Discrimination of Nonstationary Time Series using the SLEX Model

by

Hsiao-Yun Huang

B.S. Fu-Jen University, Taiwan, 1994
M.S. National Tai-Chung Teachers College, Taiwan, 1999

Submitted to the Graduate Faculty
of the Arts and Sciences
in partial fulfillment of
the requirement for the degree of
Doctor
of
Philosophy

University of Pittsburgh
2003

ABSTRACT

Signature ___________________________

Discrimination of Nonstationary Time Series using the SLEX Model

Hsiao-Yun Huang , Ph.D.

University of Pittsburgh

Statistical discrimination for nonstationary random processes have developed into a widely practiced field with various applications. In some applications, such as signal processing and geophysical data analysis, the generated processes are usually long series. In such cases, a discriminant scheme with computational efficiency and optimal property is of great interest.

In this dissertation, a discriminant scheme for nonstationary time series based on the SLEX model (Ombao, Raz, von Sachs and Guo, 2002) is presented. The SLEX model is based on the Smooth Localized complex EXponential (SLEX)[Wickerhauser, 1994] basis functions. SLEX basis functions generalize directly to a library of SLEX basis vectors that are complex-valued, orthonormal, and simultaneously localized in time and frequency domains (Wickerhauser, 1994). Thus, it provides an explicit

segmentation of the time-frequency plane and hence is able to represent discrete random processes whose spectral properties change with time. Since the SLEX basis functions can also be considered a generalization of the tapered Fourier vectors, the calculation from SLEX basis functions to a library of SLEX basis vectors (called the SLEX transform) can use the Fast Fourier Transform. That is, the SLEX transform has computational efficiency. Moreover, the SLEX model, with a structure for asymptotic theory, allows the derivation of the optimal properties of the discriminant statistic in this dissertation.

A statistical time series classification scheme can be considered a formulation with two steps: extracting features from the data and developing a decision function. For feature extraction, a fast algorithm associated with the SLEX model is formed to extract the features. For developing a decision function, an optimal discriminant statistic based on the Kullback-Leibler divergence (Kullback and Leibler, 1951) of the SLEX model is proposed. The entire scheme will be organized as an algorithm. That is, a computationally efficient and statistically optimal discriminant scheme for nonstationary time series is proposed in this dissertation.

TABLE OF CONTENTS

Page

ABSTRACT . . . . **ii**

LIST OF FIGURES . . . . **iv**

LIST OF TABLES . . . . **vi**

1.0 Introduction **1**

2.0 Literature Review **6**

2.1 Kullback-Leibler Divergence 6

2.2 Frequency Domain Discriminant Methods for Stationary Time Series 7

2.3 The Dahlhaus Model 9

2.4 Discriminant Analysis for the Dahlhaus Locally Stationary Processes 13

2.5 The SLEX Model 15

2.5.1 The SLEX Basis Vectors 15

2.5.2 The SLEX Library 17

2.5.3 The Best Basis 18

2.5.4 The SLEX Model of a Nonstationary Random Process 19

2.5.5 Asymptotic Framework 21

2.6 Feature Extraction Method 23

3.0 Discrimination for Nonstationary Time Series Using the SLEX Model **26**

3.1 Feature Extraction 27

3.2 Discriminant Statistic 28

3.3 Asymptotic Results on Consistency 33

3.4 The Algorithm of the Discriminant Scheme 40

4.0 Simulation Results and Data Analysis **42**

4.1 Simulation on Stationary Processes 43

4.2 Simulation on Piecewise Stationary Processes 45

4.3 Simulation on Slowly Varying AR(2) Processes 46

4.4 Data Analysis 49

5.0 Conclusion and Future Work **54**

APPENDIX A The Main Matlab Progrom for Section 4.4 . . . . . **56**

APPENDIX B The Matlab Subroutine for Calculating the Entropy 60

APPENDIX C The Matlab Subroutine for Calculating the Cost . **63**

APPENDIX D The Matlab Subroutine for Getting the Best Basis 65

APPENDIX E The Matlab Subroutine for Getting the SLEX Periodogram **68**

APPENDIX F The Matlab Subroutine for Getting the SLEX Periodogram of the Target Data **71**

APPENDIX G The Matlab Subroutine for Calculating the Discriminant Statistic **73**

APPENDIX H The Matlab Subroutine for Calculating the SLEX Transform **75**

APPENDIX I The Matlab Subroutine for Smoothing the SLEX Transform **80**

BIBLIOGRAPHY **84**

LIST OF FIGURES

Figure No. Page

1.1 The top panel is a plot of an earthquake. The bottom panel is a plot of an explosion. 3

2.1 The windows used in the construction of the SLEX function 17

2.2 The SLEX library where $J = 2$ 18

4.1 White Noise 43

4.2 AR(1) with coefficient 0.1 44

4.3 White Noise and AR(1) with coefficient 0.1 46

4.4 A SLEX periodogram of the White noise and AR(1) with coefficient 0.1 46

4.5 White Noise and AR(1) with coefficient 0.3 46

4.6 A SLEX periodogram of the White noise and AR(1) with coefficient 0.3 47

4.7 The coefficient $a(t)$ when Δ=0.5,0.4,0.3, and 0.2 48

4.8 A simulated slowing changing AR(2) dataset from the first category . 48

4.9 A simulated slowing changing AR(2) dataset from the second category 49

4.10 A Time-Varying SLEX periodogram of the first category 49

4.11 A Time-Varying SLEX periodogram of the second category 49

4.12 The top panel is a typical earthquake. The middle panel is a typical explosion. The bottom panel is the unknown NZ event 51

4.13 The SLEX periodogram of the earthquake 6 51

4.14 The SLEX periodogram of the explosion 6 52

4.15 The SLEX periodogram of the NZ event 52

LIST OF TABLES

Table No. Page

4.1 Simulation results for stationary data 44

4.2 The Simulation Results for Piecewise Stationary Data 47

4.3 The simulation results for slowly varying AR(2) 50

4.4 The results for real data analysis 53

Chapter 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. Time series classification problems occur under many varied circumstances in many fields. For example, the detecting of a signal embedded in a noise series has been analyzed in the engineering literature by statistical pattern recognition techniques (see Miao and Clements, 2002).

The goal of this study is to construct a scheme for the discriminant problem involving nonstationary time series. An ideal scheme should be thorough in the sense that it is theoretically well established and practically computationally efficient. I approach this problem using the SLEX model.

There have been extensive research in the discriminant analysis of stationary time series. For example, Shumway (1982) reviewed many different discriminant methods for time series including time domain and frequency domain approaches. Kaizawa, Shumway and Taniguchi (1998) proposed a method 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 approach the discriminant problem between earthquakes and explosions. These methods assume stationarity.

In many practical problems, however, the series are nonstationary random processes. A data set constructed by Blandford (1993) which are regional (100-2000 km) recordings of several typical Scandinavian earthquakes and mining explosions measured by stations in Scandinavia is an example. A plot of an earthquake and an explosion in the data set is shown in Figure 1.1. We can see that the variance of both earthquake and explosion varies over time. That is, the earthquake and the explosion are nonstationary processes.

Priestley (1965) introduced the concept of a Cramér representation with time-varying transfer function. Dahlhaus (1997) refined the ideas in Priestley and defined a model for the nonstationary time series with an evolutionary spectrum as a locally stationary time series. He also proposed estimation methods and an asymptotic framework associated with the model. With the well established theoretical structure, the Dahlhuas model is currently a tool used in many researches of nonstationary time series. For example, by ignoring some smoothness assumptions in the Dahlhaus model, Adak (1998) introduced a class of piecewise locally stationary processes and considered the applications to speech signals and earthquake data.

