). In this case, $\psi_\mathbf{t}(\mathbf{j})$ in (5.3) are the complex exponentials, $\psi_\mathbf{t}(\mathbf{j}) = \exp\{-2\pi i(\omega_{j_1} t_1 + \omega_{j_2} t_2)\}$,

where $\omega_{j_i} = j_i/n_i$ for $i = 1, 2$, and the coefficients, $\mathbf{d(j)}$, are the corresponding DFTs. In this example, thresholding was accomplished by retaining values for which the estimated spectral envelope exceeded some small α-level threshold, and by zeroing out the remaining values (this is called hard shrinkage). Of course, continuous (soft) shrinkage functions such as those discussed in Donoho and Johnstone (1994, 1995) could be employed. In Figure 20, we used a hard shrinkage method with a threshold level at $\alpha = 10^{-14}$. Next, we consider the reconstruction of the image $X_\mathbf{t}$, say $\widehat{X}_\mathbf{t}$, using the estimate $\widehat{\mathbf{Y}}_\mathbf{t}$. Let $\mathbf{g} = (g_1, \ldots, g_k)'$ denote the $k \times 1$ vector of gray levels associated with the image. In terms of the original image and the associated indicator process, we have $X_\mathbf{t} = \mathbf{g}'Y_\mathbf{t}$. A natural candidate for the reconstruction of $X_\mathbf{t}$ would be $\widehat{X}_\mathbf{t} = \mathbf{g}'\widehat{Y}_\mathbf{t}$, appropriately discretized. However, we may use the information in $\widehat{Y}_\mathbf{t}$ to further reduce the dimension of $\mathbf{g}$. To do this, let $\mathbf{e}_j$ be the indicator vector associated with gray level g_j. Then we set $\widehat{X}_\mathbf{t} = g_j$ if $\|\widehat{Y}_\mathbf{t} - \mathbf{e}_j\| \leq \|\widehat{Y}_\mathbf{t} - \mathbf{e}_i\|$, for all $i = 1, \ldots, k$, where $\| \cdot \|$ is a norm. In this example we used Euclidean distance, but statistical distance would also be appropriate. In this way, we may get a reduction in the number of gray levels needed to reproduce the image.

Figures 19b, c show reconstructions of the image in Figure 19a based on this technique. In this example, we choose $\mathscr{I}$ in (5.4) using the two-dimensional extension of (2.4) at various choices of α. In one case (Figure 19b) we used $\alpha = 10^{-18}$ and this resulted in selecting seven frequency pairs to be included in $\mathscr{I}$. In the another case (Figure 19c) we used $\alpha = 10^{-14}$ and this resulted in $\mathscr{I}$ consisting of fourteen frequency pairs. Of course, thresholding at a smaller level yields a more precise image reconstruction. In addition, this method led to a reduction in gray levels from 44 for the original image, to 26 for the reduced image displayed in Figure 19c.

6. MATCHING SEQUENCES

In this section we consider the problem of quantifying the degree to which two categorical time series are coherent [see (1.8)]. The goal is to discover whether the sequences contain similar patterns and the problem is motivated by the matching of two DNA sequences (e.g., Waterman and Vingron, 1994). This approach builds on the ideas used in defining the spectral envelope for a qualitative-valued time series; technical details can be found in Stoffer and Tyler (1998). We continue to focus on methods that are computationally simple and fast and can be applied to long sequences.
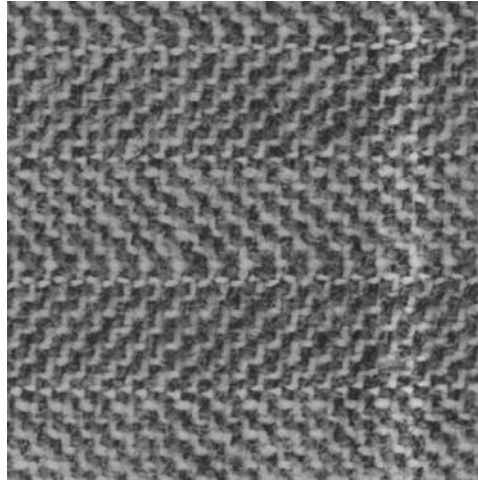
FIG. 18. *Herringbone weave from the Brodatz texture collection; this is a* 256×256 *image in 8-bit gray levels.*

6.1 The General Problem

In the general case, X_{1t} and X_{2t}, $t = 0, \pm 1, \pm 2, \ldots$, are categorical time series taking values in possibly different state-spaces of dimensions $k_1 + 1$ and $k_2 + 1$, respectively. Consider two nonconstant transformations g and h with $g(X_{1t})$ and $h(X_{2t})$ being real-valued time series such that $g(X_{1t})$ has continuous spectral density $f_{gg}(\omega)$ and $h(X_{2t})$ has continuous spectral density $f_{hh}(\omega)$. We denote the complex-valued cross-spectral density of the two series $g(X_{1t})$ and $h(X_{2t})$ by $f_{gh}(\omega)$. As discussed in Section 1.1, a measure of the degree of similarity between the sequences $g(X_{1t})$ and $h(X_{2t})$ at frequency ω is the squared coherency

$$(6.1) \qquad \rho_{gh}^2(\omega) = \frac{|f_{gh}(\omega)|^2}{f_{gg}(\omega) f_{hh}(\omega)}.$$

Of course the value of $\rho_{gh}^2(\omega)$ will depend on the choices of the transformations g and h. If X_{1t} and X_{2t} are independent, then so are $g(X_{1t})$ and $h(X_{2t})$, for any g and h, in which case $\rho_{gh}^2(\omega) = 0$ for all ω. The main goal here is to find g and h, under various constraints, to maximize the squared coherency $\rho_{gh}^2(\omega)$. If the maximized value of $\rho_{gh}^2(\omega)$ is small we can say that the two sequences X_{1t} and X_{2t} do not match at frequency ω. If the maximized value of $\rho_{gh}^2(\omega)$ is large, then the resulting transformations g and h can help in understanding the nature of the similarity between the two sequences.

Analogous to Section 2, we identify the categorical sequence X_{1t} with the multiple indicator process $\mathbf{Y}_{1t}$. Recall that $\mathbf{Y}_{1t}$ is a $k_1 \times 1$ vector with a one in the jth position if X_{1t} is in state j ($j = 1, \ldots, k_1$) at time t and zeros elsewhere. If X_{1t} is in state $k_1 + 1$, then $\mathbf{Y}_{1t}$ is the zero vector. Similarly, we identify
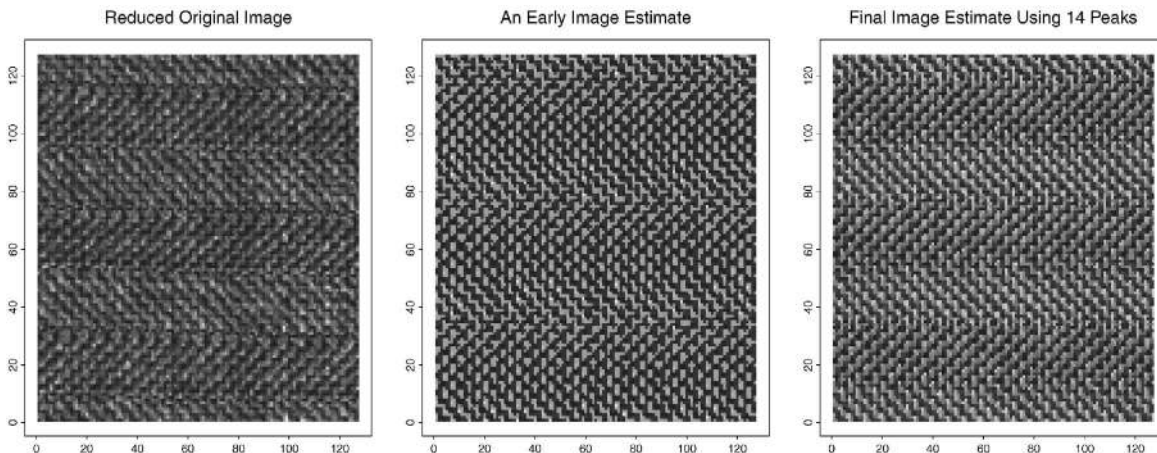

