Discrimination and Classification of Nonstationary Time Series Using the SLEX Model

Hsiao-Yun HUANG, Hernando OMBAO, and David S. STOFFER

Statistical discrimination for nonstationary random processes is important in many applications. Our goal was to develop a discriminant scheme that can extract local features of the time series, is consistent, and is computationally efficient. Here, we propose a discriminant scheme based on the SLEX (smooth localized complex exponential) library. The SLEX library forms a collection of Fourier-type bases that are simultaneously orthogonal and localized in both time and frequency domains. Thus, the SLEX library has the ability to extract local spectral features of the time series. The first step in our procedure, which is the feature extraction step based on work by Saito, is to find a basis from the SLEX library that can best illuminate the difference between two or more classes of time series. In the next step, we construct a discriminant criterion that is related to the Kullback–Leibler divergence between the SLEX spectra of the different classes. The discrimination criterion is based on estimates of the SLEX spectra that are computed using the SLEX basis selected in the feature extraction step. We show that the discrimination method is consistent and demonstrate via finite sample simulation studies that our proposed method performs well. Finally, we apply our method to a seismic waves dataset with the primary purpose of classifying the origin of an unknown seismic recording as either an earthquake or an explosion.

KEY WORDS: Kullback–Leibler divergence; Likelihood ratio; Nonstationary random process; Seismic time series; SLEX library.

1. INTRODUCTION

The extension of classical pattern-recognition techniques to experimental time series is a problem of great practical interest. A series of observations indexed in time often produces a pattern that may form a basis for discriminating between different classes of events. As an example, Figure 1 shows regional (100–2,000 km) recordings of a typical Scandinavian earthquake, a mining explosion, and an event of unknown origin measured by stations in Scandinavia. The unknown event took place near the Russian nuclear test facility in Novaya Zemlya. The problem of discriminating between mining explosions and earthquakes is a reasonable proxy for the problem of discriminating between nuclear explosions and earthquakes. This latter problem is of critical importance for monitoring a comprehensive test-ban treaty. An extensive discussion of this problem can be found in Shumway and Stoffer (2000), chapter 5; the data are available on the website for the text, *http://www.stat.pitt.edu/stoffer/tsa.html* and its mirrors. Time series classification problems are not restricted to geophysical applications, but occur under many and varied circumstances in other fields. Traditionally, detecting a signal embedded in a noise series has been analyzed in the engineering literature by statistical pattern recognition techniques.

The historical approaches to the problem of discriminating among different classes of time series can be divided into two distinct categories. The optimality approach, as found in the engineering and statistics literature, makes specific Gaussian assumptions about the probability density functions of the separate groups and then develops solutions that satisfy well-defined minimum error criteria. Typically, in the time series case, we might assume the difference between classes is expressed through differences in the theoretical mean and covariance functions, and use likelihood methods to develop an optimal classification function. A second class of techniques, which might be described as a feature extraction approach, proceeds more heuristically by looking at quantities that tend to be good visual discriminators for well-separated populations and have some basis in physical theory or intuition. Less attention is paid to finding functions that are approximations to some well-defined optimality criterion.

When analyzing time series data, both time domain and frequency domain approaches are typically available; this is true in the case of discrimination and classification. For relatively short stationary series, a time domain approach that follows conventional multivariate discriminant analysis is viable; an example is given in Shumway and Stoffer (2000), example 5.11. For longer stationary time series, a frequency domain approach is computationally easier because it reduces the dimension of the problem. For nonstationary time series, the reduction in dimension is essential for computational efficiency.

There has been extensive research on the discrimination problem of stationary time series. For example, Shumway (1982) reviewed many different discriminant methods for time series, including time domain and frequency domain approaches. Kakizawa, Shumway, and Taniguchi (1998) proposed the method of discrimination for multivariate time series by using the Kullback–Leibler discrimination information and the Chernoff information measure. Pulli (1996) considered using the ratio of spectra to process the discriminant problem between earthquakes and explosions. These methods are developed for stationary time series. A discussion of the current state of the art for stationary series can be found in Shumway and Stoffer (2000), section 5.7.

In many practical problems, however, the time series are realizations of nonstationary random processes. For example, it is clear that the seismic waves displayed in Figure 1 have variance and spectrum that change over time. Various models of nonstationary random processes have been proposed in the literature. Priestley (1965) was the first to introduce the concept of a Cramér representation with time-varying transfer function. This idea was later refined by Dahlhaus (1997), who established

Figure 1. Example of Seismic Recordings.

an asymptotic framework for locally stationary processes. Recently, Ombao, Raz, von Sachs, and Guo (2002) introduced the SLEX (smooth localized complex exponentials) model of a nonstationary random process. The SLEX model uses the SLEX vectors (orthogonal and localized Fourier vectors) as stochastic building blocks in the Cramér representation.

Sakiyama and Taniguchi (2003) and Shumway (2003) discussed discrimination for Dahlhaus locally stationary time series. We approach the problem of discrimination and classification of nonstationary time series using the SLEX model. One very important consideration in using the SLEX library is that it employs computationally efficient algorithms. In particular, it uses the fast Fourier transform algorithm to compute the SLEX transform and the best basis algorithm (BBA) of Coifman and Wickerhauser (1992) to search for the best basis for discrimination between classes.

Our discriminant scheme contains two parts: a feature extraction part and a classification part. The feature extraction step consists of selecting a basis from the SLEX library that illuminates the most differences between the groups. We follow Saito (1994), who proposed a criterion that is based on the Kullback–Leibler divergence. This method is aimed at selecting the basis from any library of orthogonal bases that can best present the time series as clouds in space with maximal distance (distance is between the normalized time-varying spectrum). Because the SLEX functions are local, they are able to extract local spectral features of the time series. After selecting a basis using training datasets, we compute the SLEX periodogram of the time series that we need to classify. We propose a classification criterion which we show to be asymptotically consistent. The essential idea is that an observed time series is assigned to a class (population or group) Π_c if the Kullback–Leibler divergence between the estimated spectrum (SLEX periodogram computed from the data) and the spectrum of Π_c is smaller than that between the estimated spectrum and the spectra of any other class.

In the next section we present our procedure for selecting the best local discriminant basis from the SLEX library. In Sec-

tion 3, we discuss our discriminant criterion. Finally, in Section 4, we present some simulation studies and an analysis of seismic recordings.

2. SELECTING THE BEST LOCAL DISCRIMINANT BASIS FROM THE SLEX LIBRARY

The first step in our proposed method is to choose a basis from the SLEX library, which is a collection of bases; each basis consists of orthogonal and localized basis vectors. The SLEX vector is a time and frequency localized orthogonal generalization of the Fourier (complex exponential) basis vectors. It is ideal for analyzing nonstationary time series. Ombao, Raz, von Sachs, and Mallow (2001) used the SLEX to estimate the spectral density matrix of a bivariate nonstationary time series. Moreover, Ombao et al. (2002) introduced a model of a nonstationary random process that has a spectral representation in terms of the SLEX. The SLEX library is a rich collection of localized bases. Thus it is able to extract local spectral features of the time series and is well suited to the problem of discrimination and classification of nonstationary time series.

2.1 The SLEX Basis Vectors

Here, we give a brief overview of the SLEX transform. For details, we refer the reader to Ombao et al. (2001) for specific applications to time series and to Wickerhauser (1994) for a broader discussion on local trigonometric transforms.

Fourier basis vectors are perfectly localized in frequency and hence are ideal for representing stationary time series. However, they cannot adequately represent nonstationary time series, that is, the time series with spectra that change over time. In this article, we will use the SLEX basis vectors, which are simultaneously orthogonal and localized in time and frequency. They are constructed by applying a projection operator on the Fourier vectors. It turns out that the action of a projection operator on any periodic vector is identical to applying two specially constructed smooth windows to the Fourier vectors. A SLEX basis vector $\psi_{S,\omega}(t)$ that has support on the discrete time block
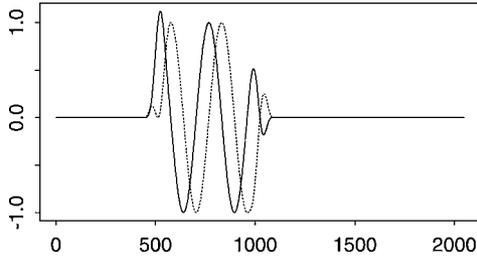
Figure 2. Real and Imaginary Parts of a SLEX Waveform With Support on Block Indexed by Time $t \in \{512, \ldots, 1{,}023\}$ With $\epsilon = 32$.