For practical applications, we need a model that is theoretically sound and computationally practical. With this consideration, the SLEX model (Ombao, Raz, von Sachs and Guo, 2002), which was built parallel to the Dahlhaus model, is an ideal choice for dealing with discrete nonstationary time series. The SLEX model has a Cramér-like representation in terms of the SLEX (Smooth Localized Complex EXponential) functions (Wickerhauser, 1994). In the theoretical aspect, the SLEX functions are orthonormal, localized in time and frequency, and a generalization of the tapered Fourier functions. Thus, the SLEX model is appropriate for random processes whose statistical properties vary over time. Moreover, under the SLEX model, one can define the SLEX spectrum, a spectrum that is a decomposition of

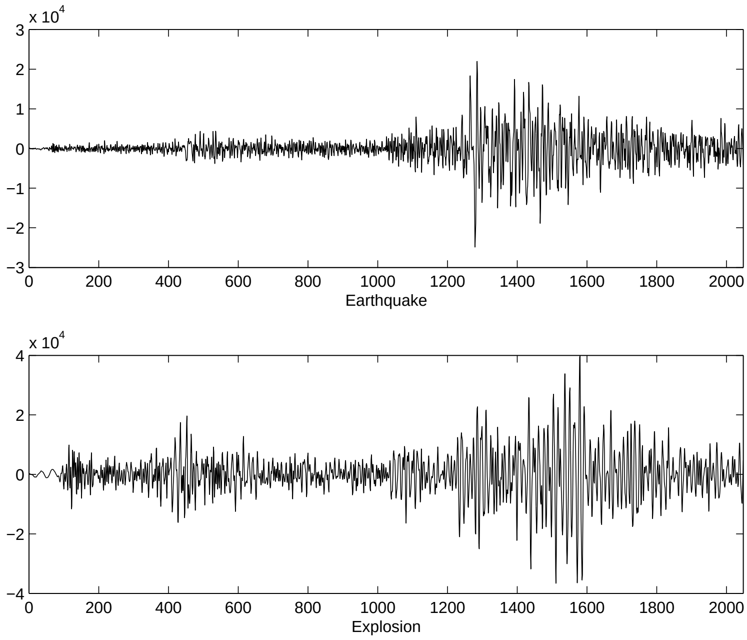

Figure 1.1 The top panel is a plot of an earthquake. The bottom panel is a plot of an explosion.

power over time and frequency. In doing this, the SLEX model remains in the spirit of traditional spectral analysis because it gives a spectrum that is a time-dependent analogue of the classical spectrum for stationary processes. The Dahlhaus model, in contrast, uses the Fourier basis functions which are not localized in time. The localization is provided by the time-varying transfer function. Thus, it does not give an explicit segmentation of the time-frequency plane. The SLEX model also provides a structure for asymptotic theory. This enables us to study the optimal property of our discriminant function. Moreover, the SLEX model is asymptotically mean square equivalent to the Dahlhaus model. Thus, the SLEX model can be related to the popular autoregressive models with time-varying coefficients, which are a subclass of the Dahlhaus model of locally stationary processes. In the computational aspect, since the SLEX functions are obtained by applying two specially constructed windows on the Fourier functions, the Fast Fourier Transform is used to calculate the SLEX transform. In addition, a fast algorithm, the Best Basis Algorithm (BBA) of

Coifman and Wickerhauser (1991), is used to search the best basis for representing processes. Thus, the SLEX model is also a model with computational efficiency.

A discriminant scheme can be considered a formulation with two steps: feature extraction and classification. For the SLEX model, the feature extraction step is to find a basis from the SLEX library which can illuminate the most differences among the classes from the library of SLEX bases. Saito (1994) proposed a method to form and calculate the cost function from a library of bases for discrimination. By this method, we can select the best basis for discrimination from the SLEX basis library which can illuminate the biggest difference among the classes. Once the basis is selected, the corresponding SLEX periodogram of the data can be obtained and serve as the extracted feature for discriminant function in the following classification step.

In the classification step, a discriminant function based on the SLEX model is proposed. Basically, this discriminant function is the difference in Kullback-Leibler divergence (Kullback and Leibler, 1951) between the periodogram of the data and spectrum of populations. The discriminant function will classify the data to the population with relatively smaller divergence. The discriminant function is constructed step by step in this dissertation including the construction of Kullback-Leibler divergence of the SLEX model. The optimal property of the discriminant function is also proved in this study. To evaluate the proposed method, some simulations and a real data analysis are performed.

The dissertation is organized as follows. In chapter 2, some background material is presented. In chapter 3, a discriminant scheme for a nonstationary time series will be derived, and an optimal property will be proven. In addition, a scheme will be organized into an algorithm so that the entire study can be applied to practical problems. In chapter 4, computer simulations will be performed and the discriminant scheme will be applied to real discriminant problem relating to the classification between the earthquakes and mining explosions in Scandinavia. In chapter 5, conclusion and future work will be discussed.

Chapter 2

Literature Review

The study is based on many previous works. Kullback-Leibler divergence (Kullback and Leibler, 1951), Frequency domain discriminant methods for stationary time series (see Shumway and Stoffer, 2000), the Dahlhaus model (Dahlhaus, 1997), the discriminant analysis proposed by Sakiyama and Taniguchi (2001), the SLEX model (Ombao, Raz, von Sachs and Guo, 2002), and Saito's feature extraction method (Saito, 1994) are six most important foundations. I will review them in this chapter.

2.1 Kullback-Leibler Divergence

The Kullback-Leibler divergence (Kullback and Leibler, 1951) is used as the measure of the "distance" or "divergence" between categories in this dissertation. It is also known as cross entropy, relative entropy or I-divergence. The definition of Kullback-Leibler divergence is as follows:

Definition 2.1 (Kullback-Leibler, 1951) *Let* $\mathbf{p}(x)$ *and* $\mathbf{q}(x)$ *denote the density functions of two different populations about the time series* x*.* $E_{\mathbf{p}}$ *denotes the expectation under the density* $\mathbf{p}(\cdot)$*. Then,*

$$\mathbf{I}(\mathbf{p};\mathbf{q}) = E_{\mathbf{p}}\left\{\log\frac{\mathbf{p}(x)}{\mathbf{q}(x)}\right\} \tag{2-1}$$

is called the Kullback-Leibler divergence.

Note that if $\mathbf{p} = \{p_i\}_{i=1}^n$, $\mathbf{q} = \{q_i\}_{i=1}^n$ are two nonnegative sequences with $\sum p_i = \sum q_i = 1$, the Kullback-Leibler divergence (relative entropy) between $\mathbf{p}$ and $\mathbf{q}$ is

$$\mathbf{I}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{2-2}$$

By convention, I set $\log 0 = -\infty$, $\log(x/0) = +\infty$ for $x > 0$, and $0 \times (\pm\infty) = 0$.

The Kullback-Leibler divergence satisfies $\mathbf{I}(\mathbf{p}, \mathbf{q}) \geq 0$ with the equality holding iff $\mathbf{p} \equiv \mathbf{q}$. This quantity is not a metric because it is not symmetric and does not satisfy the triangle inequality. In a sense, $\mathbf{I}(\mathbf{p}, \mathbf{q})$ measures how close $\mathbf{q}$ is to $\mathbf{p}$. Moreover, Kullback-Leibler divergence is an additive measure, that is, $\mathbf{I}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n \mathbf{I}(p_i, q_i)$. This property enables us to use it as the cost function in the Best Basis Algorithm which will be discussed later. If the number of populations, say C, is greater than two then $\mathbf{I}(\mathbf{p}^c)_{c=1}^C \equiv \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} \mathbf{I}(\mathbf{p}^i, \mathbf{p}^j)$.

2.2 Frequency Domain Discriminant Methods for Stationary Time Series

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.

Both time domain and frequency domain approaches to discrimination exist. For relatively short univariate series, a time domain approach that follows conventional multivariate discriminant analysis as described in conventional multivariate texts, such as Anderson (1984) or Johnson and Wichern (1992), may be preferable. We might even characterize differences by the autocovariance functions generated by different ARMA or state-space models. For longer multivariate time series that can be regarded as stationary after the common mean has been subtracted, the frequency domain approach will be easier computationally because the np dimensional vector in the time domain, represented here as $x = (x_1', x_t', \ldots, x_n')'$, with $x_t = (x_{t1}, \ldots, x_{tp})'$, will reduced to separate computations made on the p-dimensional discrete Fourier transforms (DFTs) where the DFTs of x_t is defined as follows

$$X(\nu_k) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \nu_k t} \tag{2-3}$$

with $\nu_k = k/n$ for $k = 0, 1, \ldots, n-1$. This computational benefit happens because of the approximate independence of the DFTs, $X(\nu_k), 0 \leq \nu_k \leq 1$.

The feature extraction approach often works well for discriminating between classes of univariate or multivariate series when there is a simple low-dimensional vector that seems to capture the essence of the differences between the classes.

It still seems sensible, however, to develop optimal methods for classification that exploit the differences between the multivariate means and covariance matrices in the time series case. Such methods can be based on the Whittle approximation to the log likelihood which is shown in the following.

$$\ln p_j(X) = \sum_{0<\nu_k<1/2} \{-\ln|f_j(\nu_k)| - X^*(\nu_k) f_j^{-1}(\nu_k) X(\nu_k)\}, \tag{2-4}$$

where the vector DFTs, say, $X(\nu_k)$, are assumed to be approximately normal, with means 0 and spectral matrices $f_j(\nu_k)$ for population Π_j, and $X^*(\nu_k)$ denotes the complex complement of $X(\nu_k)$. Equation (2-4) can be written in another form

$$\ln p_j(X) = \sum_{0<\nu_k<1/2} \{-\ln|f_j(\nu_k)| - \text{trace}(I(\nu_k)f_j^{-1}(\nu_k))\}, \tag{2-5}$$

where

$$I(\nu_k) = X(\nu_k)X^*(\nu_k) \tag{2-6}$$

denotes the periodogram matrix. For equal prior probabilities, we may assign an observation x into population Π_i, $i = 1, \ldots, G$ whenever

$$\ln p_i(X) > \ln p_j(X) \tag{2-7}$$

for $j \neq i$.