FIG. 19. (a) *The reduced herringbone weave image* (*compare to Figure* 18); *this is* 128×128 *using a 6-bit gray scale*; (b) *reconstruction using the seven largest spectral envelope peaks* ($\alpha = 10^{-18}$); (c) *reconstruction using the fourteen largest spectral envelope peaks* ($\alpha = 10^{-14}$).
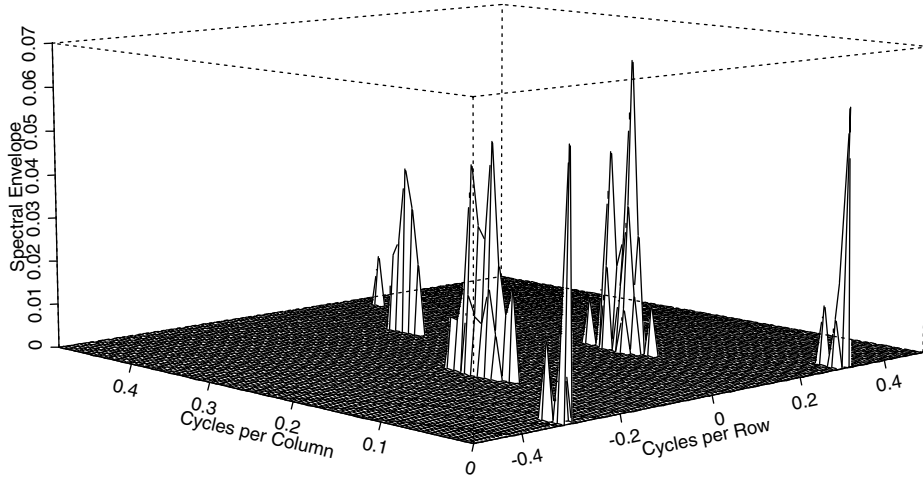
FIG. 20. *The smoothed sample spectral envelope,* $\widehat{\lambda}(\omega)$, *for the* 128×128 *herringbone weave image, using a 6-bit gray scale, as shown in Figure* 19a. *For clarity, small values are shown as zero.*

X_{2t} with the $k_2 \times 1$ multiple indicator processes $\mathbf{Y}_{2t}$. We assume the existence of the $k_i \times k_i$ $(i = 1, 2)$, nonsingular spectral density matrices $f_{11}(\omega)$ and $f_{22}(\omega)$ of $\mathbf{Y}_{1t}$ and $\mathbf{Y}_{2t}$, respectively, and denote the $k_1 \times k_2$ cross-spectral matrix between $\mathbf{Y}_{1t}$ and $\mathbf{Y}_{2t}$ by $f_{12}(\omega)$.

To describe the problem in terms of scaling categorical time series, let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{k_1})' \in \mathbf{R}^{k_1}$, $\boldsymbol{\alpha} \neq \mathbf{0}$, be a vector of scalings associated with the categories of the first series, X_{1t}, and let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{k_2})' \in \mathbf{R}^{k_2}$, $\boldsymbol{\beta} \neq \mathbf{0}$, be a vector of scalings associated with the categories of the second series, X_{2t}. That is, define the real-valued series

$$
\begin{aligned}
&X_{1t}(\boldsymbol{\alpha}) = \alpha_j \quad \text{if } X_{1t} \text{ is in state } j \\
&\qquad\qquad\qquad\qquad \text{for } j = 1, \ldots, k_1, \\
(6.2)\quad &X_{2t}(\boldsymbol{\beta}) = \beta_j \quad \text{if } X_{2t} \text{ is in state } j \\
&\qquad\qquad\qquad\qquad \text{for } j = 1, \ldots, k_2,
\end{aligned}
$$

where, in addition, $X_{1t}(\boldsymbol{\alpha}) = 0$ if X_{1t} is in state $k_1 + 1$, and $X_{2t}(\boldsymbol{\beta}) = 0$ if X_{2t} is in state $k_2 + 1$.

Since the scaled series can be written as $X_{1t}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}'\mathbf{Y}_{1t}$, and $X_{2t}(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_{2t}$, the squared-coherency between $X_{1t}(\boldsymbol{\alpha})$ and $X_{2t}(\boldsymbol{\beta})$ can be written as

$$
(6.3)\quad \rho_{12}^2(\omega; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{|\boldsymbol{\alpha}' f_{12}(\omega)\boldsymbol{\beta}|^2}{\left[\boldsymbol{\alpha}' f_{11}^{\mathrm{re}}(\omega)\boldsymbol{\alpha}\right]\left[\boldsymbol{\beta}' f_{22}^{\mathrm{re}}(\omega)\boldsymbol{\beta}\right]}.
$$

Setting $\mathbf{a} = f_{11}^{\mathrm{re}}(\omega)^{1/2}\boldsymbol{\alpha}$ and $\mathbf{b} = f_{22}^{\mathrm{re}}(\omega)^{1/2}\boldsymbol{\beta}$, subject to $\mathbf{a}'\mathbf{a} = 1$ and $\mathbf{b}'\mathbf{b} = 1$, define

$$
\begin{aligned}
(6.4)\quad Q(\omega) &= f_{11}^{\mathrm{re}}(\omega)^{-1/2} f_{12}(\omega) f_{22}^{\mathrm{re}}(\omega)^{-1/2} \\
&= Q^{\mathrm{re}}(\omega) + i Q^{\mathrm{im}}(\omega)
\end{aligned}
$$

and write (6.3) as

$$
(6.5)\quad \rho_{12}^2(\omega; \mathbf{a}, \mathbf{b}) = \left[\mathbf{a}' Q^{\mathrm{re}}(\omega)\mathbf{b}\right]^2 + \left[\mathbf{a}' Q^{\mathrm{im}}(\omega)\mathbf{b}\right]^2.
$$

The goal is to find $\mathbf{a}$ and $\mathbf{b}$ to maximize (6.5) for each ω of interest. Several approaches to the maximization are available; one approach is based on the following observation.