$S = \{\alpha_0 - \epsilon + 1, \ldots, \alpha_1 - \epsilon\}$ and oscillates at frequency ω has the form

$$\psi_{S,\omega}(t) = \Psi_{S,+}(t) \exp\left(i 2\pi \omega \frac{t}{|S|}\right)$$
$$+ \Psi_{S,-}(t) \exp\left(-i 2\pi \omega \frac{t}{|S|}\right), \quad (2.1)$$

where $\omega \in [-1/2, 1/2]$, $|S| = \alpha_1 - \alpha_0$, and ϵ is a small overlap between two consecutive time blocks (the size of ϵ is given in Sec. 2.2). The windows $\Psi_{S,+}(t)$ and $\Psi_{S,-}(t)$ are the two particularly constructed smooth windows, which take the form

$$\Psi_{S,+}(t) = r^2\left(\frac{t - \alpha_0}{\epsilon}\right) r^2\left(\frac{\alpha_1 - t}{\epsilon}\right), \quad (2.2)$$

$$\Psi_{S,-}(t) = r\left(\frac{t - \alpha_0}{\epsilon}\right) r\left(\frac{\alpha_0 - t}{\epsilon}\right)$$
$$- r\left(\frac{t - \alpha_1}{\epsilon}\right) r\left(\frac{\alpha_1 - t}{\epsilon}\right), \quad (2.3)$$

where $r(\cdot)$ is called a rising cut-off function. In our implementation, we use the sine rising cut-off function

$$r(u) = \sin\left(\frac{\pi}{4}(1 + u)\right), \quad \text{where } u \in [-1, 1]. \quad (2.4)$$

Other types of rising cut-off functions may be used. See Wickerhauser (1994) for details.

The SLEX basis vectors are defined at dyadic blocks that overlap. One important property of the SLEX vectors is that they remain orthogonal despite the overlap. Moreover, in our implementation, we use the fundamental frequencies $\omega_k = k/|S|$, $k = -|S|/2 + 1, \ldots, |S|/2$, where $|S|$ is the length of time block S. An example of a SLEX basis vector is given in Figure 2.

Figure 3. The SLEX Library Constructed With Level $J = 2$. The shaded blocks $S(1, 0) \cup S(2, 2) \cup S(2, 3)$ represent one basis (out of the five in this library).

2.2 The SLEX Library

The SLEX library is a collection of bases, each having orthogonal vectors with time support that is obtained by segmenting the time series, of length T, in a dyadic manner. The library is constructed by first specifying the finest resolution level J (smallest time block has length $T/2^J$). At resolution level j (where $j = 0, \ldots, J$), the time series is divided into 2^j overlapping blocks. The amount of overlap is set to $\epsilon = T/2^{J+1}$, which is the same for all levels j. We denote the block b on level j to be $S(j, b)$ and $M_j = T/2^j$. The SLEX vectors on block $S(j, b)$ are allowed to oscillate at different fundamental frequencies $\omega_k = k/M_j$ (where $k = -M_j/2 + 1, \ldots, M_j/2$).

To illustrate this further, consider the bottom of Figure 3, where the SLEX library is constructed by setting $J = 2$. In this example, the SLEX library consists of five orthogonal bases and we enumerate the support of these basis vectors: (1) $S(0, 0)$; (2) $S(1, 0) \cup S(1, 1)$; (3) $S(2, 0) \cup S(2, 1) \cup S(2, 2) \cup S(2, 3)$; (4) $S(1, 0) \cup S(2, 2) \cup S(2, 3)$; (5) $S(2, 0) \cup S(2, 1) \cup S(1, 1)$. Clearly, the SLEX basis vectors are allowed to have different lengths of support (or different time and frequency resolutions). Moreover, each basis captures local spectral features of the time series that are useful for discrimination and classification. In the latter part of this section, we discuss our criterion for selecting a basis from the SLEX library.

2.3 The SLEX Coefficients and Periodograms

The SLEX transform consists of the set of coefficients that correspond to all the SLEX vectors defined in the library. The SLEX coefficients on block $S = S(j, b)$ are defined by

$$\widehat{\theta}_{S,k} = \frac{1}{\sqrt{M_j}} \sum_t X_{t,T} \, \overline{\psi_{S,\omega_k}(t)}, \quad (2.5)$$

where the fundamental frequency is $\omega_k = k/M_j$ and $k = -M_j/2 + 1, \ldots, M_j/2$.

The SLEX periodogram, an analog of the Fourier periodogram, is defined to be

$$\widehat{\alpha}_{S,k} = |\widehat{\theta}_{S,k}|^2. \quad (2.6)$$

2.4 The Discriminant Criterion

We use the best local discriminant criterion that was described by Saito (1994) as a criterion for selecting the best basis. In the actual search over the SLEX library, we can employ either the algorithm used by Saito or the BBA of Coifman and Wickerhauser (1992). Both of these are computationally efficient algorithms, each of order $O(T)$.

Intuitively, a best basis for discrimination is that which gives the largest "disparity" between two or more different classes. One well-known measure of disparity is relative entropy (also known as cross-entropy, Kullback–Leibler divergence, or I divergence), which we now define. Let $\mathbf{p} = \{p_i\}_{i=1}^n$ and $\mathbf{q} = \{q_i\}_{i=1}^n$ be two nonnegative sequences that satisfy $\sum_i p_i = \sum_i q_i = 1$. The relative entropy between $\mathbf{p}$ and $\mathbf{q}$ is defined to be

$$I(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}.$$

By convention, we define $\log 0 = -\infty$, $\log(x/0) = +\infty$ for $x > 0$, and $0 \times (\pm\infty) = 0$. Relative entropy satisfies $I(\mathbf{p}, \mathbf{q}) \geq 0$ and the equality holds if and only if $\mathbf{p} \equiv \mathbf{q}$. This quantity is not a metric because it is not commutative and does not satisfy the triangle inequality. However, it measures the discrepancy of $\mathbf{p}$ from $\mathbf{q}$. Moreover, relative entropy, an additive measure in the sense that $I(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n I(p_i, q_i)$, enables us to use this divergence measure in the best basis algorithm. For $C \geq 2$ classes of time series, we define

$$I(\mathbf{p}^1, \ldots, \mathbf{p}^C) = \sum_{\ell=1}^{C-1} \sum_{k=\ell+1}^{C} I(\mathbf{p}^\ell, \mathbf{p}^k).$$

Now, let $\{\mathbf{x}_i^c\}_{i=1}^{N_c}$ be a set of N_c training signals belonging to class c, for $c = 1, \ldots, C$. Then the *time–frequency energy map* of class c, denoted by Γ_c, is a table of real values specified by the triplet (j, b, k) as

$$\Gamma_c(j, b, k) \equiv \sum_{i=1}^{N_c} \widehat{\alpha}_{(j,b),k}^i \Bigg/ \sum_{i=1}^{N_c} \|\mathbf{x}_i^c\|^2, \qquad (2.7)$$

where $\widehat{\alpha}_{(j,b),k}^i$ is the periodogram of the ith time series computed at level $j = 0, \ldots, J$ in block $b = 0, \ldots, 2^j - 1$, and frequency ω_k, where $k = -M_j/2 + 1, \ldots, M_j/2$. Note that the time–frequency energy map for a class c is in fact average "normalized" periodograms, that is, for any basis $\mathcal{B}_T$ in each class c, $\sum_{(j,b,k)\in\mathcal{B}_T} \Gamma_c(j, b, k) = 1$. The time–frequency energy map gives the "location" (in time and frequency) where the energy in a particular class is mostly concentrated. Thus, our procedure for classification is based on the time–frequency energy concentration of the signals.

After computing the time–frequency energy map of each class, we then calculate the divergence between the C signals at block $S(j, b)$ as

$$\Delta_{j,b} = \sum_{k=-M_j/2+1}^{M_j/2} I[\Gamma_1(j, b, k), \ldots, \Gamma_C(j, b, k)].$$

In this case, $\Delta_{j,b}$ gives a "local" measure of disparity between the time–frequency energy maps of the C classes in each time block $S(j, b)$. Our procedure selects the basis that consists of blocks $S(j, b)$, where, essentially, the distance between classes, $\Delta_{j,b}$, is maximized.

2.5 The Algorithm for Selecting the Best Basis

In this section, we give the specific steps for selecting the best basis. Saito (1994) developed the local discriminant basis algorithm (LDBA), which we apply to the SLEX library. Let $A_{j,b}$ represent the best basis (maximizer of the local discriminant criterion) and let $B_{j,b}$ denote the SLEX vectors at level j in block b. The algorithm for selecting the best basis is as follows:

Step 0. Specify the maximum depth of decomposition J.

Step 1. Construct time–frequency energy maps Γ_c for $c = 1, \ldots, C$.

Step 2. Set $A_{J,b} = B_{J,b}$. Determine the $A_{j,b}$ for $b = 0, \ldots, 2^j - 1$ and $j = J - 1, \ldots, 0$ by the following rule:

> If $\Delta_{j,b} \geq \Delta_{j+1,2b} + \Delta_{j+1,2b+1}$, then $A_{j,b} = B_{j,b}$, else $A_{j,b} = A_{j+1,2b} \oplus A_{j+1,2b+1}$ and set $\Delta_{j,b} = \Delta_{j+1,2b} + \Delta_{j+1,2b+1}$.

Step 3. Extract the blocks $S(j^*, b^*)$ that are defined by $A(j, b)$ in the previous step. The SLEX periodograms computed at the blocks defined by the best basis will be used to construct the classification rule.

The algorithm essentially compares a "parent" block against its "children" blocks; for example, the parent block $S(j, b)$ versus its children blocks $S(j + 1, 2b) \cup S(j + 1, 2b + 1)$. If the value of the local discriminant criterion at the parent block is smaller than the corresponding sum at the children blocks, then the children blocks separate the populations better than the parent block. Thus the children blocks are chosen in favor of their parent in the best basis B_T. If the value of the local discriminant criterion at the parent block is greater than or equal to the corresponding sum at the children blocks, then the children blocks do not separate the populations better than the parent block, and the parent block would be selected in favor of its children.