Numerous authors have considered various versions of discriminant analysis in the frequency domain. Alagón (1989) and Dargahi-Noubary and Laycock (1981) considered discriminant functions of the form

$$D_j(X) \;=\; \ln \pi_j - \sum_{0<\nu_k<1/2} \Big[\ln|f_j(\nu_k)| + X^*(\nu_k)f_j^{-1}(\nu_k)X(\nu_k)\Big] \tag{2-8}$$

in the univariate case when the means are zero and the spectra for the two groups are different. Taniguchi et al (1994) adopted (2-4) as a criterion and discussed its non-Gaussian robustness. Shumway (1982) reviews general discriminant functions in both the univariate and multivariate time series cases.

2.3 The Dahlhaus Model

Dahlhaus (1997) introduced a model for nonstationary time series whose spectral properties change slowly over time. He defined this kind of nonstationary time series to be a locally stationary process. Dahlhaus also built up a well established theoretical structure, such as estimation methods and asymptotic frameworks for the model. With these useful theoretical properties, the Dahlhaus model supports the basis of many other studies for nonstationary time series. For example, Sakiyama and Taniguchi (2001) proposed a discriminant method for the locally stationary process by applying the likelihood function derived by Dahlhaus (1997). The Dahlhaus model also served as a frame of reference in developing the SLEX model. In this section, I will introduce the definitions and theorems that will be useful in my work. For ease of discussion, I will restrict attention to the univariate zero-mean processes.

The definition of the Dahlhaus model is as follows:

Definition 2.2 *A sequence of zero mean stochastic processes* $\{X_{t,T}\}_{t=1,\ldots,T}$, $T \geq 1$, *is called locally stationary with transfer function* A^0 *if there exists a representation*

$$X_{t,T} = \int_{-pi}^{pi} A^0_{t,T}(\omega) \exp(i\omega) dZ(\omega), \tag{2-9}$$

where

1. $Z(\omega)$ *is a stochastic process on* $[-pi, pi]$ *with* $\overline{Z(\omega)} = Z(-\omega)$ *and*

$$\mathrm{cum}\{dZ(\omega_1), \ldots, dZ(\omega_k)\} = \eta(\sum_{j=1}^{k} \omega_j) \upsilon_k(\omega_1, \ldots, \omega_{k-1}) d\omega_1 \ldots d\omega_k \tag{2-10}$$

where cum$\{\ldots\}$ *denotes the cumulant of k-th order;* $\upsilon_1 = 0$, $\upsilon_2 = 1$, $|\upsilon_k(\omega_1, \ldots, \omega_{k-1})| \leq C$ *where* C *is a constant.* $\eta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ *is the period* 2π *extension of the Dirac delta function.*

2. *There exists a constant* Q *and a* 2π *periodic smooth function* $A : [0, 1] \times [-pi, pi] \to \mathbf{C}$ *with* $A(u, -\omega) = \overline{A(u, \omega)}$ *and*

$$\sup_{t,\omega} |A^0_{t,T} - A(t/T, \omega)| \leq QT^{-1} \tag{2-11}$$

for all T. $A(u, \omega)$ *is assumed to be continuous in* u.

Note that Definition 2.2 is an abstract setting for asymptotic statistic inference. Increasing T does not mean looking into the future. It means more data are available for each local structure.

With the definition of the Dahlhaus model, the corresponding time-varying spectrum in the Dahlhaus model is defined as follows:

Definition 2.3 *The time-varying or evolutionary spectrum of the Dahlhaus locally stationary process at time* $u \in [0, 1]$ *and frequency* $\omega \in [-1/2, 1/2]$ *is*

$$f(u, \omega) = |A(u, \omega)|^2. \tag{2-12}$$

The asymptotic Kullback-Leibler information divergence for a locally stationary process has been derived in Dahlhaus (1997). Dahlhaus used this information divergence as the measure of distance between models and be applied to the general minimum distance estimation procedure. However, rather than using information divergence for estimation, information divergence can also be used for discriminant analysis of locally stationary processes. Sakiyama et al (2001) proposed a discriminant analysis method for locally stationary processes based on the Kullback-Leibler information divergence of the Dahlhaus model.

The asymptotic Kullback-Leibler information divergence of Dahlhuas model is as follows: Suppose $f(u, \lambda)$ denotes the true spectral density and $f_\theta(u, \lambda)$ denotes the spectral density of the Dahlhaus model where $\theta \in \Theta$ and Θ is the parameter space of the model. The asymptotic Kullback-Leibler information divergence for the Dahlhaus model is

$$\mathcal{L}(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left\{ \log f_\theta(u, \lambda) + \frac{f(u, \lambda)}{f_\theta(u, \lambda)} \right\} d\lambda du. \tag{2-13}$$

An estimate, say $\mathcal{L}_T(\theta)$, of $\mathcal{L}(\theta)$ is proposed by Dahlhaus as follows:

$$\mathcal{L}_T(\theta) = \frac{1}{4\pi} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{ \log f_\theta(u_j, \lambda) + \frac{I_N(u_j, \lambda)}{f_\theta(u_j, \lambda)} \right\} d\lambda, \tag{2-14}$$

where

1. $d_N(u, \lambda) = \sum_{s=0}^{N-1} h(\frac{s}{N}) X_{[uT]-N/2+s+1,T} \exp(-i\lambda s)$, $h : \mathbf{R} \to \mathbf{R}$ is a data taper with $h(x) = 0$ for $x \notin [0, 1)$ and N is the segment length with midpoint $[uT]$. $[uT]$ denotes the largest integer less than uT.

2. $H_{k,N}(\lambda) = \sum_{s=0}^{N-1} h(\frac{s}{N})^k \exp(-i\lambda s)$.

3. $I_N(u, \lambda) = \frac{1}{2\pi H_{2,N}(0)} |d_N(u, \lambda)|^2$.

Thus, $I_N(u, \lambda)$ is the periodogram over a segment of length N with midpoint $[uT]$. The shift from segment to segment is denoted by S. So, the local periodogram I_N is calculated for every midpoint $t_j \equiv S(j-1) + N/2$ $(j = 1, \ldots, M)$ where $T = S(M-1) + N$. M denotes the numbers of blocks. Note that t_j can also be expressed as a rescaled time, $u_j = t_j/T$.

One of the very useful theorems introduced in the Dahlhaus (1997) is the central limit theorem. The proof of the consistency in the discriminant problem proposed by Sakiyama and Taniguchi (2001) is based on this central limit theorem. Some of my proofs are also inspired by this central limit theorem.

First, for $\phi : [0, 1] \times [-\pi, \pi] \to \mathbf{C}$, set

$$J_T(\phi) \equiv \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi(u_j, \lambda) I_n(u_j, \lambda)\, d\lambda \tag{2-15}$$

and

$$J(\phi) \equiv \frac{1}{M} \int_0^1 \int_{-\pi}^{\pi} \phi(u, \lambda) I_n(u, \lambda)\, d\lambda. \tag{2-16}$$

Then, the central limit theorem for the Dahlhaus model is as follows:

Theorem 2.1 (Dahlhaus, 1997) *Let $X_{t,T}$ be a locally stationary process with mean $\mu(u) = 0$. Suppose that the functions $A(u, \lambda)$* [from Definition 2.2] *and $\phi(u, \lambda)$ is 2π-periodic in λ and the periodic extensions are differentiable in u and λ with uniformly bounded derivative $(\partial/\partial u)(\partial/\partial\lambda)A$ and $(\partial/\partial u)(\partial/\partial\lambda)\phi$, and v_4 which was mentioned in (2-10) is continuous. Then*

$$\sqrt{T}(J_T(\phi) - J(\phi)) \xrightarrow{D} \xi, \tag{2-17}$$

where $\xrightarrow{D}$ means converging in distribution, ξ is a Gaussian random variable with mean zero, and

$$\begin{aligned} \mathrm{Var}(\xi) &= 2\pi c_h \int_0^1 \left[\int_{-\pi}^{\pi} \phi(u,\lambda)\left\{\overline{\phi(u,\lambda)} + \overline{\phi(u,-\lambda)}\right\} f(u,\lambda)^2\, d\lambda + \right. \\ &\left. \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \phi(u,\lambda)\overline{\phi(u,-\mu)} f(u,\lambda) f(u,\mu)\nu_4(\lambda,-\lambda,\mu)\, d\lambda d\mu\right] du \end{aligned}$$

with $c_h = (\int_0^1 h(u)^4 du)/(\int_0^1 h(u)^2 du)^2$ if $S = N$ and $c_h = 1$ if $S/N \to 0$.