RESULT 1. *Fix* ω *and drop it from the notation. Then* (6.5) *can be written as*

$$
\begin{aligned}
(6.6)\quad \rho_{12}^2(\mathbf{a}, \mathbf{b}) &= \mathbf{a}'\left(Q^{\mathrm{re}}\mathbf{b}\mathbf{b}'Q^{\mathrm{re}} + Q^{\mathrm{im}}\mathbf{b}\mathbf{b}'Q^{\mathrm{im}}\right)\mathbf{a} \\
&= \mathbf{b}'\left(Q^{\mathrm{re}}\mathbf{a}\mathbf{a}'Q^{\mathrm{re}} + Q^{\mathrm{im}}\mathbf{a}\mathbf{a}'Q^{\mathrm{im}}\right)\mathbf{b}.
\end{aligned}
$$

Let $\mathbf{b}_0$ *be an arbitrary real-valued* $k_2 \times 1$ *unit length vector. Define the sequence of vectors* $\mathbf{a}_j$ *to be the eigenvector corresponding to the largest root of the at most* rank 2, *nonnegative definite matrix*

$$
(6.7)\quad Q^{\mathrm{re}}\mathbf{b}_{j-1}\mathbf{b}_{j-1}'Q^{\mathrm{re}'} + Q^{\mathrm{im}}\mathbf{b}_{j-1}\mathbf{b}_{j-1}'Q^{\mathrm{im}'}
$$

and the sequence $\mathbf{b}_j$ *to be the eigenvector corresponding to the largest root of the at most* rank 2, *nonnegative definite matrix*

$$
(6.8)\quad Q^{\mathrm{re}'}\mathbf{a}_j\mathbf{a}_j'Q^{\mathrm{re}} + Q^{\mathrm{im}'}\mathbf{a}_j\mathbf{a}_j'Q^{\mathrm{im}},
$$

for $j = 1, 2, \ldots$. *Then, from the first part of* (6.6) *it follows that* $\rho^2(\mathbf{a}_{j+1}, \mathbf{b}_j) \geq \rho^2(\mathbf{a}, \mathbf{b}_j)$ *for any* $\mathbf{a}$ *of unit length, and from the second part of* (6.6) *it follows that* $\rho^2(\mathbf{a}_{j+1}, \mathbf{b}_{j+1}) \geq \rho^2(\mathbf{a}_{j+1}, \mathbf{b})$ *for any* $\mathbf{b}$ *of unit length. Thus,*

$$(6.9) \quad \rho^2(\mathbf{a}_{j+1}, \mathbf{b}_{j+1}) \geq \rho^2(\mathbf{a}_{j+1}, \mathbf{b}_j) \geq \rho^2(\mathbf{a}_j, \mathbf{b}_j).$$

The algorithm described by Result 1 can be used to find the optimal scalings at each frequency, ω, of interest. The algorithm is initialized by setting $\mathbf{b}_0$ equal to either $\mathscr{E}_1[Q^{\mathrm{re}}(\omega)'Q^{\mathrm{re}}(\omega)]$ or $\mathscr{E}_1[Q^{\mathrm{im}}(\omega)'Q^{\mathrm{im}}(\omega)]$, depending on which vector [we denote the eigenvector corresponding to the largest eigenvalue of matrix A by $\mathscr{E}_1(A)$] produces the larger value of (6.5) for arbitrary $\mathbf{a}$. In turn, $\boldsymbol{\alpha}(\omega)$ and $\boldsymbol{\beta}(\omega)$ can be taken proportional to $f_{11}^{\mathrm{re}}(\omega)^{-1/2}\mathbf{a}(\omega)$ and $f_{22}^{\mathrm{re}}(\omega)^{-1/2}\mathbf{b}(\omega)$, respectively, where $\mathbf{a}(\omega)$ and $\mathbf{b}(\omega)$ maximize (6.5). Note that the algorithm requires only the computation of latent roots and vectors of at most rank 2, nonnegative definite matrices, regardless of the dimension of the state-spaces. Moreover, by (6.9), the objective function increases with each step. Unfortunately, it does not guarantee convergence to the global maximum. From simulations, however, it appears to be the case that the algorithm usually converges; Stoffer and Tyler (1998) also provide tight bounds for the maximum of (6.5).

6.2 Common Scalings

In many cases, the processes X_{1t} and X_{2t} are defined on the same state-space, $\mathscr{S} = \{c_1, \ldots, c_{k+1}\}$, for example, DNA sequences. To enhance the interpretation in such cases, it would be appropriate to choose *common scalings*. Henceforth, set $k_1 = k_2 = k$, and assume that $\mathbf{Y}_{1t}$ and $\mathbf{Y}_{2t}$ have the same spectra, that is, $f_{11}(\omega) = f_{22}(\omega) = f(\omega)$, for at least all ω of interest. Realistic models that satisfy these conditions will be discussed later.

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)' \in \mathbf{R}^k$, $\boldsymbol{\beta} \neq \mathbf{0}$, be a vector of scalings common to the categories of both series, that is, the real-valued series are

$$X_{1t}(\boldsymbol{\beta}) = \beta_j \quad \text{if } X_{1t} = c_j,$$
$$X_{2t}(\boldsymbol{\beta}) = \beta_j \quad \text{if } X_{2t} = c_j,$$

for $j = 1, \ldots, k$. The scale associated with category c_{k+1} is held fixed at zero for both sequences. We restrict attention to the frequencies ω for which $f_{11}(\omega) = f_{22}(\omega) = f(\omega)$. Since the scaled series are $X_{it}(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_{it}$, for $i = 1, 2$, the squared coherency between $X_{1t}(\boldsymbol{\beta})$ and $X_{2t}(\boldsymbol{\beta})$ can be written as

$$(6.10) \qquad \rho_{12}^2(\omega; \boldsymbol{\beta}) = \frac{|\boldsymbol{\beta}' f_{12}(\omega)\boldsymbol{\beta}|^2}{[\boldsymbol{\beta}' f^{\mathrm{re}}(\omega)\boldsymbol{\beta}]^2}.$$

Setting $\mathbf{b} = f^{\mathrm{re}}(\omega)^{1/2}\boldsymbol{\beta}$, subject to the standardization $\mathbf{b}'\mathbf{b} = 1$, and writing

$$(6.11) \quad \begin{aligned} Q(\omega) &= f^{\mathrm{re}}(\omega)^{-1/2}f_{12}(\omega)f^{\mathrm{re}}(\omega)^{-1/2} \\ &= Q^{\mathrm{re}}(\omega) + i\, Q^{\mathrm{im}}(\omega), \end{aligned}$$

we may write (6.10) as

$$(6.12) \quad \rho_{12}^2(\omega; \mathbf{b}) = [\mathbf{b}' Q^{\mathrm{re}}(\omega)\mathbf{b}]^2 + [\mathbf{b}' Q^{\mathrm{im}}(\omega)\mathbf{b}]^2.$$

Although $Q^{\mathrm{re}}(\omega)$ and $Q^{\mathrm{im}}(\omega)$ in (6.12) are not necessarily symmetric, we may assume, without loss of generality, that they are since

$$(6.13) \quad \begin{aligned} \rho_{12}^2(\omega; \mathbf{b}) &= [\mathbf{b}' Q^{\mathrm{re}}(\omega)\mathbf{b}]^2 + [\mathbf{b}' Q^{\mathrm{im}}(\omega)\mathbf{b}]^2 \\ &= [\mathbf{b}' Q_s^{\mathrm{re}}(\omega)\mathbf{b}]^2 + [\mathbf{b}' Q_s^{\mathrm{im}}(\omega)\mathbf{b}]^2, \end{aligned}$$

where

$$(6.14) \quad \begin{aligned} Q_s^{\mathrm{re}}(\omega) &= [Q^{\mathrm{re}}(\omega) + Q^{\mathrm{re}}(\omega)']/2 \quad \text{and} \\ Q_s^{\mathrm{im}}(\omega) &= [Q^{\mathrm{im}}(\omega) + Q^{\mathrm{im}}(\omega)']/2, \end{aligned}$$

Our goal is to find $\mathbf{b}$ to maximize (6.13) for each ω of interest. The maximization can still be accomplished iteratively via the algorithm (6.7) and (6.8) in conjunction with the following result.