Finally, we point out the connection between LDBA and the BBA. The basis that maximizes the I divergence is equivalent to the basis that minimizes the negative I divergence. One may use the BBA to search for the best local discriminant basis in the SLEX library. Thus, our procedure is computationally efficient and can handle massive time series datasets.

2.6 The SLEX Model

We now describe the SLEX model for a nonstationary time series $X_{t,T}$ for $t = 0, \ldots, T - 1$. Let B_T be the best local discriminant basis selected from the SLEX library and let $\bigcup S_i$ be the blocks in B_T. Note that $\bigcup S_i$ is a particular dyadic segmentation of the time series. Define M_i to be the number of points on the block S_i. Let J_T be the highest time resolution level in B_T, that is, the smallest time block in B_T has length $T/2^{J_T}$. The frequencies defined on S_i are the grid frequencies $\omega_{k_i} = k_i/M_i$ for $k_i = -M_i/2 + 1, \ldots, M_i/2$. The spectral representation of $X_{t,T}$ is

$$X_{t,T} = \sum_{\bigcup S_i \sim B_T} \frac{1}{\sqrt{M_i}} \sum_{k=-M_i/2+1}^{M_i/2} \theta_{i,k,T} \psi_{i,k}(t) z_{i,k}, \qquad (2.8)$$

where $\theta_{i,k,T}$ is the transfer function on time block S_i and frequency k; $\psi_{i,k}$ is the SLEX basis vector oscillating at frequency k and having support at block S_i; and $z_{i,k}$ is an orthonormal random process with finite fourth moment.

2.6.1 The SLEX Spectrum. The SLEX spectrum is defined analogously to the spectrum of a stationary process. It is the square of the modulus of the time-varying transfer function. It is defined on rescaled time $[0, 1]$. Let u be in an interval $I \subset [0, 1]$ such that $[uT]$ is in some time block S_i on B_T. The SLEX spectrum is $f_T(u, \omega_k) = |\theta_{i,k,T}|^2 \iff [uT] \in S_i$. Note that for a fixed frequency ω_k, $f_T(u, \omega_k)$ is constant within each time subinterval (or block). This is because for each fixed T, the SLEX model gives an explicit partitioning of the time–frequency plane as determined by the blocks $\bigcup_i S_i$ in the basis B_T.

Ombao et al. (2002) developed asymptotic theory on the SLEX model. In this article, we discuss only the ideas that are essential in proving the consistency of our discrimination procedure. As $T \to \infty$, $f_T(u, \omega_k)$ approaches the "limiting spectrum" $f(u, \omega)$, which is independent of T. The limiting log-spectrum is defined subsequently. As T increases, the number of observations in each block also increases. However, the number of blocks in B_T should also be allowed to increase if we want to model a limiting spectrum that, as a function of time, has an infinite wavelet expansion. Thus, we need to allow J_T to increase as T increases, but at a rate that is slower than $K_T = \log_2(T)$. The infinite wavelet expansion in Ombao et al. (2002) was defined on $\log f(u, \omega)$ instead of $f(u, \omega)$ to guarantee positivity of the spectrum. Let the Haar wavelet basis of $L_2([0, 1])$ be $\{\varphi_{0,0}^*\} \cup \{\varphi_{\ell,m}\}_{\ell \geq 0, m \geq 0}$, where φ^* is the "father" Haar wavelet and φ is the "mother" wavelet. The infinite wavelet expansion of the log-limiting SLEX spectrum is

$$\log f(u, \omega) = \lim_{T \to \infty} \log f_T(u, \omega)$$

$$= \lim_{T \to \infty} \left(\beta_{-1,0}(\omega) \varphi_{0,0}^*(u) \right.$$

$$\left. + \sum_{\ell=0}^{J_T - 1} \sum_{m=0}^{2^\ell - 1} \beta_{\ell,m}(\omega) \varphi_{\ell,m}(u) \right), \quad (2.9)$$

where the coefficients are

$$\beta_{-1,0}(\omega) = \int_0^1 \log f(u, \omega) \varphi_{0,0}^*(u) \, du,$$

$$\beta_{\ell,m}(\omega) = \int_0^1 \log f(u, \omega) \varphi_{\ell,m}(u) \, du.$$

Ombao et al. (2002) demonstrated that for a given T, $\log f(u, \omega)$, and SLEX basis B_T with finest time resolution level J_T, $\log f_T(u, \omega)$ arises as a finite wavelet expansion of $\log f(u, \omega)$ truncated to a level J_T.

As a final remark, Ombao et al. (2002) showed that the SLEX model and the Dahlhaus (1997) model of a locally stationary process are asymptotically mean-squared equivalent. The equivalence implies nonstationary processes that have smoothly time-varying spectrum that can be modeled using the SLEX model. For example, one may model autoregressive moving average processes with parameters that vary smoothly over time by a SLEX model with a spectrum that varies with time but is piecewise constant within small time blocks.

2.6.2 The Likelihood of the SLEX Periodograms. We now derive the likelihood of the SLEX periodograms. When the orthonormal random increment processes, $z_{i,k}$, are complex Gaussian, then the SLEX coefficients, $\widehat{\theta}_{i,k}$, are independent complex Gaussian. Moreover, the SLEX periodograms $\widehat{\alpha}_{i,k} = |\widehat{\theta}_{i,k}|^2$ are distributed as $f_T(u_i, \omega_k) V_{i,k}$, where $V_{i,k}$ are independent over i and $k \geq 0$ and are distributed as $\chi_2^2/2$ when $k \neq 0, M_i/2$, and χ_1^2 when $k = 0, M_i/2$. Because the periodograms at each level j and block b are symmetric about the zero frequency ($k = 0$), we consider only the periodograms at nonnegative frequencies.

Denote $\widehat{\alpha} = \{\widehat{\alpha}_{i,k}\}$. Let $p(\widehat{\alpha})$ be the joint density of the periodograms under the SLEX process Π and let f_T be the spectrum. Then the log-likelihood of the periodograms is

$$\ell(f_T | \widehat{\alpha})$$

$$= -\sum_{\bigcup S_i \sim B_T} \sum_{k=0}^{M_i/2} \left\{ \log f_T(u_i, \omega_k) + \frac{\widehat{\alpha}_{i,k}}{f_T(u_i, \omega_k)} \right\}. \quad (2.10)$$

In the next section, we develop a classification rule that is based on the log-likelihood ratio.

3. THE CLASSIFICATION RULE

Let $\mathbf{x} = \{x_0, \ldots, x_{T-1}\}$ be a time series that we wish to classify as a realization of either Π_1 or Π_2 with time-varying spectra f_T^1 or f_T^2, respectively. We consider only the case where there are $C = 2$ classes of time series, although our method is applicable for $C > 2$. Our classification criterion is based on a frequency domain approach, that is, we use the spectrum as the basic signature for classifying time series. We extract the SLEX periodograms computed from the blocks in the basis selected in the feature extraction step. In this section, we derive our criterion using both the log-likelihood ratio and the Kullback–Leibler divergence. We then show that this criterion is consistent, that is, the probability of incorrectly classifying a time series goes to zero as the length of the time series $T \to \infty$.

3.1 Classification Rule as Log-Likelihood Ratio

The basic idea of our approach is as follows. We first extract the SLEX periodograms, which we denote as $\widehat{\alpha}(\mathbf{x})$, from the blocks of the best basis selected. We then form the likelihood under Π_1 and Π_2, denoted as $p_1[\widehat{\alpha}(\mathbf{x})]$ and $p_2[\widehat{\alpha}(\mathbf{x})]$, respectively. Our rule is to classify the time series $\mathbf{x}$ into Π_1 if $p_1[\widehat{\alpha}(\mathbf{x})] \geq p_2[\widehat{\alpha}(\mathbf{x})]$. In terms of the log-likelihood ratio, the discriminant statistic is

$$D_T(f_T^1, f_T^2; \mathbf{x}) = \frac{1}{T} \log \frac{p_1[\widehat{\alpha}(\mathbf{x})]}{p_2[\widehat{\alpha}(\mathbf{x})]} \quad (3.1)$$

and the classification rule is to assign $\mathbf{x}$ as a realization of Π_1 if $D \geq 0$. Otherwise, it is assigned to Π_2.

In this section, we derive the discrimination statistic (log-likelihood ratio) and then show that this criterion is consistent, that is, the probability of incorrectly classifying a time series goes to zero as the length of the time series $T \to \infty$.