2.4 Discriminant Analysis for the Dahlhaus Locally Stationary Processes

Sakiyama and Taniguchi (2001) presented a discriminant method for multivariate locally stationary processes based on Dahlhaus (1997). They introduced a discriminant statistic based on the difference of Kullback-Leibler information divergence. Essentially, a process will be classified by their discriminant statistic to a category with relatively smaller divergence. In their paper (Sakiyama and Taniguchi, 2001), the misclassification probability is shown to converge to zero as the size of the time series $T \to \infty$, i.e., the discriminant method is consistent. My discriminant method is prompted by the idea of Sakiyama and Taniguchi. So, I briefly review their work here.

Let $X_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(d)})'$ be a d dimensional locally stationary process. Let Π_1 and Π_2 denote two categories with time varying spectral density matrix $f(u, \lambda)$ and $g(u, \lambda)$, respectively [for the definition of $f(u, \lambda)$ and $g(u, \lambda)$, see (2-12)]. As introduced in the previous section, Dahlhuas derived an asymptotic Kullback-Leibler information divergence between two different models as $\mathcal{L}(\theta)$ in (2-13). Dahlhaus also proposed an estimator $\mathcal{L}_T(\theta)$ [see (2-14)] for $\mathcal{L}(\theta)$. Based on Dahlhaus' work, Sakiyama and Taniguchi proposed a discriminant statistic $D(f; g)$ as follows:

$$D(f;g) = \frac{1}{4\pi M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left[\log \left\{ \frac{|g(u_j, \lambda)|}{|f(u_j, \lambda)|} \right\} + \text{trace} \left\{ I_N(u_j, \lambda) \left(g_{u_j,\lambda}^{-1} - f_{u_j,\lambda}^{-1} \right) \right\} \right] \tag{2-18}$$

where M denotes the number of blocks, N denotes the size of each block, M and N satisfy $T = NM$, and I_N, as in the previous section, denotes the periodogram of the data $X_{t,T}$.

Let $\mathcal{L}_{T,g}$ denote the Kullback-Leibler information divergence between I_N and $g(u, \lambda)$. Let $\mathcal{L}_{T,f}$ denote the Kullback-Leibler information divergence between I_N and $f(u, \lambda)$. Then $D(f; g)$ is equal to $\mathcal{L}_{T,g} - \mathcal{L}_{T,f}$. Note that, in Dahlhaus (1997), $\mathcal{L}_T(\theta)$ estimates the distance between two different models. But, in Sakiyama and Taniguchi (2001), the $\mathcal{L}_T(\theta)$ is used to estimate the divergence between the periodogram of the $X_{t,T}$ and spectrum of category. To prove the consistency of the discriminant statistic $D(f; g)$, the following theorem is needed.

Theorem 2.2 (Sakiyama and Taniguchi, 2001) *Assuming that the minimum eigenvalues of $f(u_j, \lambda)$ and $g(u_j, \lambda)$ are greater than a positive constant C for all u and λ, then under* Π_1

$$\sqrt{T}[D(f;g) - E\{D(f;g)|\pi_1\}] \xrightarrow{D} N\{0, \sigma^2(f,g)\} \tag{2-19}$$

and under Π_2

$$\sqrt{T}[D(f;g) - E\{D(f;g)|\pi_2\}] \xrightarrow{D} N\{0, \sigma^2(g,f)\} \tag{2-20}$$

as $T \to \infty$. Here, $\sigma^2(f;g)$ and $\sigma^2(g;f)$ represent the variance matrices between $f(u,\lambda)$ and $g(u,\lambda)$. (see Sakiyama and Taniguchi, 2001, for details).

This property leads directly to the consistency of the discriminant method, i.e.,

$$\lim_{T\to\infty} P\{D(f;g) \le 0|\Pi_1\} = 0 \tag{2-21}$$

and

$$\lim_{T\to\infty} P\{D(f;g) > 0|\Pi_2\} = 0 \tag{2-22}$$

2.5 The SLEX Model

In the SLEX model (Ombao, Raz, von Sachs and Guo, 2002), a nonstationary time series can be represented as a linear combination of functions in the SLEX basis. The basis is chosen from a library of orthogonal SLEX functions which are orthogonal time-localized generalized versions of the Fourier complex exponentials. The discriminant scheme proposed in this dissertation is based on the SLEX model. In this section we will introduce how to construct the SLEX library, the SLEX model, and some of its asymptotic frameworks.

2.5.1 The SLEX Basis Vectors

The Fourier basis functions are perfectly localized in frequency and hence are ideal at representing a stationary time series. However, they cannot adequately represent a nonstationary time series, i.e., the time series with spectra that change over time. The most common approach to alleviating the time localization problem is to apply smooth compactly supported windows (tapers) to the Fourier basis functions. Windowed Fourier functions, however, are generally no longer orthogonal. In fact, the Balian-Low theorem says that there does not exist a smooth window such that the windowed Fourier basis functions are both orthogonal and localized in time and frequency (Wickerhauser, 1994). In this dissertation, we will use the SLEX basis functions which are simultaneously orthogonal and localized in time and frequency. They evade the Balian-Low obstruction because they are constructed by applying a projection operator, rather than a window, on the complex exponentials. It turns out that the action of a projection operator on a periodic vector is identical to applying two specially constructed smooth windows to the Fourier vectors.

Let $\alpha_1 > \alpha_0$ be two integer time points, let $|S| = \alpha_1 - \alpha_0$, and let the overlap $\epsilon = [\eta|S|]$, where $[.]$ denotes the greatest integer less than or equal to its argument, where $0 < \eta < -.5$. A SLEX basis vector $\psi_{S,\omega}(t)$ having support on the discrete time block $S = \{\alpha_0 - \epsilon + 1, \ldots, \alpha_1 - \epsilon\}$ and oscillating at frequency ω is defined to be

$$\psi_{S,\omega}(t) = \Psi_{S,+}(t)\text{exp}\left(i\omega\frac{t}{|S|}\right) + \Psi_{S,-}(t)\text{exp}\left(-i\omega\frac{t}{|S|}\right) \tag{2-23}$$

where $\omega \in [-pi, pi]$, $|S| = \alpha_1 - \alpha_0$. The functions $\Psi_{S,+}(t)$ and $\Psi_{S,-}(t)$ are two smooth windows which take the form

$$\Psi_{S,+}(t) = r^2\left(\frac{t-\alpha_0}{\epsilon}\right) r^2\left(\frac{\alpha_1 - t}{\epsilon}\right) \tag{2-24}$$

$$\Psi_{S,-}(t) = r\left(\frac{t-\alpha_0}{\epsilon}\right) r\left(\frac{\alpha_0 - t}{\epsilon}\right) - r\left(\frac{t-\alpha_1}{\epsilon}\right) r\left(\frac{\alpha_1 - t}{\epsilon}\right) \tag{2-25}$$

where $r(.)$ is called a rising cut-off function that is defined in Wickerhauser (1994). In our implementation, we use the sine rising cut-off function

$$r(u) = \sin\left[\frac{\pi}{4}(1+u)\right], \text{ where } u \in [-1, 1]. \tag{2-26}$$

A plot of $\Psi_{S,+}(t)$ and $\Psi_{S,-}(t)$ is shown in the Figure 2.1.

The SLEX basis vectors are defined at dyadic blocks that overlap. Though they overlap, they remain orthogonal. Moreover, the SLEX basis vectors are defined at frequencies $\omega_k = 2\pi k/|S|,\ k = -|S|/2 + 1, \ldots, |S|/2$.

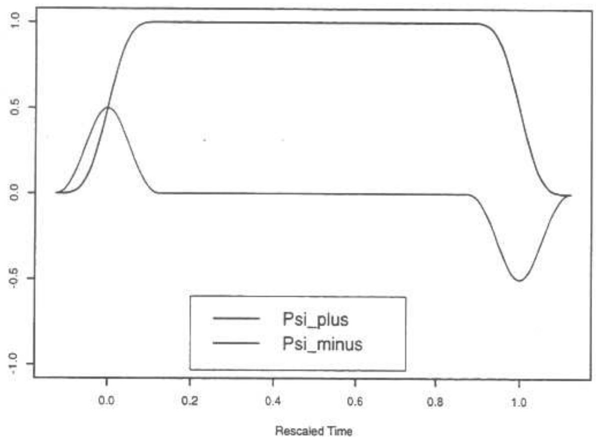

Figure 2.1 The windows used in the construction of the SLEX function

2.5.2 The SLEX Library

The SLEX library is a collection of orthogonal bases that consist of vectors having time support that is obtained by segmenting the time series in a dyadic manner. The library is constructed by first specifying the finest resolution level J. At resolution level j (where $j = 0, \dots, J$), the time series is divided into $M_j = 2^j$ overlapping blocks. We denote the block b on level j to be $S(b, j)$. The SLEX vectors on block $S(b, j)$ are allowed to oscillate at different fundamental frequencies $\omega_k = 2\pi k/M_j$ (where $k = -M_j/2 + 1, \dots, M_j/2$). To illustrate this further, consider Figure 2.2 where the SLEX library is constructed by setting $J = 2$. In this example library consists of 5 orthogonal bases and we enumerate these: (i.) the SLEX basis consists of vectors with support on the whole block denoted $S(0, 0)$; (ii.) vectors with support on two half blocks, namely $S(1, 0)$ and $S(1, 1)$; (iii.) vectors with support on four quarter blocks, namely $S(2, 0), S(2, 1), S(2, 2), S(2, 3)$; (iv.) vectors with support on $S(1, 0), S(2, 2), S(2, 3)$; (v.) vectors with support on $S(2, 0), S(2, 1), S(1, 1)$. Clearly, the SLEX basis vectors are allowed to have different lengths of support.