RESULT 2. *Under the conditions described in Section 4.1, if in Result 1, $k_1 = k_2 = k$ and the matrices Q^{re} and Q^{im} are symmetric, the maximum value of $\rho_{12}^2(\mathbf{a}, \mathbf{b})$ is attained when $\mathbf{a} = \mathbf{b}$.*

Since $Q_s^{\mathrm{re}}(\omega)$ and $Q_s^{\mathrm{im}}(\omega)$ are symmetric, Result 2 can be used to maximize (6.13), initializing the algorithm by setting $\mathbf{b}_0$ equal to either $\mathscr{E}_1[Q_s^{\mathrm{re}}(\omega)^2]$ or $\mathscr{E}_1[Q_s^{\mathrm{im}}(\omega)^2]$, depending on which vector produces the larger value of $\rho_{12}^2(\omega, \mathbf{b}_0)$. The sequence

$$(6.15) \quad \begin{aligned} \mathbf{b}_j = \mathscr{E}_1\big[&Q_s^{\mathrm{re}}(\omega)\mathbf{b}_{j-1}\mathbf{b}_{j-1}'Q_s^{\mathrm{re}}(\omega) \\ &+ Q_s^{\mathrm{im}}(\omega)\mathbf{b}_{j-1}\mathbf{b}_{j-1}'Q_s^{\mathrm{im}}(\omega)\big], \end{aligned}$$

for $j = 1, 2, \ldots$, replaces the alternating sequences defined in (6.7) and (6.8). Note that $\rho_{12}^2(\omega; \mathbf{b}_j) \geq \rho_{12}^2(\omega; \mathbf{b}_{j-1})$. The optimal scaling, $\boldsymbol{\beta}(\omega)$, is chosen proportional to $f^{\mathrm{re}}(\omega)^{-1/2}\mathbf{b}(\omega)$, where $\mathbf{b}(\omega)$ maximizes (6.13). Another important consequence of Result 2 is that it gives sufficient conditions under which choosing common scales is not only parsimonious but optimal. Specifically, if $Q^{\mathrm{re}}(\omega)$ and $Q^{\mathrm{im}}(\omega)$ are both symmetric, then the maximum of $\rho_{12}^2(\omega, \mathbf{a}, \mathbf{b})$ [see (6.5)] is achieved when $\mathbf{a} = \mathbf{b}$.

6.3 Models and Applications

For practical applications of the theory presented in Section 4.2, we address two problems. First is the case where the two sequences under investigation are in phase and contain at most one common pattern. They may be subsequences of larger sequences. This will be termed *local alignment*. The case of *global alignment*, where we do not assume

that the sequences are in phase will be discussed next. The local model is

$$(6.16) \qquad \mathbf{Y}_{it} = \mathbf{p}_i + \mathbf{S}_t + \mathbf{e}_{it},$$

where $\mathbf{p}_i = (p_{i1}, \ldots, p_{ik})'$ is the vector of positive probabilities $p_{ij} = \Pr(X_{it} = c_j)$, for $i = 1, 2$ and $j = 1, \ldots, k$. In addition, $\mathbf{S}_t$ is a realization of a stationary $k \times 1$ vector-valued time series that is uncorrelated with the stationary $k \times 1$ vector-valued series $\mathbf{e}_{it}$, $i = 1, 2$. There may be some dependence structure between $\mathbf{S}_t$ and $\mathbf{e}_{it}$; refer to Stoffer (1987) for details. Furthermore, $\mathbf{S}_t$ has $k \times k$ spectral density matrix $\mathbf{f}_{ss}(\omega)$, and $\mathbf{e}_{it}$, $i = 1, 2$, have common $k \times k$ spectra denoted by $\mathbf{f}_{ee}(\omega)$. It is hypothesized that the process $\mathbf{S}_t$ is common to both sequences.

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)' \in \mathbf{R}^k$, $\boldsymbol{\beta} \neq \mathbf{0}$, be a vector of scalings associated with the categories $\{c_1, \ldots, c_k\}$. As before, define the real-valued series $X_i(t, \boldsymbol{\beta}) = \beta_j$ if $X_{it} = c_j$, $j = 1, \ldots, k$, and $X_i(t, \beta) = 0$ if $X_{it} = c_{k+1}$, for $i = 1, 2$. It can be shown that, in this case, the conditions of Result 2 are met, and hence, *the optimal strategy is to select the common scales for the sequences X_{1t} and X_{2t}.*

Note that $X_{it}(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_{it} = \boldsymbol{\beta}'\mathbf{p}_i + \boldsymbol{\beta}'\mathbf{S}_t + \boldsymbol{\beta}'\mathbf{e}_{it}$, for $i = 1, 2$. Let $f_{11}(\omega; \boldsymbol{\beta})$ be the spectrum of scaled process $X_{1t}(\boldsymbol{\beta})$; similarly, let $f_{22}(\omega; \boldsymbol{\beta})$ denote the spectrum of $X_{2t}(\boldsymbol{\beta})$ and let $f_{12}(\omega; \boldsymbol{\beta})$ denote the cross-spectrum between $X_{1t}(\boldsymbol{\beta})$ and $X_{2t}(\boldsymbol{\beta})$. The following conditions hold:

$$(6.17) \quad \begin{aligned} f_{ii}(\omega; \boldsymbol{\beta}) &= \boldsymbol{\beta}'\{f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)\}\boldsymbol{\beta}, \quad i = 1, 2, \\ f_{12}(\omega; \boldsymbol{\beta}) &= \boldsymbol{\beta}' f_{ss}^{\mathrm{re}}(\omega)\boldsymbol{\beta}. \end{aligned}$$

The coherence between $X_{1t}(\boldsymbol{\beta})$ and $X_{2t}(\boldsymbol{\beta})$ is seen to be

$$(6.18) \qquad \rho_{12}(\omega; \boldsymbol{\beta}) = \frac{\boldsymbol{\beta}' f_{ss}^{\mathrm{re}}(\omega)\boldsymbol{\beta}}{\boldsymbol{\beta}'[f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)]\boldsymbol{\beta}}.$$

If $f_{ss}(\omega) = 0$, then $\rho_{12}(\omega; \boldsymbol{\beta}) = 0$ for any scaling $\boldsymbol{\beta}$. Thus, the detection of a common signal can be achieved by considering the maximal coherency under the model conditions. Setting $\mathbf{b} = [f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)]^{1/2}\boldsymbol{\beta}$, subject to $\mathbf{b}'\mathbf{b} = 1$, write (6.18) as

$$(6.19) \quad \begin{aligned} \rho_{12}(\omega; \mathbf{b}) = {}&\mathbf{b}'\big[f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)\big]^{-1/2} f_{ss}^{\mathrm{re}}(\omega) \\ &\times \big[f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)\big]^{-1/2}\mathbf{b}. \end{aligned}$$

Hence, the problem is again an eigenvalue problem, and the maximum value of (6.19) is the largest scalar $\lambda(\omega)$ such that

$$(6.20) \quad \begin{aligned} \big[f_{ss}^{\mathrm{re}}(\omega) &+ f_{ee}^{\mathrm{re}}(\omega)\big]^{-1/2} f_{ss}^{\mathrm{re}}(\omega) \\ &\times \big[f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)\big]^{-1/2}\mathbf{b}(\omega) = \lambda(\omega)\mathbf{b}(\omega). \end{aligned}$$