3.1.1 The SLEX Likelihood Ratio. Denote $\widehat{\alpha} = \{\widehat{\alpha}_{i,k}\}$. Let $p_1(\widehat{\alpha})$ and $p_2(\widehat{\alpha})$ be the density under processes Π_1 and Π_2, respectively, and let f_T^1 and f_T^2 be the spectra under these two processes. Then the log-likelihoods under these two densities are, respectively,

$$\ell(f_T^1 | \widehat{\alpha})$$

$$= -\sum_{\bigcup S_i \sim B_T} \sum_{k=0}^{M_i/2} \left\{ \log f_T^1(u_i, \omega_k) + \frac{\widehat{\alpha}_{i,k}}{f_T^1(u_i, \omega_k)} \right\} \quad (3.2)$$

and

$$\ell(f_T^2 | \widehat{\alpha})$$

$$= -\sum_{\bigcup S_i \sim B_T} \sum_{k=0}^{M_i/2} \left\{ \log f_T^2(u_i, \omega_k) + \frac{\widehat{\alpha}_{i,k}}{f_T^2(u_i, \omega_k)} \right\}. \quad (3.3)$$

Our classification rule is based on the likelihood ratio, that is, we classify $\mathbf{x}$ into Π_1 if $\ell(f_T^1 | \widehat{\alpha}) \geq \ell(f_T^2 | \widehat{\alpha})$; otherwise, it is classified into Π_2. Following (3.1)–(3.3), the discriminant sta-

tistic is

$$D_T(f_T^1, f_T^2; \mathbf{x}) = \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \left\{ \log \frac{f_T^2(u_i, \omega_k)}{f_T^1(u_i, \omega_k)} \right.$$
$$\left. + \widehat{\alpha}_{i,k} \left[\frac{1}{f_T^2(u_i, \omega_k)} - \frac{1}{f_T^1(u_i, \omega_k)} \right] \right\}. \quad (3.4)$$

We classify $\mathbf{x}$ into Π_1 if $D_T(f_T^1, f_T^2; \mathbf{x}) \geq 0$.

Remark 1. First, we note that $D_T(f_T^1, f_T^2; \mathbf{x}) \approx \frac{2}{T} [\ell(f_T^1 | \widehat{\boldsymbol{\alpha}}) - \ell(f_T^2 | \widehat{\boldsymbol{\alpha}})]$. In the computation of the log-likelihood, we do not explicitly include the negative frequencies, because the periodograms are symmetric about $k = 0$.

Remark 2. To simplify the computation of the log-likelihood $\ell(f_T^1)$ and $\ell(f_T^2)$, we ignored the fact that the distribution of $\widehat{\alpha}_{i,k}/f_T(u_i, \omega_k)$ at $k = 0$ and $M_i/2$ is χ_1^2, whereas for $k = 1, \ldots, M_i/2 - 1$, the ratio is $\chi_2^2/2$. However, for M_i sufficiently large, we expect the difference to be practically negligible.

Remark 3. The derivation of $D_T(f_T^1, f_T^2; \mathbf{x})$ in the SLEX approach is inspired by the work of Sakiyama and Taniguchi (2003), who derived a discriminant statistic for the Dahlhaus (1997) model of a locally stationary random process. Denote the spectrum for the Dahlhaus model as $f_D(u, \omega)$. The likelihood is obtained by segmenting the time series into N time blocks, each having length M, and then computing the Fourier periodograms $P_M(b, \omega)$, at blocks $b = 1, \ldots, N$. Their classification criterion is

$$D_T^*(f^1, f^2; \mathbf{x})$$
$$= \frac{1}{N} \sum_{j=1}^{N} \int_{-1/2}^{1/2} \log \frac{f_D^2(u_j, \omega)}{f_D^1(u_j, \omega)}$$
$$+ P_M(u_j, \omega) \left[\frac{1}{f_D^2(u_j, \omega)} - \frac{1}{f_D^1(u_j, \omega)} \right] d\omega, \quad (3.5)$$

which is comparable to (3.4).

Remark 4. In practice, the true spectra, f_T^1 and f_T^2, are unknown. We replace these quantities in D_T by the SLEX periodograms that are averaged across time series replicates. Suppose that there are N_ℓ independent time series from class Π_ℓ for $\ell = 1, 2$. Denote the periodograms computed from the mth time series in the training dataset from class Π_ℓ to be $\widehat{\alpha}_{i,k}^{\ell,m}$. Then the estimate of $f_T^\ell(u_i, \omega_k)$ is

$$\widehat{f_T^\ell}(u_i, \omega_k) = \frac{1}{N_\ell} \sum_{m=1}^{N_\ell} \widehat{\alpha}_{i,k}^{\ell,m}, \qquad \ell = 1, 2. \quad (3.6)$$

Essentially, we perform averaging across subjects, which gives the similar desired effect of reducing the variance when smoothing over frequency. However, if N_1 is small, we might not achieve the desired reduction in the variance and thus we may need to do an extra smoothing step, which is over frequency. Hence, we may use the weighted form

$$\widetilde{f_T^\ell}(u_i, \omega_k) = \sum_{s=-M}^{M} W_s \widehat{f_T^\ell}(u_i, \omega_{k-s}),$$

where the $2M + 1$ is the span of the weights, $W_s \geq 0$, and $\sum_{s=-M}^{M} W_s = 1$. The estimate $\widetilde{f_T^2}(u_i, \omega_k)$ can be obtained in a similar manner if N_2 is small.

3.2 The Classification Rule as Kullback–Leibler Divergence

The discriminant statistic, D_T given in (3.4), can be derived using the Kullback–Leibler divergence. As before, let f_T^1 and f_T^2 represent the spectra of Π_1 and Π_2, respectively. Let the periodogram of the time series $\mathbf{x}$, to be classified into either Π_1 and Π_2, be denoted by $\widehat{\boldsymbol{\alpha}} = \{\widehat{\alpha}_{i,k}\}$. The Kullback–Leibler divergences between $\widehat{\boldsymbol{\alpha}}$ and f_T^1 and f_T^2 are

$$\text{KL}(\widehat{\boldsymbol{\alpha}}, f_T^1) = \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \left\{ -\log \frac{\widehat{\alpha}_{i,k}}{f_T^1(u_i, \omega_k)} \right.$$
$$\left. + \widehat{\alpha}_{i,k} \left[\frac{1}{f_T^1(u_i, \omega_k)} - \frac{1}{\widehat{\alpha}_{i,k}} \right] \right\}$$

and

$$\text{KL}(\widehat{\boldsymbol{\alpha}}, f_T^2) = \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \left\{ -\log \frac{\widehat{\alpha}_{i,k}}{f_T^2(u_i, \omega_k)} \right.$$
$$\left. + \widehat{\alpha}_{i,k} \left[\frac{1}{f_T^2(u_i, \omega_k)} - \frac{1}{\widehat{\alpha}_{i,k}} \right] \right\}.$$

We classify the series $\mathbf{x}$ into Π_1 if its periodogram $\widehat{\boldsymbol{\alpha}}$ is "closer" to f_T^1 than f_T^2; that is, if $\text{KL}(\widehat{\boldsymbol{\alpha}}, f_T^1) \leq \text{KL}(\widehat{\boldsymbol{\alpha}}, f_T^2)$ or if

$$D_T(f_T^1, f_T^2; \mathbf{x}) = \text{KL}(\widehat{\boldsymbol{\alpha}}, f_T^2) - \text{KL}(\widehat{\boldsymbol{\alpha}}, f_T^1) \geq 0.$$

This leads to the same classification rule discussed in Section 3.1.1 following (3.4), namely classify $\mathbf{x}$ into Π_1 if $D_T \geq 0$; otherwise, classify $\mathbf{x}$ into Π_2.

3.3 Consistency Theorem

We now state the results pertinent to the consistency of our classification scheme based on the SLEX model. Denote the probability of misclassifying a time series $\mathbf{x}$ to process Π_2, given that the true process is Π_1, to be $P_T(2|1) = P\{D_T(f_T^1, f_T^2; \mathbf{x}) < 0|\Pi_1\}$. Similarly, we denote $P_T(1|2) = P\{D_T(f_T^1, f_T^2; \mathbf{x}) \geq 0|\Pi_2\}$. Then, under appropriate conditions,

$$\lim_{T \to \infty} P_T(2|1) = 0, \quad (3.7)$$
$$\lim_{T \to \infty} P_T(1|2) = 0. \quad (3.8)$$

Moreover, (3.7) and (3.8) hold when f_T^1 and f_T^2 are replaced by the consistent estimates defined in (3.6) if, in addition, we have $N_\ell \to \infty$ for $\ell = 1, 2$. The conditions and proofs of these results are given in Appendix B.

4. SIMULATION RESULTS AND DATA ANALYSIS

4.1 Simulation Results

A number of numerical experiments were performed to test how well our proposed method was able to classify time series.

Table 1. Misclassification Rate for the First Simulation Study Where Π_1 Is Gaussian White Noise and Π_2 Is AR(1) With Parameter ϕ

ϕ	−.5	−.3	−.1	.1	.3	.5
Error rate	.00	.00	.02	.02	.00	.00

We discuss a few of simulation experiments here. In the first experiment we had two classes, where processes from Π_1 were Gaussian white noise and processes from Π_2 were stationary first-order autoregressive with parameter ϕ. The length of each time series was set to $T = 1,024$ and the training set for each group consisted of eight time series. Several simulation studies were performed by varying ϕ. Using the training datasets, we selected the best local discriminant basis and then set up the discriminant criterion. We then generated 50 time series datasets from each of Π_1 and Π_2. In Table 1 we report the error rate, which is the number of incorrect classifications out of a total of 100 trials. Note that there is a small misclassification rate when the value of ϕ is close to 0. For example, when $\phi = \pm.1$, there is some misclassification error because these AR(1) processes are close to white noise.