The SLEX transform consists of the coefficients corresponding to all the SLEX vectors constructed in the library. The SLEX coefficients on block $S = S(j, b)$ are

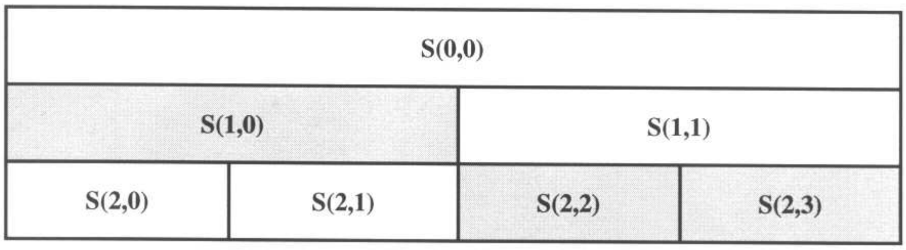

Figure 2.2 The SLEX library where $J = 2$

then computed as:

$$\hat{\theta}_{S,k,T} = \frac{1}{\sqrt{M_j}} \sum_{t \in \tilde{S}} X_{t,T} \, \overline{\phi_{S,\omega_k}(t)}, \quad k = -M_j/2+1, \ldots, M_j/2. \tag{2-27}$$

The SLEX periodogram, an analogue of the Fourier periodogram, is defined to be $\hat{\alpha}_{S,k,T} = |\hat{\theta}_{S,k,T}|^2$.

As stated, the SLEX transform divides the time axis in a dyadic manner. The dyadic partitioning allows for computational efficiency which is necessary for processing very large datasets. Moreover, the SLEX transform also smoothly partitions the time axis by allowing the SLEX vectors defined on adjacent blocks to have overlapping support. Despite the overlap, the SLEX vectors are still orthogonal under very mild conditions on the amount of overlap, which we now discuss. Let ϵ_j be the amount of overlap between the SLEX vectors at adjacent blocks at level j. In order for the SLEX basis vectors to be orthogonal, it is sufficient to set $\epsilon_j = \epsilon_J = |S(J)|/2$ for all $j = 0, \ldots, J-1$. When computing the SLEX coefficients on block $S(j,b)$, the range of the summation is over the "extended" block, say $\tilde{S} = \tilde{S}(j,b)$.

2.5.3 The Best Basis

The SLEX transform forms a library of orthonormal bases. In SLEX model, one basis will be selected from the SLEX library for representing a nonstationary random process. The Local Discriminant Basis Algorithm (LDBA) of Saito (1994) is applied to search for the desired basis in this study . A selected basis is called the "best basis" in the sense that it maximizes the cost function in the algorithm which is applied for searching the desired basis. The details of cost function and LDBA will be discussed in Section 2.6. Note that every basis from the SLEX library defines a particular segmentation in the SLEX library. The segmentation is implied in the support of the SLEX vectors that are included in the basis. Suppose that the selected best basis is defined to be B_T, we use $\cup_i S_i \sim B_T$ to denote the collection of the blocks S_i that are defined in the segmentation of basis B_T through this dissertation.

2.5.4 The SLEX Model of a Nonstationary Random Process

We now briefly define the SLEX model of a nonstationary random process. For details, refer to Ombao, Raz, von Sachs and Guo (2002). In a nutshell, the SLEX model gives a Crámer-like representation of a non-stationary random process. The elements of the model are the following: (i.) time-varying transfer function, (ii.) SLEX basis vectors and (iii.) orthogonal increment random process that is analogous to $Z(\omega)$ in Definition 2.2.

Suppose that we have a nonstationary time series $X_{t,T}, t = 0, \ldots, T-1$, and we want to model it by a representation of a collection of orthogonal SLEX basis vectors corresponding to a basis B_T. Let J_T be the finest resolution in B_T, i.e., the scale on which the basis vectors that have the narrowest time support live. We define $M_j := |S(j)| = T/2^j$ to be the number of points on each block on scale j where $j = 0, \ldots, J_T$. Correspondingly, the frequencies defined on this level are the grid frequencies $\omega_{k_j} = 2\pi k_j/M_j$ for $k_j = -M_j/2+1, \ldots, M_j/2$. The SLEX model is defined to be

$$X_{t,T} = \sum_{\bigcup S_i \sim B_T} \frac{1}{\sqrt{M_i}} \sum_{k=-M_i/2+1}^{M_i/2} \theta_{i,k,T} \phi_{i,k}(t) z_{i,k} \tag{2-28}$$

where $\theta_{i,k,T}$ is the transfer function defined at time block S_i and frequency k; $\phi_{i,k}$ 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. I summarize some facts about the SLEX model that will be relevant in deriving the theoretical results of my discriminant criterion.

Remarks:

(1.) Under the SLEX model, the transfer function $\theta_{S_i,k_i,T}$ is piecewise constant in time (in the block S_i) and is only a function of the frequency k_i. This is similar to the Dahlhaus model where the transfer function $A^0_{t,T}(\omega)$ is also approximately constant in a small interval about t/T.

(2.) The spectrum of the SLEX model, defined to be the square of the modulus of the transfer function, is also piecewise constant in time. To state this formally, for a given T and basis B_T, the SLEX spectrum $f_T(u, \omega_k)$ is a function of the rescaled time u in $[0, 1]$ and is defined on a particular set of grid frequencies $\omega_k = \omega_k(u)$ in $[-\pi, \pi]$. We first give the relationship between the rescaled time $u \in [0, 1]$ and the "time index" $t \in \{0, 1, \ldots, T-1\}$. To determine this, suppose that u is contained in a subinterval I of $[0, 1]$. There is some block S in $\bigcup_i S_i$ that is equivalent to I in the following manner. We write that $I \sim S$ if we have $t/T = u \in I \iff [uT] \in S$. Suppose now that $[uT] \in S_{i^*}$ for some block $S_{i^*} \in \bigcup_i S_i$. Then the corresponding grid frequencies are $\omega_{k_{i^*}} = k_{i^*}/M_{i^*}, \;\; k_{i^*} = -M_{i^*}/2+1, \ldots, M_{i^*}/2$. The spectrum is now defined $f_T(u, \omega_k) = |\theta_{i,k,T}|^2 \iff [uT] \in S_i$.

(3.) Under the SLEX model, with Gaussian increments $\{z_{S,k}\}$, the distribution of the SLEX coefficients $\hat{\theta}_{S,k,T}$ is complex normal, and the SLEX coefficients are independent.

(4.) The SLEX model and the Dahlhaus (1997) model of a locally stationary process are asymptotically mean-squared equivalent, see Ombao, et al. (2002). The equivalence implies that the SLEX model is capable of modeling locally stationary processes that have a smoothly time-varying spectrum despite being a sequence of piecewise stationary processes with spectra being piecewise constant in time.

2.5.5 Asymptotic Framework

The SLEX model is allowed to become more complex as $T \to \infty$. That is, the segments and level J_T can increase to as large as we want but not as fast as $\log_2 T$ increases, i.e., $2^{J_T}/T \to 0$ as $T \to \infty$. To introduce the asymptotic framework, we need some assumptions.

Assumption 2.1 *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 $L > 0$*

$$|f(u, \omega) - f(u, \omega^*)| \leq L|\omega - \omega^*|^{\mu} \tag{2-29}$$

Assumption 2.2 *There exists a hierarchical collection, $\mathcal{I}$, of dyadic subintervals of [0,1],*

$$\mathcal{I} = \{I_{lm} = [2^{-l}m, 2^{-l}(m+1)) : l = 0, 1, \ldots; m = 0, \ldots, 2^l - 1\}$$

and a subset of intervals I_v, $I_v[u_v, u_{v+1}) \in \mathcal{I}$ such that $\cup_v I_v = [0, 1]$ 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, in the transition points u_v between I_{v-1} and I_v allow for a finite number of possible jumps of finite height.