The optimal scaling, $\boldsymbol{\beta}(\omega)$, is taken proportional to $[f_{ss}^{\mathrm{re}}(\omega) + f_{ee}^{\mathrm{re}}(\omega)]^{-1/2}\mathbf{b}(\omega)$. This value will maximize the coherency at frequency ω between the two sequences, with the maximum value being $\lambda(\omega)$. That is, $\rho_{12}(\omega; \boldsymbol{\beta}) \leq \rho_{12}(\omega; \boldsymbol{\beta}(\omega)) = \lambda(\omega)$,

with equality only when $\boldsymbol{\beta}$ is proportional to $\boldsymbol{\beta}(\omega)$. Estimation proceeds in an obvious way: given consistent (smoothed) estimates $\widehat{f}_{ij}(\omega)$, for $i, j = 1, 2$ as described in Section 1.1, put

$$(6.21) \quad \begin{aligned} \widehat{f}_{ss}^{\mathrm{re}}(\omega) &= \big[\widehat{f}_{12}^{\mathrm{re}}(\omega) + \widehat{f}_{21}^{\mathrm{re}}(\omega)\big]/2 \quad \text{and} \\ \widehat{f}_{ss}^{\mathrm{re}}(\omega) + \widehat{f}_{ee}^{\mathrm{re}}(\omega) &= \big[\widehat{f}_{11}^{\mathrm{re}}(\omega) + \widehat{f}_{22}^{\mathrm{re}}(\omega)\big]/2. \end{aligned}$$

A frequency-based test of the null hypothesis that $f_{ss}(\omega) = 0$ was also developed for this model. We will not discuss the test here, but details can be found in Stoffer and Tyler (1998).

The model can be extended to include the possibility that there are more than one signal common to each sequence and that the sequences are not necessarily aligned. The *global model* is

$$(6.22) \quad \begin{aligned} \mathbf{Y}_{1t} &= \mathbf{p}_1 + \sum_{j=1}^{q} \mathbf{S}_{jt} + \mathbf{e}_{1t} \quad \text{and} \\ \mathbf{Y}_{2t} &= \mathbf{p}_2 + \sum_{j=1}^{q} \mathbf{S}_{j, t-\tau_j} + \mathbf{e}_{2t}, \end{aligned}$$

where $\mathbf{S}_{jt}$, $j = 1, \ldots, q$, are zero-mean realizations of stationary $k \times 1$ vector-valued time series that are mutually uncorrelated, and in addition are uncorrelated with the zero-mean, stationary $k \times 1$ vector-valued series $\mathbf{e}_{1t}$ and $\mathbf{e}_{2t}$. Furthermore, $\mathbf{S}_{jt}$ has $k \times k$ spectral density matrix $f_{S_j}(\omega)$, $j = 1, \ldots, q$, and $\mathbf{e}_{it}$, $i = 1, 2$, have common $k \times k$ spectra denoted by $f_{ee}(\omega)$. It is hypothesized that the processes $\mathbf{S}_{jt}$ are (stochastic) signals that are common to both time series X_{1t} and X_{2t}, or equivalently, $\mathbf{Y}_{1t}$ and $\mathbf{Y}_{2t}$.

There is no need to specify the phase shifts, $\tau_1, \ldots, \tau_q$, or the integer $q \geq 0$; however, the problem of their estimation is interesting. We consider the following method to help decide whether or not $q = 0$. Using the notation established in this section with $\omega \in [-\pi, \pi]$,

$$(6.23) \quad \begin{aligned} f_{11}(\omega) = f_{22}(\omega) &= \sum_{j=1}^{q} f_{S_j}(\omega) + f_{ee}(\omega) \quad \text{and} \\ f_{12}(\omega) &= \sum_{j=1}^{q} f_{S_j}(\omega) \exp(i\omega\tau_j). \end{aligned}$$

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)' \in \mathbf{R}^k$, $\boldsymbol{\beta} \neq 0$, be a vector of scalings, write $X_{it}(\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{Y}_{it}$, for $i = 1, 2$, so that the squared coherency between $X_{1t}(\boldsymbol{\beta})$ and $X_{2t}(\boldsymbol{\beta})$ is

$$(6.24) \quad \rho_{12}^2(\omega; \boldsymbol{\beta}) = \frac{\left|\sum_{j=1}^{q} \boldsymbol{\beta}' f_{S_j}^{\mathrm{re}}(\omega)\boldsymbol{\beta} \exp(i\omega\tau_j)\right|^2}{\left|\boldsymbol{\beta}' f^{\mathrm{re}}(\omega)\boldsymbol{\beta}\right|^2},$$

where $f(\omega) = f_{11}(\omega) = f_{22}(\omega)$. Setting $\mathbf{b} = f^{\mathrm{re}}(\omega)^{1/2}\boldsymbol{\beta}$, with the constraint $\mathbf{b}'\mathbf{b} = 1$, write

(6.24) as

$$
\rho_{12}^2(\omega; \mathbf{b}) = \left| \mathbf{b}' \left\{ \sum_{j=1}^{q} f^{\mathrm{re}}(\omega)^{-1/2} f_{S_j}^{\mathrm{re}}(\omega) \right. \right.
$$

(6.25)

$$
\left. \left. \times f^{\mathrm{re}}(\omega)^{-1/2} \exp(i\omega\tau_j) \right\} \mathbf{b} \right|^2 .
$$

Define the complex-valued matrix $Q(\omega)$ as

$$
Q(\omega) = \sum_{j=1}^{q} f^{\mathrm{re}}(\omega)^{-1/2} f_{S_j}^{\mathrm{re}}(\omega)
$$

(6.26)

$$
\times f^{\mathrm{re}}(\omega)^{-1/2} \exp(i\omega\tau_j)
$$

$$
= Q^{\mathrm{re}}(\omega) + i \, Q^{\mathrm{im}}(\omega),
$$

and note that both $Q^{\mathrm{re}}(\omega)$ and $Q^{\mathrm{im}}(\omega)$ are symmetric matrices (but not necessarily positive definite). As noted in Section 4.2, Result 2, it follows that *the optimal strategy is to select the scalings to be the same for both sequences*. Now, write (6.25) as

(6.27) $\rho_{12}^2(\omega; \mathbf{b}) = \left[\mathbf{b}' Q^{\mathrm{re}}(\omega) \mathbf{b} \right]^2 + \left[\mathbf{b}' Q^{\mathrm{im}}(\omega) \mathbf{b} \right]^2.$

Given consistent spectral estimates $\widehat{f}_{ij}(\omega)$, we can estimate $f(\omega)$ by $\widehat{f}(\omega) = \frac{1}{2}[\widehat{f}_{11}(\omega) + \widehat{f}_{22}(\omega)]$ so that consistent estimates of $Q^{\mathrm{re}}(\omega)$ and $Q^{\mathrm{im}}(\omega)$ are, respectively,

$$
\widehat{Q}^{\mathrm{re}}(\omega) = \left[\widehat{f}_{11}^{\mathrm{re}}(\omega) + \widehat{f}_{22}^{\mathrm{re}}(\omega) \right]^{-1/2}
$$