In the second experiment, processes from Π_1 were defined by

$$Y_t = \begin{cases} Y_t^{(1)} & \text{if } 1 \leq t \leq T/2 \\ Y_t^{(2)} & \text{if } T/2 + 1 \leq t \leq T, \end{cases} \quad (4.1)$$

where $Y_t^{(1)}$ is white noise and $Y_t^{(2)}$ is AR(1) with parameter $\phi = .1$. The processes from Π_2 were defined by

$$X_t = \begin{cases} X_t^{(1)} & \text{if } 1 \leq t \leq T/2 \\ X_t^{(2)} & \text{if } T/2 + 1 \leq t \leq T, \end{cases} \quad (4.2)$$

where $X_t^{(1)}$ is white noise and $X_t^{(2)}$ is AR(1) with parameter $\phi = -.1$. We generated $N = 8, 16$ training datasets from each group, Π_1 and Π_2, with $T = 512, 1,024, 2,048$. In each case we generated 50 time series datasets from Π_1 and another 50 from Π_2. The misclassification rates are displayed in Table 2, and again we note the favorable results.

In the third experiment, the processes from Π_1 and Π_2 were slowly time-varying AR(2) processes. The processes from Π_1 were generated as $Y_t = a_{t;\tau} Y_{t-1} - .81 Y_{t-2} + \epsilon_t$, $t = 1, \ldots, 1,024$, where $a_{t;\tau} = .8[1 - \tau \cos(\pi t/1,024)]$, $\tau = .5$, and ϵ_t were iid standard normal. We performed three subexperiments, where Π_2 was defined by the processes $Y_t = a_{t;\tau} Y_{t-1} - .81 Y_{t-2} + \epsilon_t$, $t = 1, \ldots, 1,024$, where τ takes on the three different values $\tau = .4, .3, .2$. The plots of $a_{t;\tau}$ for the various values of τ are shown in Figure 4. A simulated dataset from process Π_1 is shown in Figure 5 and that from Π_2 with $\tau = .3$ is shown in Figure 6.

Table 2. Misclassification Rate for the Second Simulation Study Where Π_1 Is (4.1) and Π_2 Is (4.2)

T	512	512	1,024	1,024	2,048	2,048
N	8	16	8	16	8	16
Error rate	.06	.05	.03	.00	.00	.00
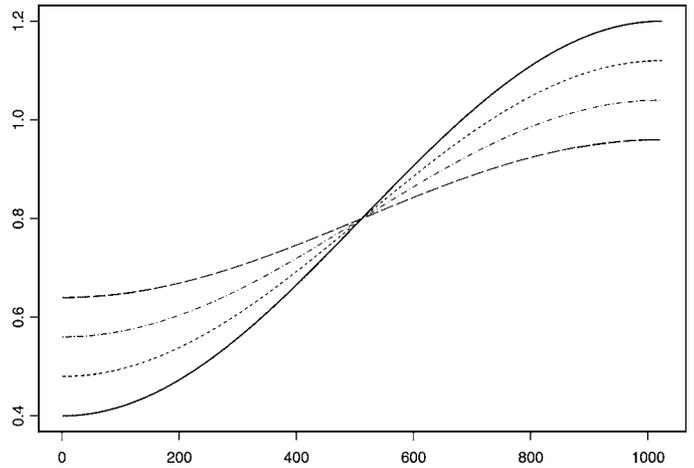
Figure 4. The Coefficient $a_{t;\tau}$ versus t for $\tau = .5$ (——), $\tau = .4$ (········), $\tau = .3$ (─·─·─), and $\tau = .2$ (─ ─ ─).
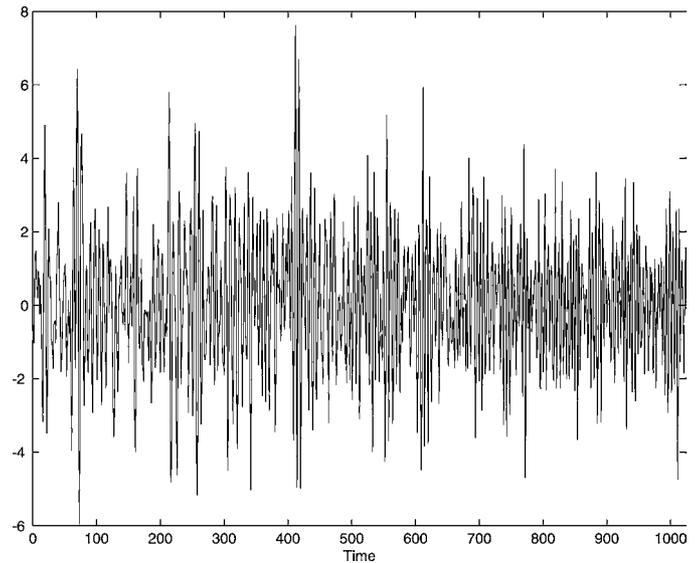
Figure 5. A Simulated Slowly Changing AR(2) Dataset From Π_1.
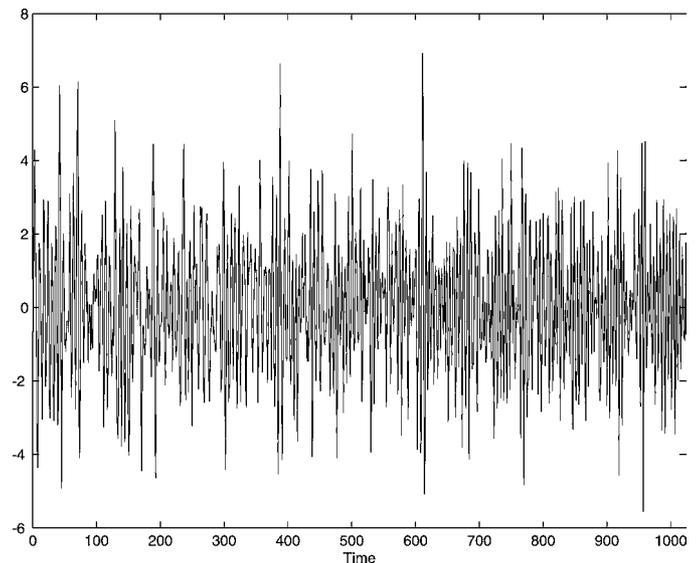
Figure 6. A Simulated Slowly Changing AR(2) Dataset From Π_2 With $\tau = .3$.

Table 3. The Simulation Results for the Slowly Varying AR(2) Experiment

τ	.4	.3	.2
Error rate	.08	.00	.00

We generated 10 datasets from each of Π_1 and Π_2 as the training data. From the training dataset, we obtained the best basis and the estimated spectra. Then we generated 10 new data for each category to evaluate the error rate. The simulation results are shown in Table 3. We can see that the error rates are zero when the $\tau = .3, .2$ in Π_2. There was a misclassification error for the case where the τ's are close for both processes, that is, $\tau = .4$ for Π_2 and $\tau = .5$ for Π_1.

4.2 Data Analysis

As discussed in the Introduction, discriminating between nuclear explosions and earthquakes is a problem of critical importance for monitoring a comprehensive test-ban treaty. We apply the proposed SLEX methodology to construct the discriminant rule for classifying a time series as either an explosion or an earthquake. Because the proliferation of nuclear explosions is monitored in regional distances (100–2,000 km) nowadays, the data on mining explosions can serve as a reasonable proxy. A dataset constructed by Blandford (1993) that comprises regional (100–2,000 km) recordings of several typical Scandinavian earthquakes and mining explosions measured by stations in Scandinavia are used in this study. A list of these events (eight earthquakes and eight explosions) and an extra event of uncertain origin that was located in the Novaya Zemlya region of Russia (called the NZ event) were given by Kakizawa et al. (1998). The problem was discussed in detail by Shumway and Stoffer (2000, chap. 5) and the data are availiable online from the website of the text (the URL is given in the Introduction). The earthquake and explosion seismic recordings are given in Figures 7 and 8, respectively. The SLEX spectral estimates for one earthquake and one explosion time series are given in Figures 9 and 10, respectively, whereas the SLEX spectral estimate of the NZ event is given in Figure 11.