Assumption 2.3 *(a) As $K = K_T = log_2(T) \to \infty$, either J_T is fixed or $J_T \to \infty$ such that $K_T - J_T \to \infty$ (i.e. $2^{J_T}/T \to 0$). Furthermore, $J_T \leq J_{2T}$. (b) For the length M_j of each segment $S(j), j = 1, \ldots, J_T$, it is assumed that $M_{J_T}/M_j = c_j$ where $0 < c_j \leq 1$.*

With these assumptions, the evolutionary SLEX spectrum $f(u, \omega)$ can then be defined as follows:

Definition 2.4 (Ombao et al, 2002) *Let $\{\varphi_{0,0}\} \cup \{\psi_{l,m}\}_{l\geq 0, m\geq 0}$ denote the Haar wavelet basis of $L_2([0,1])$. The evolutionary SLEX spectrum $f(u, \omega)$ at frequency $\omega \in [-1/2, 1/2]$ is defined by the set of wavelet coefficients (i.e. the scaling coefficient $\beta_{-1,0}(\omega)$ and mother wavelet coefficients $\beta_{l,m}(\omega)$, $l = 0, 1, \ldots; m = 0, \ldots, 2^l - 1$) of its logarithmic wavelet expansion:*

$$\log f(u, \omega) = \beta_{-1,0}(\omega)\varphi_{0,0}(u) + \sum_{l=0}^{\infty} \sum_{m=0}^{2^l-1} \beta_{l,m}(\omega)\phi_{l,m}(u), \tag{2-30}$$

where

$$\beta_{-1,0}(\omega) = \int_0^1 \log f(u, \omega)\varphi_{0,0}(u)\, du, \tag{2-31}$$

$$\beta_{l,m}(\omega) = \int_0^1 \log f(u, \omega)\psi_{l,m}(u)\, du. \tag{2-32}$$

Once the evolutionary SLEX spectrum f is defined, how the SLEX spectrum f_T tends to f is very important for the proof of the consistency of my discriminant statistic. The following theorem shows that f_T actually converges to f.

Theorem 2.3 (Ombao et al, 2002) *Given the SLEX model as (2-28), with its SLEX spectrum f_T and evolutionary SLEX spectrum f, if Assumptions 2.1, 2.2, and 2.3 hold, and a sequence of frequencies $\omega_{k,T} \to \omega$, we have following:*

1. *Let* $u \in I_v = [u_v, u_{v+1}) \in \mathcal{I}$ *(with* $\cup_v I_v = [0,1]$*), let* $b_T = 2^{-J_T} \sim T^{\frac{-\mu}{\mu+s_v}}$ *as* $T \to \infty$*, let* $s \equiv \inf_v s_v$*. Then*

$$\begin{aligned} |\log f_T(u,\omega_{k,T}) - \log f(u,\omega)| &= O(T^{-\frac{s_v\mu}{\mu+s_v}}) && \text{(2-33)} \\ &= O(b_T^s), && \text{(2-34)} \end{aligned}$$

uniformly in $u \in I_v$

2. *Let* $s \equiv \inf_v s_v$ *and* $2^{J_T} \sim T^{\frac{\mu}{\mu+s}}$*. Then, as* $T \to \infty$*,*

$$\left| \int_0^1 du (\log f_T(u,\omega_{k,T}) - \log f(u,\omega))^2 \right| = O(T^{-\frac{2s\mu}{\mu+s}}) \qquad \text{(2-35)}$$

2.6 Feature Extraction Method

Saito (1994) proposed the Local Discriminant Basis Algorithm (LDBA) to select a basis from a library of orthonormal bases. The selected basis is called a local discriminant basis (LDB). The algorithm is fast [O(T)] and it can be proved that the selected basis maximizes the discriminant measure among all the bases. Since SLEX transform also forms a library of orthonormal bases, LDBA is an ideal algorithm for selecting the desired basis in my dissertation.

The essential idea of the LDBA is to search for a basis which can maximize the selected cost. In the LDBA, various possible costs are discussed. Suppose the Kullback-Leibler divergence $\mathbf{I}(\mathbf{p},\mathbf{q})$ [see (2-2)] is used as the cost in the algorithm, the question is what should we use as the **p** and **q** in the Kullback-Leibler divergence. In LDBA, Saito proposed the *time-frequency energy map* to serve as the densities **p** and **q**. I now describe how we may obtain **p** and **q**:

Definition 2.5 *Let* $\{\mathbf{x_i}^c\}_{i=1}^{N_c}$ *be a set of training signals belonging to class 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,k,l) *as*

$$\Gamma_c(j,b,k) \equiv \sum_{i=1}^{N_c} (\omega_{j,k,l}^T x_i^c)^2 / \sum_{i=1}^{N_c} \|\mathbf{x_i}^c\|^2 \tag{2-36}$$

for $j = 0,\ldots,J$ $k = 0,\ldots,2^j-1,$ $k = -M_j/2+1,\ldots,M_j/2.$

where $\omega_{j,k,l}^T x_i^c$ denotes the expanding coefficients of the i^{th} training data with respect to a selected basis. The *time-frequency energy map* satisfies $\sum_{j,b,k}\Gamma_c(j,b,k) = 1$ for all $c = 1,\ldots,C$.

Once we get the *time-frequency energy map* for each class, we can calculate the cost $(\mathbf{I}(\{\Gamma_c(j,b,\cdot)\}_{c=1}^C))$. Then, the LDBA can be applied to get the desired best basis for discrimination.

Set

$$\mathbf{I}(\{\Gamma_c(j,b,\cdot)\}_{c=1}^C) = \sum_{k=-M_j/2+1}^{M_j/2} \mathbf{I}(\Gamma_1(j,b,k),\Gamma_2(j,b,k),\ldots,\Gamma_C(j,b,k)). \tag{2-37}$$

Let $A_{j,b}$ represent the desired best basis, $B_{j,b}$ denote a set of basis vectors at level j and block b of a library of orthonormal bases, and $A_{j,b} = A_{j+1,2b} \oplus A_{j+1,2b+1}$ denotes that $A_{j,b}$ is set to be the combination of $A_{j+1,2b}$ and $A_{j+1,2b+1}$. Then, the LDBA algorithm for selecting the best basis is as follows:

Local Discriminant Basis Algorithm : (Staito, 1994)

Given a training dataset consisting of C classes of data $\{\{\mathbf{x}_i^{(c)}\}_{i=1}^{N_c}\}_{c=1}^C$,

Step 0: Choose a dictionary of orthonormal bases and specify the maximum depth of decomposition J.

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

Step 2: Set $A_{J,b} = B_{J,b}$ and $\triangle_{J,b} = \mathbf{I}(\{\Gamma_c(j,b,\cdot)\}_{c=1}^C)$ for $b = 0,\ldots,2^J-1$.

Step 3: Determine the $A_{j,k}$ for $j = J-1,\ldots,0$, $b = 0,\ldots,2^J-1$ by the following rule:

Set $\triangle_{j,b} = \mathbf{I}(\{\Gamma_c(j,b,\cdot)\}_{c=1}^C)$.

If $\triangle_{j,b} \geq \triangle_{j+1,2b} + \triangle_{j+1,2b+1}$, then $A_{j,b} = B_{j,b}$;

else $A_{j,b} = A_{j+1,2b} \oplus A_{j+1,2b+1}$ and set $\triangle_{j,b} = \triangle_{j+1,2b} + \triangle_{j+1,2b+1}$.

Chapter 3

Discrimination for Nonstationary Time Series Using the SLEX Model

In this chapter, I will introduce the discriminant scheme step by step. As mentioned previously, a discriminant scheme is composed of two sequential steps. The first is the feature extraction and the second is the classification. In Section 3.1, I will discuses how to extract the local features of the time series, i.e., how to select a best basis for the discriminant purpose from the SLEX library. In this step, I apply the local discriminant basis selection algorithm in Saito (1994). Then, in Section 3.2, I will derive the discriminant statistic for classification. In Section 3.3, I will prove that the discriminant statistic in Section 3.2 is consistent, i.e., the misclassification probability converges to zero as $T \to \infty$. The entire discriminant scheme will be combined as an algorithm in Section 3.4.

3.1 Feature Extraction

Choosing a desired basis from a library of bases is the so-called feature extraction. In this study, an algorithm proposed by Saito (1994) is employed to select the desired best basis which can best discriminate among classes from the library of SLEX bases.

Saito (1994) proposed a local discriminant basis selection algorithm (LDBA) [see Section 2.6] for choosing the "best" basis for discriminant purpose from an orthonormal bases library. This method is fast [O(T)] and it can be proven that the selected basis maximizes the discriminant measure on the time-frequency energy distributions between all bases in the library. As introduced in Section 2.5, the SLEX transform forms a library of orthonormal SLEX bases. Thus, the LDBA can be applied to this study.

The essential idea of BBA and LDBA is to choose a basis that can either minimize or maximize the overall selected cost. For example, if the selected cost is the divergence between two different categories, the algorithm (for discriminant purpose) should be set to choose the basis which can maximize the cost. In this study, the Kullback-Leibler divergence $\mathbf{I}(\mathbf{p}, \mathbf{q})$ [see Section 2.1] is employed as the cost.