(6.28)

$$
\times \left[\widehat{f}_{12}^{\mathrm{re}}(\omega) + \widehat{f}_{21}^{\mathrm{re}}(\omega) \right]
$$

$$
\times \left[\widehat{f}_{11}^{\mathrm{re}}(\omega) + \widehat{f}_{22}^{\mathrm{re}}(\omega) \right]^{-1/2};
$$

$$
\widehat{Q}^{\mathrm{im}}(\omega) = \left[\widehat{f}_{11}^{\mathrm{re}}(\omega) + \widehat{f}_{22}^{\mathrm{re}}(\omega) \right]^{-1/2}
$$

(6.29)

$$
\times \left[\widehat{f}_{12}^{\mathrm{im}}(\omega) - \widehat{f}_{21}^{\mathrm{im}}(\omega) \right]
$$

$$
\times \left[\widehat{f}_{11}^{\mathrm{re}}(\omega) + \widehat{f}_{22}^{\mathrm{re}}(\omega) \right]^{-1/2} .
$$

The estimated squared coherency can be maximized via Result 2 with $Q^{\mathrm{re}}(\omega)$ and $Q^{\mathrm{im}}(\omega)$ in (6.27) replaced by their estimates (6.28) and (6.29), respectively. In particular, the recursion (6.15), with $Q_s^{\mathrm{re}}(\omega)$ and $Q_s^{\mathrm{im}}(\omega)$ replaced by (6.28) and (6.29), can be employed. The estimated optimal scaling vector at any particular frequency, $\widehat{\boldsymbol{\beta}}(\omega)$, is taken proportional to $\widehat{f}^{\mathrm{re}}(\omega)^{-1/2}\widehat{\mathbf{b}}(\omega)$, where $\widehat{\mathbf{b}}(\omega)$ is the maximizing vector. Some discussion of the finite sample null behavior in this case is given in Stoffer and Tyler (1998).

We illustrate the concepts in this section by matching two DNA sequences. In Section 2 we saw a rather strange result about the gene BNRF1 of the Epstein–Barr virus (EBV). There, it was found that although a cycle of 1/3 could be found in most of the gene, the last 1000 bp appeared to contain no cyclic behavior and might be considered to be non-coding; see Figures 5 and 6. Herpesvirus saimiri (HVS) also contains a gene labeled BNRF1 because

similarities between the two have been noted by molecular biologists. The spectral envelope of the entire HVS-BNRF1 gene looks similar to Figure 5, but, unlike Figure 6, HVS-BNRF1 has considerable power at frequency 1/3 in the final 1000 bp. It is of interest to know if the two genes match in the final 1000 bp even though there is no evidence that the last part of EBV-BNRF1 is actually coding. Figure 21 shows the maximum squared coherency using two models, the local model [thick line], (6.16), and the global model [thin line], (6.22). Because the local model is contained in the global model, it will always be the case that the result from the global model will envelope the result from the local fit, assuming that the spectra are estimated in the same way (that is, in Figure 21, the thick line is never above the thin line). In both cases, triangular smoothing with $m = 15$ was used. The two methods are in agreement, but the evidence that the series match at $\omega = 1/3$ appears stronger in the local case. In fact, there is a significant peak at $\omega = 1/3$ (at the 0.01 level when using the maximum F-statistic approach; this is detailed in Stoffer and Tyler, 1998). Thus, using the local model, we are lead to conclude that there is a significant match between the two genes in the final 1000 bp. The estimated optimal common scaling at $\omega = 1/3$ for the local model was A = 59.4, C = 0.8, G = 64.9, T = 0 (the global model had A = 60.8, C = 5.6, G = 67.1, T = 0) which indicates that the match is in the purine-pyrimidine (A = G, C = T) alphabet.

7. DISCUSSION AND OTHER REFERENCES

There are a number of texts devoted primarily to spectral analysis, for example Bloomfield (1976), Brillinger (1975), Hannan (1970), Priestley (1981) and Percival and Walden (1993). There are also many texts on time series analysis that present spectral analysis integrated with time domain (regression) analysis, for example, Brockwell and Davis (1991), Chatfield (1989), Fuller (1995) and Shumway and Stoffer (2000). The material on scaling time series is rather sparse and we do not know of any particular references besides those already mentioned. The basic idea, however, has been extensively used for the analysis of contingency tables and regression with qualitative variables; these come under a number of different titles such as *dual scaling*, for example, Nishisato (1980) and *correspondence analysis*, for example, Greenacre (1984). These and related topics are also discussed in Breiman and Friedman (1985) where the focus is on obtaining optimal transformations, numeri-
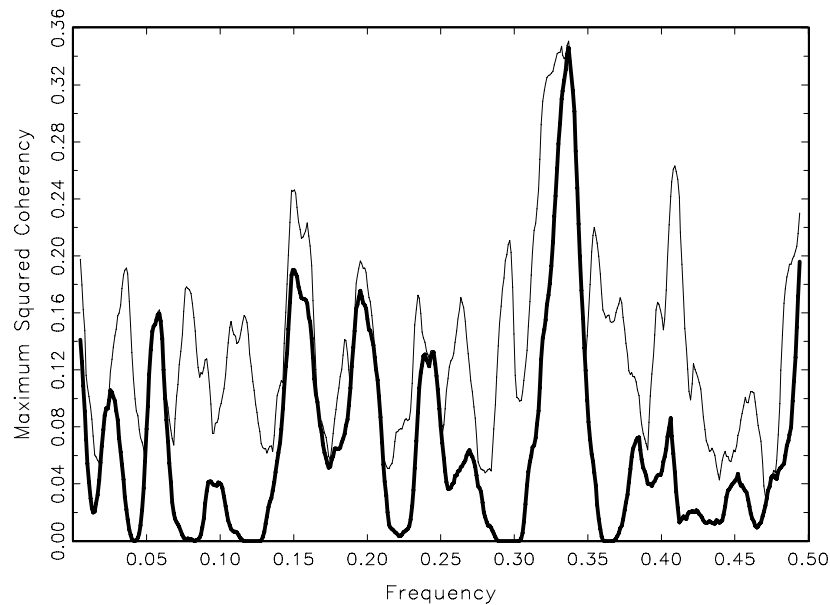
FIG. 21. *Maximum squared coherency between* EBV-BNRF1 *and* HVS-BNRF1 *using two models, the local model* [*thick line*], (6.16), *and the global model* [*thin line*], (6.22).

cally, in various situations. For a recent survey, see Michailidis and De Leeuw (1998).

Fourier analysis of categorical time series has been applied successfully in molecular genetics for quite some time. For example, McLachlan and Stewart (1976) and Eisenberg, Weiss and Tewillger (1984) studied the periodicity in proteins with Fourier analysis. They used predefined scales (or alphabets) and observed the $\omega = 1/3.6$ frequency of amphipatic helices. Because predetermination of the scaling is arbitrary and may not be optimal, Cornette et al. (1987) reversed the problem, and started with a frequency of $\omega_0 = 1/3.6$ and proposed a method to establish an "optimal" scaling at $\omega_0 = 1/3.6$. In this setting, optimality roughly refers to the fact that the scaled (numerical) sequence is maximally correlated with the sinusoid that oscillates at a frequency of ω_0. Viari, Soldano and Ollivier (1990) generalized this approach to a systematic calculation of a type of spectral envelope (which they called λ-graphs) and of the corresponding optimal scalings over all fundamental frequencies. While the aforementioned authors dealt exclusively with amino acid sequences, various forms of harmonic analysis have been applied to DNA by, for example, Tavaré and Giddings (1989), and in connection to nucleosome positioning by Satchwell, Drew and Travers (1986) and Bina (1994). The basic technique of the spectral envelope for categorical time series is similar to the methods established in Tavaré and Giddings (1989) and Viari, Soldano and Ollivier (1990); however, there

are some differences. In particular, the techniques differ by the optimality criterion used. Also, the spectral envelope methodology is developed in a statistical (rather than visual) setting to allow the investigator to distinguish between significant results and those results that can be attributed to chance.