To evaluate our method, we used the holdout procedure; that is, we removed one time series (at a time) to be used for classification and used all the remaining time series in the training dataset (excluding the unknown NZ event) to construct the classification rule. For each holdout time series, we performed the following steps in sequence: We selected a basis for discrimination, we extracted the SLEX periodograms computed at the blocks in this basis, we constructed the classification rule, and, finally, we assigned the holdout time series. We repeated the process for each of the holdout time series. In the
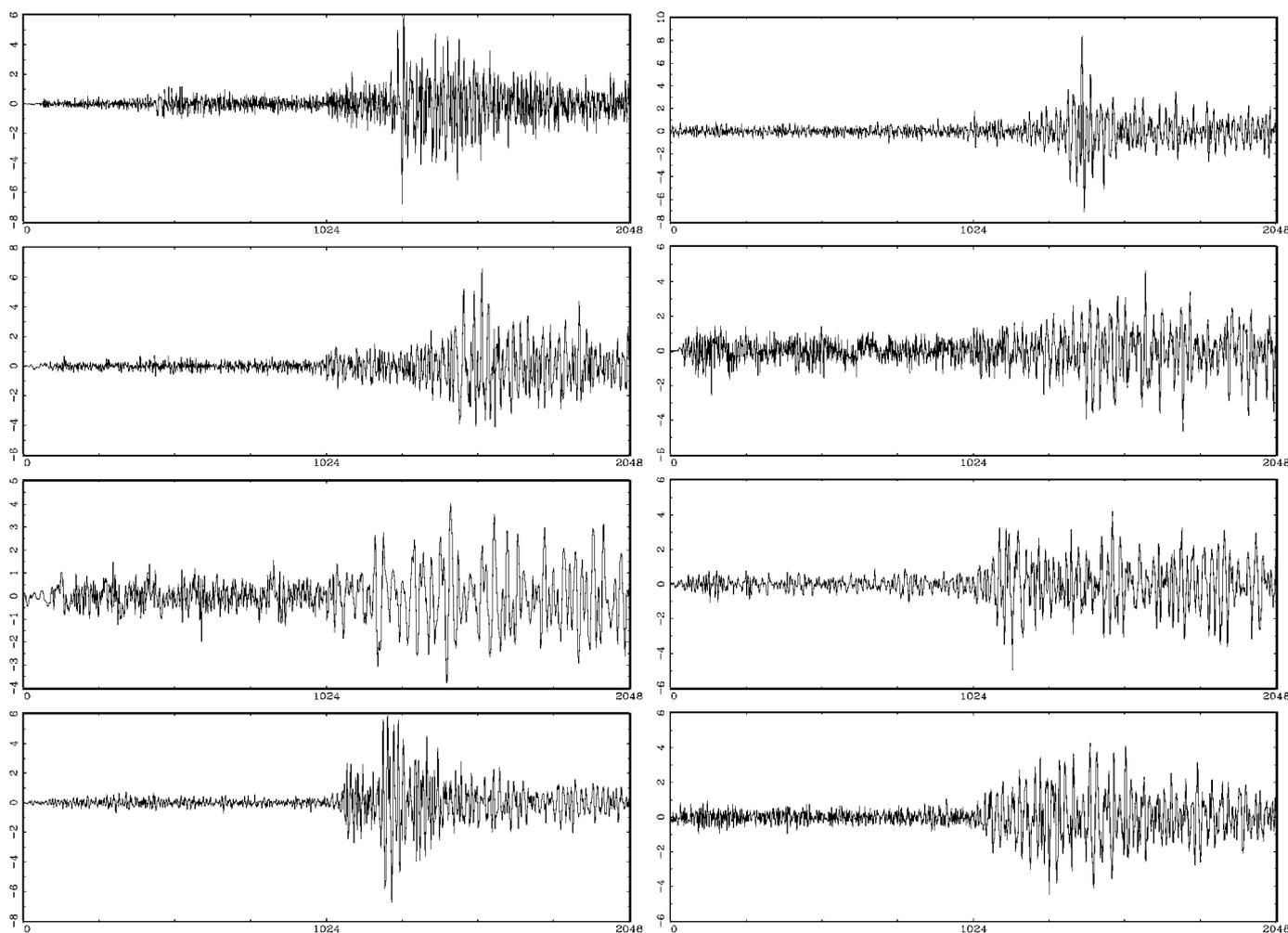
Figure 7. Seismic Recordings of Earthquake Origin.
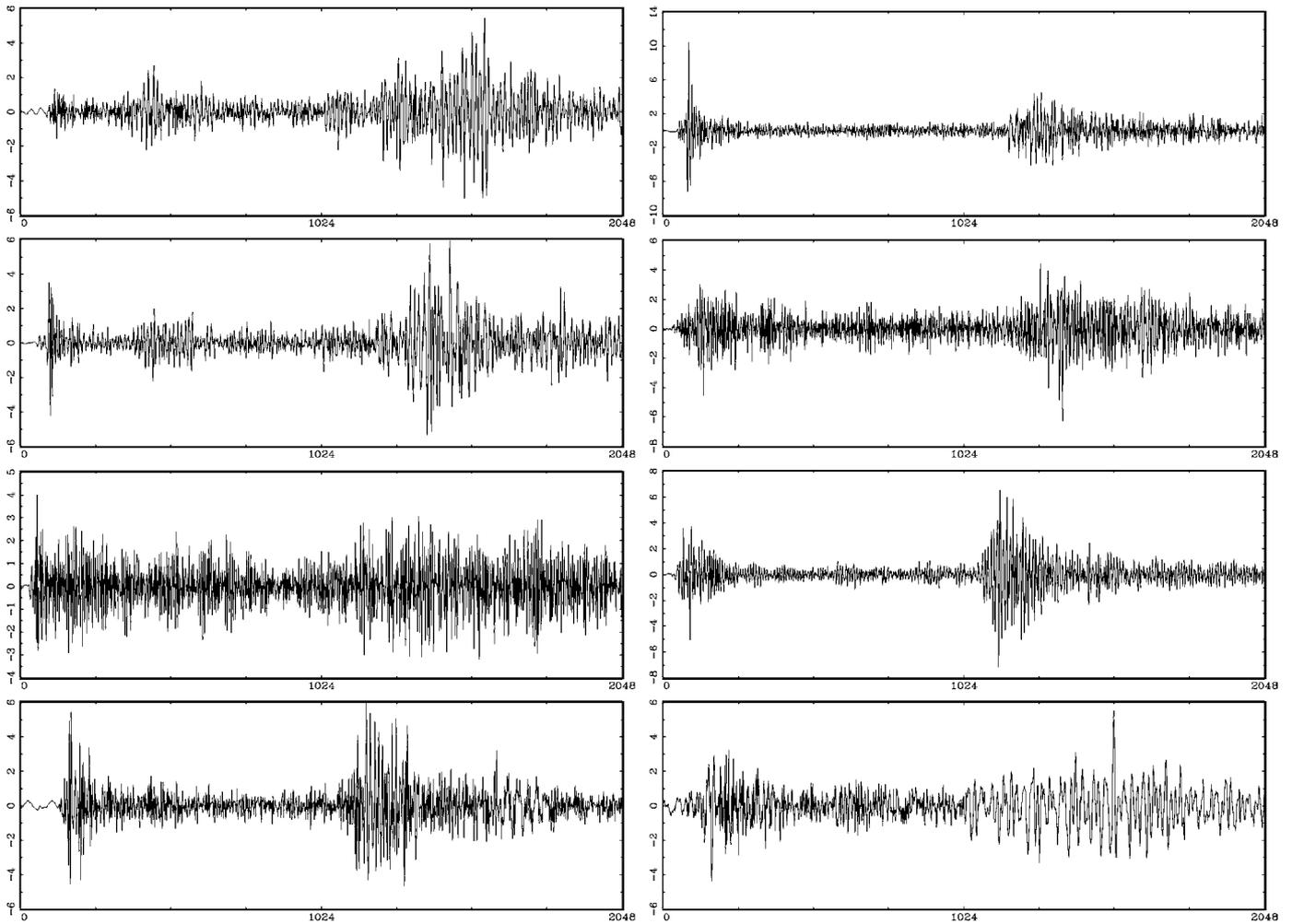
Figure 8. Seismic Recordings of Explosion Origin.