As indicated in Section 2, the spectral envelope and related topics could come under the general titles of principal component analysis or canonical analysis of time series in the spectral domain. These topics are discussed in detail in Chapters 9 and 10 of Brillinger (1975) and there is a connection between Brillinger's work and ours. Specifically, the spectral envelope can be viewed as a special case of Brillinger's principal components. In the language of Brillinger (1975, Section 9.3), suppose we want to approximate $\mathbf{Y}_t$, a $k \times 1$ stationary time series with mean $\boldsymbol{\mu}_Y$, variance-covariance matrix V, and spectral matrix $f_Y(\omega)$, by finding a scalar process, Z_t, defined by

$$(7.1) \qquad Z_t = \sum_{j=-\infty}^{\infty} \mathbf{b}_{t-j}' \mathbf{Y}_j,$$

and absolutely summable $k \times 1$ filters $\{\mathbf{b}_t\}$ and $\{\mathbf{c}_t\}$, so that the error of approximation, $\mathbf{Y}_t - \widehat{\mathbf{Y}}_t$ is small relative to mean squared error, where $\widehat{\mathbf{Y}}_t = \boldsymbol{\mu}_Y + \sum_{j=-\infty}^{\infty} \mathbf{c}_{t-j} Z_j$. If $\mathbf{b}(\omega)$ is the transform of $\mathbf{b}_t$, and $f_Z(\omega)$ the spectral density of Z_t, then the problem becomes one of finding a *complex* vector $\mathbf{b}(\omega)$, subject to the constraint that $\mathbf{b}^*(\omega)V\mathbf{b}(\omega) = 1$ such

that

$$f_Z(\omega) = \mathbf{b}^*(\omega) f_Y(\omega) \mathbf{b}(\omega)$$

is maximized. The solution, of course, is that $\mathbf{b}(\omega)$ is the eigenvector corresponding to the largest eigenvalue of $f_Y(\omega)$ in the metric of V, say $\lambda(\omega)$, and $f_Z(\omega) = \lambda(\omega)$ with $\mathbf{b}(\omega)$ so chosen in (7.1) [also, the transform of $\mathbf{c}_t$ is $\mathbf{b}^*(\omega)$].

In the language of scaling, we would state the same problem as: given a vector process $\mathbf{Y}_t$, find a complex vector $\mathbf{b}$ such that, at a given frequency ω, the time series $Z_t(\mathbf{b}) = \mathbf{b}^*\mathbf{Y}_t$ has the largest possible spectrum (subject to $\mathbf{b}^*V\mathbf{b} = 1$). The solution is to choose $\mathbf{b} = \mathbf{b}(\omega)$, that is, the eigenvector corresponding to the largest eigenvalue of $f_Y(\omega)$ in the metric of V. In this case the spectrum of $Z_t(\mathbf{b}(\omega))$, say $f_Z(\omega, \mathbf{b}(\omega))$, attains the largest possible value, $\lambda(\omega)$.

Hence, Brillinger's approach can be seen as a scaling problem with complex-valued scales. If we restrict $\mathbf{b}(\omega)$ to be real, then Sections 2 and 5 have $\mathbf{Y}_t$ being the multiple indicator process associated with a categorical-valued process. In Section 3, $\mathbf{b}(\omega)$ was also restricted to be real, but the $\mathbf{Y}_t$ process was generated by a space of transformations. In Section 4, we took the scalings $\mathbf{b}(\omega)$ to be real or complex depending on whether or not it was advantageous to do so. In the real case, the material in Sections 2 and 3 apply, and in the complex case, Brillinger (1975, chapter 9) applies.

The discussion of matching sequences, Section 6, is related to canonical analysis. Let $\mathbf{Y}_{1t}$ and $\mathbf{Y}_{2t}$ be as defined in Section 6. Brillinger (1975, chapter 10) discusses a time series extension of canonical correlation analysis that could be used in Section 6 as a special case. Briefly, consider real constants μ_i and $k_i \times 1$ linear filters $\{\mathbf{b}_{it}\}$ such that $\sum \|\mathbf{b}_{it}\| < \infty$, $i = 1, 2$. The real-valued univariate series

$$Z_{1t} = \mu_1 + \sum_{j=-\infty}^{\infty} \mathbf{b}'_{1, \mathbf{t}-j} \mathbf{Y}_{1j}$$

and

$$Z_{2t} = \mu_2 + \sum_{j=-\infty}^{\infty} \mathbf{b}'_{2, t-j} \mathbf{Y}_{2j}$$

having maximum squared-coherency $\rho_{12}^2(\omega)$ at each ω subject to $\mathbf{b}_i^*(\omega) f_{ii}(\omega) \mathbf{b}_i(\omega) = 1$, for $i = 1, 2$, where $\mathbf{b}_i(\omega)$ is the Fourier transform of $\{\mathbf{b}_{it}\}$, are given by finding the largest scalar $\lambda(\omega)$ such that, with $f_{21}(\omega) = f_{12}^*(\omega)$,

$$f_{22}(\omega)^{-1/2} f_{21}(\omega) f_{11}(\omega)^{-1} f_{12}(\omega) f_{22}(\omega)^{-1/2} \mathbf{v}(\omega)$$

$$= \lambda(\omega) \mathbf{v}(\omega).$$

The maximum squared-coherency achieved between Z_{1t} and Z_{2t} is $\lambda(\omega)$, and $\mathbf{b}_1(\omega)$ and $\mathbf{b}_2(\omega)$ are taken proportional to $f_{11}(\omega)^{-1} f_{12}(\omega) f_{22}(\omega)^{-1/2} \mathbf{v}(\omega)$ and $f_{22}(\omega)^{-1/2} \mathbf{v}(\omega)$, respectively.

An interpretation of the maximal squared coherency that is consistent with the notion of scaling and the spectral envelope established in Section 6 can be given. Specifically, we may regard $\lambda(\omega)$ as a "coherency envelope" in the following sense. Let $\mathbf{b}_i$ be complex-valued $k_i \times 1$ vectors, $i = 1, 2$, and consider the scaled complex-valued processes $Z_{1t}(\mathbf{b}_1) = \mathbf{b}_1^*\mathbf{Y}_{1t}$ and $Z_{2t}(\mathbf{b}_2) = \mathbf{b}_2^*\mathbf{Y}_{2t}$ having squared coherency $\rho_{12}^2(\omega, \mathbf{b}_1, \mathbf{b}_2)$. If we find $\mathbf{b}_1$ and $\mathbf{b}_2$ so that the squared coherency between $Z_{1t}(\mathbf{b}_1)$ and $Z_{2t}(\mathbf{b}_2)$ is maximized at a particular frequency $\omega = \omega_0$, then the maximum squared coherency at frequency ω_0 is $\lambda(\omega_0)$, and the complex-valued scalings for which the maximum squared coherency is achieved are proportional to $\mathbf{b}_1(\omega_0) = f_{11}(\omega_0)^{-1} f_{12}(\omega_0) f_{22}(\omega_0)^{-1/2} \mathbf{v}(\omega_0)$, and $\mathbf{b}_2(\omega_0) = f_{22}(\omega_0)^{-1/2} \mathbf{v}(\omega_0)$. Thus, for any nonzero complex vectors $\mathbf{b}_1$ and $\mathbf{b}_2$, $\rho_{12}^2(\omega, \mathbf{b}_1, \mathbf{b}_2) \leq \lambda(\omega)$, with equality when $\mathbf{b}_1$ is proportional to $\mathbf{b}_1(\omega)$ and $\mathbf{b}_2$ is proportional to $\mathbf{b}_2(\omega)$. The material in Section 6 focuses on the case where $\mathbf{b}_1$ and $\mathbf{b}_2$ are restricted to be real-valued, and in some cases, they are further restricted to be equal.