classification criterion, we estimated the spectra for each group by first averaging the SLEX periodograms across subjects and then smoothing the averaged SLEX periodograms across frequency. Denote the estimated spectra for the earthquake and the explosion classes to be, respectively, $\widetilde{f}_T^1$ and $\widetilde{f}_T^2$. The classification criterion was $D_T(\widetilde{f}_T^1, \widetilde{f}_T^2; \mathbf{x})$ and the testing data was classified as an earthquake event if $D_T(\widetilde{f}_T^1, \widetilde{f}_T^2; \mathbf{x}) \geq 0$ and as an explosion if $D_T(\widetilde{f}_T^1, \widetilde{f}_T^2; \mathbf{x}) < 0$. The results are reported in
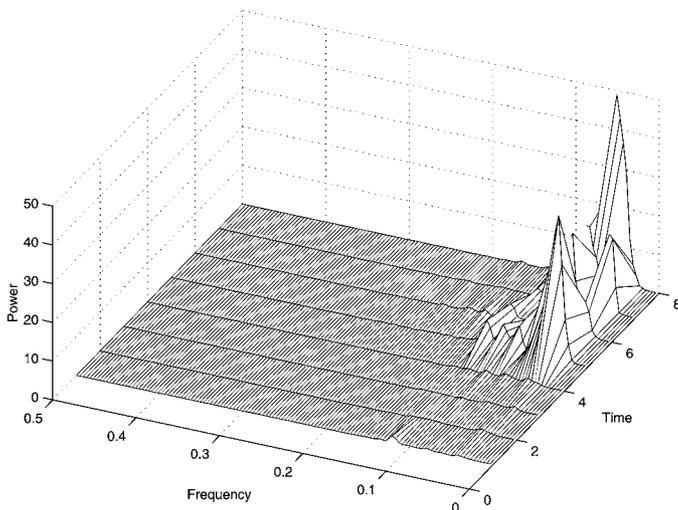
Figure 9. The SLEX Spectral Estimate of an Earthquake Time Series.
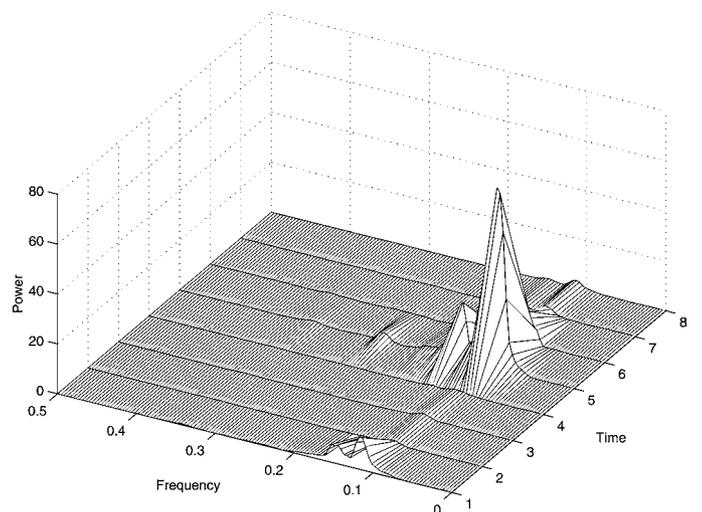
Figure 10. The SLEX Spectral Estimate of an Explosion Time Series.
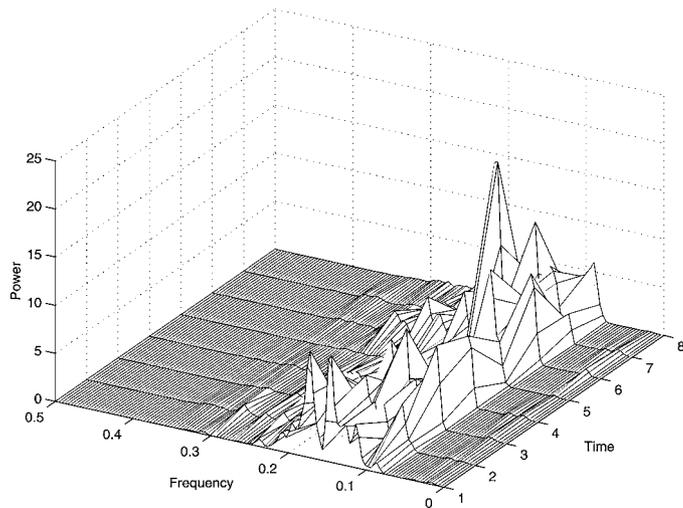
Figure 11. The SLEX Spectral Estimate of the NZ Event.

Table 4, where we denote Eqk to be the kth earthquake in the dataset and Exp$k$ to be the kth explosion in the dataset. The proposed method had a zero misclassification rate. Moreover, the unknown NZ event here is classified as an explosion, which agrees with the result in Kakizawa et al. (1998).

5. CONCLUSION

In this article, we proposed the SLEX method for discrimination and classification of nonstationary time series that is based on the SLEX transform. The SLEX transform is localized in both time and frequency domains; thus our method is able to extract local features of the data. Moreover, our method is computationally efficient and hence is able to handle large datasets. Our procedure computes the Kullback–Leibler distance of the time–frequency energy maps (normalized SLEX periodograms or the estimated normalized spectrum) of the classes and selects the basis from the SLEX library that can best discriminate between classes of nonstationary time series. Our classification rule, which is based on the Kullback–Leibler distance between the estimated spectra of the groups and the time series that needs to be classified, is shown to be consistent; that is, the probability of misclassification goes to zero as the length of the time series goes to infinity. Finite sample simulation studies and data analysis demonstrate that the method performs well in practice.

Table 4. Result of Holdout Procedure for Classifying Seismic Recordings

Holdout series	$D_T(\widetilde{f}_T^1, \widetilde{f}_T^2; \boldsymbol{x})$	Holdout series	$D_T(\widetilde{f}_T^1, \widetilde{f}_T^2; \boldsymbol{x})$
Eq1	.1526	Exp1	−.5096
Eq2	.6271	Exp2	−18.6391
Eq3	.8132	Exp3	−1.4676
Eq4	.0131	Exp4	−17.5373
Eq5	.5283	Exp5	−12.5065
Eq6	.7189	Exp6	−10.3111
Eq7	.6116	Exp7	−2.1453
Eq8	.6469	Exp8	−.1044
		NZ	−3.1552

APPENDIX A: ASSUMPTIONS ON THE LIMITING SPECTRUM OF THE SLEX MODEL

As mentioned in Section 2.6, the existence of the limiting SLEX spectrum depends on three assumptions, which we state now for completeness. Details can be found in Ombao et al. (2002).

Assumption 1. For $f(u, \omega)$ as a function of $\omega \in [-1/2, 1/2]$, uniformly in $u \in [0, 1]$, we assume a Hölder condition of order $\mu \in (0, 1]$ with constant $L > 0$:

$$|f(u, \omega) - f(u, \omega^*)| \leq L|\omega - \omega^*|^\mu.$$

Assumption 2. There exists a hierarchical collection $\mathcal{I}$ of dyadic subintervals of $[0, 1]$,

$$\mathcal{I} = \left\{\left[2^{-\ell}m, 2^{-\ell}(m+1)\right) : \ell = 0, 1, \ldots; \; m = 0, \ldots, 2^\ell - 1\right\},$$

and a subset of intervals, say $I_v = [u_v, u_{v+1}) \in \mathcal{I}$, such that $\bigcup_v I_v = [0, 1]$ and that $f(u, \omega)$, as a function of u, is Hölder of order $0 < s_v < 1$ on I_v for all $\omega \in [-1/2, 1/2]$. Moreover, the transition points u_v between the intervals I_{v-1} and I_v allow for a finite number of possible jumps of finite height.

Assumption 3. (a) Either the maximum depth J is fixed or, if $J = J_T$ is allowed to grow as a function of T (i.e., $J_T \to \infty$), then we must have $2^{J_T}/T \to 0$ as $T \to \infty$. Furthermore, $J_T \leq J_{2T}$.

(b) The length M_j of each segment S_j satisfies $M_{J_T}/M_j \leq 1$ for $j = 0, 1, \ldots, J_T$.

APPENDIX B: PROOF OF CONSISTENCY

In Section 3.3 we claimed that our discrimination rule is consistent in the sense of (3.7) and (3.8). We now prove these results under Assumptions 1–3. First, we state a theorem that relates the SLEX spectrum, f_T to the SLEX limiting spectrm, f (recall Remark 4 in Sec. 2.6).

Theorem 1 (Ombao et al. 2002). Under Assumptions 1–3, given the SLEX model as defined in Section 2.6, with SLEX spectrum f_T and SLEX limiting spectrum f, and a sequence of frequencies $\omega_{k,T} \to \omega$ as $T \to \infty$, we have the following situations:

1. Let $u \in I_v$ and let $2^{J_T} \sim T^{\mu/(\mu+s_v)}$. Then

$$|\log f_T(u, \omega_{k,T}) - \log f(u, \omega)| = O\left(T^{-s_v\mu/(\mu+s_v)}\right)$$

uniformly in $u \in I_v$.

2. Define $s := \inf_v s_v$ and let $2^{J_T} \sim T^{\mu/(\mu+s_v)}$. Then

$$\left|\int_0^1 \left[\log f_T(u, \omega_{k,T}) - \log f(u, \omega)\right]^2 du\right| = O\left(T^{-s_v\mu/(\mu+s_v)}\right).$$

To aid in the proof of the consistency of our method, we first define the following functions. Let

$$\phi_T(u, \omega) = 1/f_T^2(u, \omega) - 1/f_T^1(u, \omega)$$

and

$$\phi(u, \omega) = 1/f^2(u, \omega) - 1/f^1(u, \omega).$$

Define

$$H_T(\phi_T) = \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \phi_T(u_i, \omega_k)\widehat{\alpha}_{i,k},$$

$$H(\phi) = \int_0^1 \int_{-1/2}^{1/2} \phi(u, \omega) f^*(u, \omega) \, d\omega \, du,$$

where $\widehat{\alpha}_{i,k}$ is the SLEX periodogram of $\mathbf{x}$, $\bigcup_i S_i$ represents the blocks corresponding to the best basis B_T, and f^* is the SLEX limiting spectrum that corresponds, generically, to class Π_* (i.e., $\mathbf{x} \in \Pi_*$). Similarly, f_T^* represents the SLEX spectrum that corresponds to Π_*. The

proofs of the following lemmas are brief to save space; explicit details can be obtained from the authors.

Lemma 1. Under the established notation and conditions,

$$\mathrm{E}_* H_T(\phi_T) = H(\phi) + O\left(\left[\frac{M}{T}\right]^\mu\right) + O\left(\left[\frac{2^{J_T}}{T}\right]^\mu\right) + O(b_T^s),$$

where E_* denotes expection with respect to (wrt) Π_*, $M := \sup_i M_i$, $b_T := M_{J_T}/T$, and $s := \inf_v s_v$.

Proof.

$$\mathrm{E}_* H_T(\phi_T) = \sum_{\bigcup_i S_i \sim B_T} \frac{M_i}{T} \frac{1}{M_i} \sum_{k=-M_i/2+1}^{M_i/2} \phi_T(u_i, \omega_k) f_T^*(u_i, \omega_k)$$

$$= \sum_{\bigcup_i S_i \sim B_T} \frac{M_i}{T} \frac{1}{M_i} \sum_{k=-M_i/2+1}^{M_i/2} [\phi(u_i, \omega_k) + O(b_T^s)]$$

$$\times [f^*(u_i, \omega_k) + O(b_T^s)]$$

$$= \sum_{\bigcup_i S_i \sim B_T} \frac{M_i}{T} \left[\int_{-1/2}^{1/2} \phi(u_i, \omega) f^*(u_i, \omega) \, d\omega \right.$$

$$\left. + O\left(\left[\frac{2^{J_T}}{T}\right]^\mu\right) + O(b_T^s) \right]$$

$$= \int_0^1 \int_{-1/2}^{1/2} \phi(u, \omega) f^*(u, \omega) \, d\omega \, du$$

$$+ O\left(\left[\frac{M}{T}\right]^\mu\right) + O\left(\left[\frac{2^{J_T}}{T}\right]^\mu\right) + O(b_T^s).$$

Lemma 2. Under the established notation and conditions,

$$\mathrm{Var}_*\{H_T(\phi_T)\} = O(T^{-1}) + O\left(\frac{b_T^s}{T}\right) + O\left(\frac{2^{J_T}}{T}\right),$$

where Var_* denotes variance wrt Π_*.

Proof. Let $P_i = \sum_{k=-M_i/2+1}^{M_i/2} \phi_T(u_i, \omega_k)\widehat{\alpha}_{i,k}$. Then

$$\mathrm{Var}_*\{H_T(\phi_T)\} = \mathrm{Var}_*\left\{ \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \phi_T(u_i, \omega_k)\widehat{\alpha}_{i,k} \right\}$$

$$= \frac{1}{T^2} \sum_i \mathrm{Var}_*(P_i) + \frac{1}{T^2} \sum_{i \neq j} \mathrm{Cov}_*(P_i, P_j)$$

in obvious notation. Using arguments similar to the proof of Lemma 1 and the result of Lemma 1, we can show

$$\frac{1}{T^2} \sum_i \mathrm{Var}_*\{P_i\} = O(T^{-1}) + O\left(\frac{b_T^s}{T}\right)$$

and

$$\frac{1}{T^2} \sum_{i \neq j} \mathrm{Cov}_*(P_i, P_j) = O\left(\frac{2^{J_T}}{T}\right).$$

Lemma 3. Let $D_T(f_T^1, f_T^2; \mathbf{x})$ denote the discriminant statistic defined in (3.4), where f_T^1 and f_T^2 are the SLEX spectra in classes Π_1 and Π_2, respectively. Then under the established notation and conditions, as $T \to \infty$,

$$D_T(f_T^1, f_T^2; \mathbf{x}) - \mathrm{E}\{D_T(f_T^1, f_T^2; \mathbf{x})|\Pi_i\} \xrightarrow{P} 0 \quad \text{for } \mathbf{x} \in \Pi_i, \ i = 1, 2.$$

Proof. Without loss of generality, we prove the result for $\mathbf{x} \in \Pi_1$. Set

$$R_T = \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \log \frac{f_T^2(u_i, \omega_k)}{f_T^1(u_i, \omega_k)}.$$

Then

$$D_T(f_T^1, f_T^2; \mathbf{x}) = R_T + H_T(\phi_T)$$

and

$$\mathrm{E}\{D_T(f_T^1, f_T^2; \mathbf{x})|\Pi_1\} = R_T + \mathrm{E}\{H_T(\phi_T)|\Pi_1\}.$$

Hence,

$$\mathrm{E}\{[D_T(f_T^1, f_T^2; \mathbf{x}) - \mathrm{E}\{D_T(f_T^1, f_T^2; \mathbf{x})\}]^2|\Pi_1\}$$
$$= \mathrm{Var}\{H_T(\phi_T)|\Pi_1\} \to 0$$

as $T \to \infty$ by Lemma 2.

We are now ready to establish the main result on the consistency of our discrimination rule.

Theorem 2. Under the established notation and conditions,

$$\lim_{T \to \infty} P_T(2|1) = \lim_{T \to \infty} P\{D_T(f_T^1, f_T^2; \mathbf{x}) < 0|\Pi_1\} = 0,$$

$$\lim_{T \to \infty} P_T(1|2) = \lim_{T \to \infty} P\{D_T(f_T^1, f_T^2; \mathbf{x}) \geq 0|\Pi_2\} = 0.$$

Proof. We prove the first result; the second result follows in a similar manner. Using Lemma 1, we have

$$\mathrm{E}[D_T(f_T^1, f_T^2; \mathbf{x})|\Pi_1]$$

$$= \frac{1}{T} \sum_{\bigcup_i S_i \sim B_T} \sum_{k=-M_i/2+1}^{M_i/2} \left[\log \frac{f_T^2(u_i, \omega_k)}{f_T^1(u_i, \omega_k)} + \frac{f_T^1(u_i, \omega_k)}{f_T^2(u_i, \omega_k)} - 1 \right]$$

$$= \int_0^1 \int_{-1/2}^{1/2} \left[\log \frac{f^2(u, \omega)}{f^1(u, \omega)} + \frac{f^1(u, \omega)}{f^2(u, \omega)} - 1 \right] d\omega \, du$$

$$+ O\left(\left[\frac{M}{T}\right]^\mu\right) + O\left(\left[\frac{2^{J_T}}{T}\right]^\mu\right) + O\left(\left[\frac{M_T}{T}\right]^s\right).$$

However,

$$\log \frac{f^2(u, \omega)}{f^1(u, \omega)} + \frac{f^1(u, \omega)}{f^2(u, \omega)} - 1 \geq 0,$$

so that

$$\mathrm{E}[D_T(f_T^1, f_T^2; \mathbf{x})|\Pi_1] \to C \geq 0 \qquad (\mathrm{A.1})$$

as $T \to \infty$. In light of (A.1) and Lemma 3, the result follows immediately.

Corollary. Let $\widehat{f_T^\ell}(u_i, \omega_k)$ for $\ell = 1, 2$ be the estimates given in (3.6) and let $N = \min\{N_1, N_2\}$. Then, under the established notation and conditions,

$$\lim_{T \to \infty} \lim_{N \to \infty} P\{D_T(\widehat{f_T^1}, \widehat{f_T^2}; \mathbf{x}) < 0|\Pi_1\} = 0,$$

$$\lim_{T \to \infty} \lim_{N \to \infty} P\{D_T(\widehat{f_T^1}, \widehat{f_T^2}; \mathbf{x}) \geq 0|\Pi_2\} = 0.$$

Proof. By the strong law of large numbers, $D_T(\widehat{f_T^1}, \widehat{f_T^2}; \mathbf{x}) \xrightarrow{\text{as}} D_T(f_T^1, f_T^2; \mathbf{x})$ as $N \to \infty$, and the result follows immediately from Theorem 2.

REFERENCES

Blandford, R. R. (1993), "Discrimination of Earthquakes and Explosions," Report AFTAC-TR-93-044 HQ, Air Force Technical Applications Center, Patrick Air Force Base, FL.

Coifman, R., and Wickerhauser, M. (1992), "Entropy Based Algorithms for Best Basis Selection," *IEEE Transactions on Information Theory*, 32, 712–718.

Dahlhaus, R. (1997), "Fitting Time Series Models to Nonstationary Processes," *The Annals of Statistics*, 25, 1–37.

Kakizawa, Y., Shumway, R. H., and Taniguchi, M. (1998), "Discrimination and Clustering for Multivariate Time Series," *Journal of the American Statistical Association*, 93, 328–340.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002), "The SLEX Model of Nonstationary Processes," *Annals of the Institute of Statistical Mathematics*, 32, 201–225.

Ombao, H., Raz, J., von Sachs, R., and Malow, B. (2001), "Automatic Statistical Analysis of Bivariate Nonstationary Time Series," *Journal of the American Statistical Association*, 96, 543–560.

Priestley, M. (1965), "Evolutionary Spectra and Nonstationary Processes," *Journal of the Royal Statistical Society*, Ser. B, 28, 228–240.

Pulli, J. J. (1996), "Extracting and Processing Signal Parameters for Regional Seismic Event Identification," in *Monitoring a Comprehensive Test Ban Treaty*, eds. E. S. Husebye and A. M. Dainty, Dordrecht: Kluwer, pp. 791–803.

Saito, N. (1994), "Local Feature Extraction and Its Applications," Ph.D. dissertation, Yale University, Dept. of Mathematics.

Sakiyama, K., and Taniguchi, M. (2003), "Discrimination for Locally Stationary Random Processes," *Journal of Multivariate Analysis*, in press.

Shumway, R. H. (1982), "Discriminant Analysis for Time Series," in *Classification, Pattern Recognition and Reduction of Dimensionality*, *Handbook of Statistics*, Vol. 2, eds. P. R. Krishnaiah and L. N. Kanal, Amsterdam: North-Holland, pp. 1–46.

———— (2003), "Time–Frequency Clustering and Discriminant Analysis," *Statistics and Probability Letters*, 63, 307–314.

Shumway, R. H., and Stoffer, D. S. (2000), *Time Series Analysis and Its Applications*, New York: Springer-Verlag.

Wickerhauser, M. (1994), *Adapted Wavelet Analysis From Theory to Software*, New York: IEEE Press.