REFERENCES

ANTOGNINI, J. F., BUONOCORE, M. H., DISBROW, E. A. and CARSTENS, E. (1998). Isoflurane anesthesia blunts cerebral responses to noxious and innocuous stimuli: A functional MRI study. *Life Sciences* **61** PL349–PL354.

BINA, M. (1994). Periodicity of dinucleotides in nucleosomes derived from siraian virus 40 chromatin. *J. Molecular Biology* **235** 198–208.

BLOOMFIELD, P. (1976). *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

BREIMAN, L. and FRIEDMAN, J. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619.

BRILLINGER, D. R. (1975). *Time Series*: *Data Analysis and Theory*. (2nd ed., 1981.) Holden-Day, San Francisco.

BRILLINGER, D. R. (1980). Analysis of variance problems under time series models. *Handbook of Statistics* (P. R. Krishnaiah, ed.) **1** 237–278, North Holland, Amsterdam.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

BRODATZ, P. (1966). *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York.

BRUCE, A. and GAO, H-Y. (1996). *Applied Wavelet Analysis with S-PLUS*. Springer, New York.

CHATFIELD, C. (1989). *The Analysis of Time Series: An Introduction*, 4th ed. Chapman and Hall, London.

COOLEY, J. W. and TUKEY, J. W. (1965). An algorithm for the machine calculation of the complex Fourier series. *Math. Comp.* **19** 297–301.

CORNETTE, J. L., CEASE, K. B., MARGAHT, H., SPOUGE, J. L., BERZOFSKY, J. A. and DELISI, C. (1987). Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Molecular Biology* **195** 659–685.

DING, Z., GRANGER, C. W. J. and ENGLE, R. F. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance* **1** 83–106.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.

EISENBERG, D., WEISS, R. M. and TERWILLGER, T. C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Nat. Acad. Sci. U.S.A.* **81** 140–144.

FERRYANTO, SG. (1995). On estimation of the Walsh-Fourier spectral density of two dimensional strictly homogeneous random fields. *J. Nonparametr. Statist.* **5** 391–407.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

FULLER, W. A. (1995). *Introduction to Statistical Times Series*, 2nd ed. Wiley, New York.

GREENACRE, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press, London.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

IOSHIKHES, I., BOLSHOY, A. and TRIFONOV, E. N. (1992). Preferred positions of AA and TT dinucleotides in aligned nucleosomal DNA sequences. *J. Biomolecular Structure and Dynamics* **9** 1111–1117.

LI, C.-S. and TUREK, J. (1996). Content-based indexing of earth observing satellite image database with fuzzy attributes. In *Symposium on Electronic Imaging: Science and Technology—Storage and Retrieval for Image and Video Databases* IV **2670** 438–449. IS&T/SPIE.

LIU, F. and PICARD, R. W. (1996). Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence* **18** 722–733.

MACNEILL, I. (1977). A test of whether several time series share common periodicities. *Biometrika* **64** 495–508.

MCDOUGALL, A. J., STOFFER, D. S. and TYLER, D. E. (1997). Optimal transformations and the spectral envelope for real-valued time series. *J. Statist. Plann. Inference* **57** 195–214.

MCLACHLAN, A. D. and STEWART, M. (1976). The 14-fold periodicity in alpha-tropomyosin and the interaction with actin. *J. Molecular Biology* **103** 271–298.

MICHAILIDIS, G. and DE LEEUW, J. (1988). The Gifi system of descriptive multivariate analysis. *Statist. Sci.* **18** 307–336.

NIBLACK, W., BARBER, R., EQUITZ, W., FLICKNER, M., GLASMAN, E., PETKOVIC, D., YANKER, P. and FALOUTSOS, C. (1993). The QBIC project: Querying images by content using color, texture, and shape. *IBM RJ* 9203 (81511), February 1993.

NISHISATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. Univ. Toronto Press.

OGAWA, S. and LEE, T. M. (1990). Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magnetic Resonance in Medicine* **16** 9–18.

OGAWA, S., LEE, T. M., NAYAK, A. and GLYNN, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine* **14** 68–78.

PERCIVAL, D. B. and WALDEN, A. T. (1993). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge Univ. Press.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series* **1** and **2**. Academic Press, London.

RAO, K. R. and YIP, P. (1990). *Discrete Cosine Transform*. Academic Press, San Diego.

SATCHWELL, S. C., DREW, H. R. and TRAVERS, A. A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Molecular Biology* **191** 659–675.

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

STOFFER, D. S. (1987). Walsh-Fourier analysis of discrete-valued time series. *J. Time Ser. Anal.* **8** 449–467.

STOFFER, D. S. (1999). Detecting common signals in multiple time series using the spectral envelope. *J. Amer. Statist. Assoc.* **94** 1341–1356.

STOFFER, D. S., SCHER, M., RICHARDSON, G., DAY, N. and COBLE, P. (1988). A Walsh–Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *J. Amer. Statist. Assoc.* **83** 954–963.

STOFFER, D. S. and TYLER, D. E. (1998). Matching sequences: Cross-spectral analysis of categorical time series. *Biometrika* **85** 201–213.

STOFFER, D. S., TYLER, D. E. and MCDOUGALL, A. J. (1993). Spectral analysis for categorical time series: Scaling and the spectral envelope. *Biometrika* **80** 611–622.

STOFFER, D. S., TYLER, D. E., MCDOUGALL, A. J. and SCHACHTEL, G. (1993). Spectral analysis of DNA sequences (with discussion). *Bulletin of the International Statistical Institute* **I** 345–361; Discussion: **IV** 63–69 (1994).

TAVARÉ, S. and GIDDINGS, B. W. (1989). Some statistical aspects of the primary structure of nucleotide sequences. In *Mathematical Methods for DNA Sequences* (M. S. Waterman, ed.) 117–131. CRC Press, Boca Raton.

TIAO, G. C. and TSAY, R. S. (1994). Some advances in nonlinear and adaptive modeling in time series analysis. *J. Forecasting* **13** 109–131.

TIAO, G. C., TSAY, R. S. and WANG, T. (1993). Usefulness of linear transformations in multivariate time series analysis. *Empirical Economics* **18** 567–593.

VIARI, A., SOLDANO, H. and OLLIVIER, E. (1990). A scale-independent signal processing method for sequence analysis. *Computer Applications in the Biosciences* **6** 71–80.

WATERMAN, M. S. and VINGRON, M. (1994). Sequence comparison significance and Poisson approximation. *Statist. Sci.* **9** 367–381.

WENDT, D. A. (1999). Analysis of multidimensional categorical data. Ph.D. dissertation, Univ. Pittsburgh.