After choosing the Kullback-Leibler divergence as the divergence measure, the following question is what should we use as the $\mathbf{p}$ and $\mathbf{q}$ in the relative entropy $\mathbf{I}(\mathbf{p}, \mathbf{q})$. The *time-frequency energy map* (Saito, 1994) is applied as the $\mathbf{p}$ and $\mathbf{q}$ in this study.

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

$$\Gamma_c^S(j,b,k) \equiv \sum_{i=1}^{N_c} (\hat{\theta}_{S_{(j,b)},k,T})^2 / \sum_{i=1}^{N_c} \|\mathbf{x}_i^c\|^2 \qquad (3\text{-}1)$$

where $\hat{\theta}_{S_{(j,b)},k,T}$ is the SLEX coefficient of level j, block b, and frequency $k/|S|$ [see (2-27)], $j = 0, \ldots, J$, $k = 0, \ldots, 2^j - 1$, and $k = -M_j/2 + 1, \ldots, M_j/2$. The *time-frequency energy map* can be treated as a density because $\sum_{j,b,k} \Gamma_c(j, b, k) = 1$ for each c. Moreover, this *time-frequency energy map* is a very ideal choice for this study because it fully utilizes the time-frequency localization characteristics of the dictionary of the SLEX basis.

Once the *time-frequency energy map* for each class is obtained, the cost function can be calculated by $\mathbf{I}(\{\Gamma_c^S(j, b, \cdot)\}_{c=1}^C)$, where

$$\mathbf{I}(\{\Gamma_c^S(j, b, \cdot)\}_{c=1}^C) = \sum_{i=1}^{C-1} \sum_{k=i+1}^{C} \mathbf{I}(\Gamma_c^S(j, b, i), \Gamma_c^S(j, b, k)). \tag{3-2}$$

Thus we can use the LDBA to get the desired best basis from the SLEX bases library for discrimination. Finally, the SLEX periodogram $\hat{\alpha}_{S_{(j,b)},k,t} = |\hat{\theta}_{S_{(j,b)},k,T}|^2$ can be obtained directly from the SLEX basis that was selected.

3.2 Discriminant Statistic

After obtaining the best basis and the SLEX periodogram, the next step is to construct a discriminant statistic. The idea of the Kullback-Leibler information is employed to build the discriminant statistic in this dissertation. Suppose there are two categories of time series, Π_1 and Π_2. Let $f_{1,T}$ and $f_{2,T}$ be the spectra for the respective categories and let f_1 and f_2 be the corresponding evolutionary spectra. The basic idea for constructing the discriminant statistic is that a series $\{X_t\}$ will be classified to the category Π_1 if the divergence between the periodogram of $\{X_t\}$ and $f_{1,T}$ is relatively smaller than the divergence between the periodogram of $\{X_t\}$ and $f_{2,T}$. To achieve this goal, first, I will show that the SLEX transform is asymptotically independent between different blocks of the selected basis. Second, I will derive a divergence measure between two categories, say $L(f_1, f_2)$, based on the asymptotic Kullback-Leibler divergence of the SLEX model. Third, I will propose an asymptotic unbiased estimator $L^*(f_{1,T}, f_{2,T})$ for $L(f_1, f_2)$. Last, I will construct the discriminant statistic with $L^*(f_{1,T}, f_{2,T})$.

The following lemma shows that the SLEX transform is asymptotically independent between different blocks of the selected basis.

Lemma 3.1 *Let $\{X_t\}$ be a zero mean time series. Suppose we observe $X_1, \ldots, X_{T_1}$ and $X_{m+T_1+1}, \ldots, X_{m+2T_1}$ for $T_1 \geq 1$ and $m \geq 0$ from two different blocks of the basis of the SLEX model, say S_1, S_2. Let the length of the support of S_1 and S_2 be L, $\gamma(s,t) = \mathrm{Cov}(X_s, X_{t+m})$ for $s, t = 0, \ldots, T_1 - 1$, and $\hat{\theta}_{S_1,j}$ and $\hat{\theta}_{S_2,j}$ denote the SLEX transform for blocks S_1 and S_2 respectively. Define a function $g(h)$ such that $|\gamma(s,t)| < g(|s-t|)$ and $\sum g(h) < \infty$. Then, $\hat{\theta}_{S_1,j}$ and $\hat{\theta}_{S_2,k}$ are asymptotically ($T_1 \to \infty$) independent for any $j, k = -T_1/2+1, \ldots, T_1/2$.*

Proof.

Since $E(\hat{\theta}_{S_1,j}) = 0$, $E(\hat{\theta}_{S_2,j}) = 0$, and $|\overline{\phi_{S_1,\omega_j}(t)}\phi_{S_2,\omega_k}(s)| \leq 1$,

$$
\begin{aligned}
|E\{\hat{\theta}_{S_1,j}\overline{\hat{\theta}_{S_2,k}}\}| &\leq \frac{1}{T_1}\sum_{t=1}^{L}\sum_{s=1}^{L}|\gamma(s,t)| \\
&\leq \frac{1}{T_1}\sum_{t=1}^{L}\sum_{s=1}^{L} g(|s-t|) \\
&= \sum_{j=1}^{L}\frac{j}{T_1}g(j) + \sum_{j=L+1}^{2L-1}\frac{2L-j}{T_1}g(j) \\
&\leq \sum_{j=1}^{L}\frac{j}{T_1}g(j) + \sum_{j=L+1}^{2L-1} g(j)
\end{aligned}
$$

Let $T_1 \to \infty$, then

$$
\sum_{j=1}^{L}\frac{j}{T_1}g(j) \to 0
$$

by Kronecker's Lemma, then

$$\sum_{j=L+1}^{2L-1} g(j) \to 0$$

∎

Before moving forward to derive the divergence measure between two categories in the SLEX model, I also need the following lemma:

Lemma 3.2 *Let f_T denotes the SLEX spectrum and f denotes the corresponding evolutionary SLEX spectrum. Under Assumption 2.1, 2.2, and 2.3 in Section 2.5.5*

$$\begin{aligned}\frac{1}{T}\sum_{i:\cup_i S_i = B_T}\sum_{k_i=|S_i|/2+1}^{|S_i|/2} \log f_T(u_i,\omega_{k_i}) &= \int_0^1\int_{-\pi}^{\pi}\log f(u,\omega)d\omega du + O\left\{\left(\frac{\sup|S_i|}{T}\right)^{\mu}\right\} \\ &\quad + O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\} + O(b_T^s)\end{aligned}$$

where $b_T = M_{J_T}/T = 2^{-J_T}$, $s := \inf_v s_v$, and μ is the order of Hölder condition in Assumption 2.1.

Proof.

$$\begin{aligned}\frac{1}{T}\sum_{i:\cup_i S_i = B_T}\sum_{k_i=|S_i|/2+1}^{|S_i|/2} \log f_T(u_i,\omega_{k_i}) &= \frac{1}{T}\sum_{i:\cup_i |S_i| = B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|} (\log f(u_i,\omega_{k_i}) + O(b_T^s)) \\ &= \frac{1}{T}\sum_{i:\cup_i |S_i| = B_T}\left\{\sum_{k_i=-|S_i|/2+1}^{|S_i/2|} \log f(u_i,\omega_{k_i}) + O(b_T^s)\right\} \\ &= \sum_{i:\cup_i |S_i| = B_T}\frac{|S_i|}{T}\left[\int_{-\pi}^{\pi}\log f(u_i,\omega)\,d\omega + \right. \\ &\quad \left. O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\} + O(b_T^s)\right] \\ &= \sum_{i:\cup_i |S_i| = B_T}\frac{|S_i|}{T}\int_{-\pi}^{\pi}\log f(u_i,\omega)\,d\omega +\end{aligned}$$

$$
\begin{aligned}
& O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\} + O(b_T^s) \\
= & \int_0^1 \int_{-\pi}^{\pi} \log f(u_i, \omega)\, d\omega d\mu + O\left\{\left(\frac{\sup |S_i|}{T}\right)^{\mu}\right\} \\
& + O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\} + O(b_T^s)
\end{aligned}
$$

∎

Under the assumption that the increments $z_{i,k}$ are Gaussian, the SLEX transform is complex normal. (The property still holds asymptotically without the Gaussian assumption.) By Lemma 3.1, assuming the mean is zero, the Whittle approximation to the log likelihood function of SLEX transform under population Π_j is

$$
\begin{aligned}
\log p_j(\hat{\theta}) &= \sum_{Si}\sum_{k_i}\left\{-\log f_T(u_i, \omega_{k_i}) - \{\overline{\hat{\theta}_{S_i,k_i,T}}\, f_T^{-1}(u_i, \omega_{k_i})\, \hat{\theta}_{S_i,k_i,T}\}\right\} && \text{(3-3)} \\
&= \sum_{Si}\sum_{k_i}\left\{-\log f_T(u_i, \omega_{k_i}) - \{\hat{\alpha}_{S_i,k_i,T}\, f_T^{-1}(u_i, \omega_{k_i})\}\right\} && \text{(3-4)}
\end{aligned}
$$

where $\hat{\theta}_{S_i,k_i,T}$ is the SLEX coefficients, $f_T(u_i, \omega_{k_i})$ is the SLEX spectrum, and $\hat{\alpha}_{S_i,k_i,T}$ is the SLEX periodogram. The divergence measure between two categories in the SLEX model is derived in the following theorem:

Theorem 3.1 *Suppose there are two categories,* Π_1 *and* Π_2*. Let* $f_{1,T}(u_i, \omega_{k_i})$ *and* $f_{2,T}(u_i, \omega_{k_i})$ *denote the SLEX spectrum of* Π_1 *and* Π_2 *at time recaled* u_i *and frequency* ω_{k_i} *respectively. Let* $f_1(u, \omega)$ *and* $f_2(u, \omega)$ *denote the corresponding evolutionary SLEX spectrum. Then, the divergence measure between two categories of the SLEX model is*

$$
L(f_1, f_2) = \int_0^1 \int_{-\pi}^{\pi} \left\{\log f_2(u, \omega) + \frac{f_1(u, \omega)}{f_2(u, \omega)}\right\} d\omega du \tag{3-5}
$$

PROOF.

Let $\log p_1(\hat{\theta})$ and $\log p_2(\hat{\theta})$ denote the log likelihood function of SLEX coefficients under each category. Then, the asymptotic Kullback-Leibler information divergence of SLEX model is

$$
\begin{aligned}
\lim_{T\to\infty} 1/T\, E_1\left\{\frac{\log p_1(\hat{\theta})}{\log p_2(\hat{\theta})}\right\} &= \lim_{T\to\infty} 1/T\, E_1\left\{\log p_1(\hat{\theta}) - \log p_2(\hat{\theta})\right\} \\
&= \lim_{T\to\infty}\left[1/T\, E_1 \sum_{S_i}\sum_{k_i}\left\{-\log\frac{f_{1,T}(u_i,\omega_{k_i})}{f_{2,T}(u_i,\omega_{k_i})}\right.\right. \\
&\quad \left.\left.+\hat{\alpha}_{S_i,K_i,T}(f_{2,T}^{-1}(u_i,\omega_{k_i}) - f_{1,T}^{-1}(u_i,\omega_{k_i}))\right\}\right] \\
&= \lim_{T\to\infty}\left[1/T \sum_{S_i}\sum_{k_i}\left\{-\log\frac{f_{1,T}(u_i,\omega_{k_i})}{f_{2,T}(u_i,\omega_{k_i})}\right.\right. \\
&\quad \left.\left.+f_{1,T}(u_i,\omega_{k_i})(f_{2,T}^{-1}(u_i,\omega_{k_i}) - f_{1,T}^{-1}(u_i,\omega_{k_i}))\right\}\right] \\
&= \lim_{T\to\infty}\left[1/T \sum_{S_i}\sum_{k_i}\left\{-\log\frac{f_{1,T}(u_i,\omega_{k_i})}{f_{2,T}(u_i,\omega_{k_i})}\right.\right. \\
&\quad \left.\left.+\frac{f_{1,T}(u_i,\omega_{k_i})}{f_{2,T}(u_i,\omega_{k_i})} - 1\right\}\right] \\
&= \lim_{T\to\infty}\left[1/T \sum_{S_i}\sum_{k_i}\left\{\log f_{2,T}(u_i,\omega_{k_i}) + \frac{f_{1,T}(u_i,\omega_{k_i})}{f_{2,T}(u_i,\omega_{k_i})}\right\}\right] \\
&\quad +\text{constant}
\end{aligned}
$$

By Lemma 3.2,

$$
\lim_{T\to\infty} 1/T\, E_1\left\{\log\frac{p_1(k_i)}{p_2(k_i)}\right\} = \int_0^1\int_{-\pi}^{\pi}\left\{|\log f_2(u,\omega)| + \frac{f_1(u,\omega)}{f_2(u,\omega)}\right\} d\omega du + \text{constant}
$$

So, we can regard $L(f_1,f_2) = \int_0^1\int_{-\pi}^{\pi}\{\log|f_2(u,\omega)| + \frac{f_1(u,\omega)}{f_2(u,\omega)}\}d\omega du$ as a distance measure between the two categories Π_1 and Π_2 in the SLEX model. ∎

Let

$$
L^*(\hat{\alpha}_T, f_{2,T}) = \frac{1}{T}\sum_{i:\cup_i S_i = B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i|/2}\left\{\log f_{2,T}(u_i,\omega_{k_i}) + \frac{\hat{\alpha}_{S_i,k_i,T}}{f_{2,T}(u_i,\omega_{k_i})}\right\}
$$

where $\hat{\alpha}_{S_i,k_i,T}$ denotes the SLEX periodogram of a time series from a category whith spectrum $f_{1,T}$. Then, by Lemma 3.2,

$$E_1[L^*(\hat{\alpha}_T, f_{2,T})] = L(f_1, f_2) + O\left\{\left(\frac{\sup|S_i|}{T}\right)^{\mu}\right\} + O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\} + O(b_T^s)$$

That is, $L^*(\hat{\alpha}_T, f_{2,T})$ is an asymptotic unbiased estimator of $L(f_1, f_2)$. So I use L^* to define my discriminant statistic in the following definition.

Definition 3.1 *Suppose there are two categories,* Π_1 *and* Π_2*. Let* $f_{1,T}(u_i, \omega_{k_i})$ *and* $f_{2,T}(u_i, \omega_{k_i})$ *denote the SLEX spectrum of* Π_1 *and* Π_2 *respectively and* $\hat{\alpha}_{S_i,k_i,T}$ *is the SLEX periodogram of a process* x*. Then,*

$$\begin{aligned} D(f_{1,T}, f_{2,T}; x) &\equiv L^*(\hat{\alpha}_T, f_{2,T}) - L^*(\hat{\alpha}_T, f_{1,T}) && \text{(3-6)} \\ &= \frac{1}{T} \sum_{i:\cup_i S_i = B_T} \sum_{k_i=-|S_i|/2+1}^{|S_i|/2} \left\{ \log \frac{|f_{2,T}(u_i, \omega_{k_i})|}{|f_{1,T}(u_i, \omega_{k_i})|} \right. \\ &\quad \left. + \hat{\alpha}_{S_i,k_i,T} \left[f_{2,T}^{-1}(u_i, \omega_{k_i}) - f_{1,T}^{-1}(u_i, \omega_{k_i}) \right] \right\} && \text{(3-7)} \end{aligned}$$

is the discriminant statistic for this study. If $D(f_{1,T}, f_{2,T}; x) \geq 0$ *we choose category* Π_1*. Otherwise we choose category* Π_2*.*

Remarks:

Since $\hat{\alpha}_{S_i,k_i,T}$ are distributed as $f_T(u_i, \omega_{k_i})\chi_2^2/2$ (except at frequencies $k = 0, M_i/2$ where we have χ_1^2), the discriminant statistic $D(f_{1,T}, f_{2,T}; x)$ can also be approached via the likelihood ratio.

3.3 Asymptotic Results on Consistency

The SLEX model with a structure of asymptotic theory enables us to develop the asymptotic results of the discriminant statistic $D(f_{1,T}, f_{2,T}; x)$. The assumptions and theory that will be used in the proofs of this section are described in Section 2.5.5. For further details, we refer to Ombao, Raz, von Sachs and Guo (2002).

To be clear about the expression in the following derivation, I first define the following functions. Let $\phi_T(u,\omega) = 1/f_{2,T}(u,\omega) - 1/f_{1,T}(u,\omega)$ and $\phi(u,\omega) = 1/f_2(u,\omega) - 1/f_1(u,\omega)$ where $f_{i,T}$ denotes the spectrum of population Π_i, and f_i denotes the corresponding evolutionary spectrum, for $i = 1, 2$. Define

$$H_T(\phi_T) = \frac{1}{T} \sum_{i:\cup_i |S_i| = B_T} \sum_{k_i = -|S_i|/2+1}^{|S_i/2|} \phi_T(u_i, \omega_{k_i}) \hat{\alpha}_{S_i,k_i,T} \tag{3-8}$$

and

$$H(\phi) = \int_0^1 \int_{-\pi}^{\pi} \phi(u,\omega) f(u,\omega)\, d\omega du \tag{3-9}$$

where $\hat{\alpha}_{S_i,k_i,T}$ is the SLEX periodogram, $\cup_i |S_i| = B_T$ is the selected best basis, and $f(u,\omega)$ is the evolutionary SLEX spectrum.

Before proving the consistency of the $D(f_{1,T}, f_{2,T}; x)$, I present some properties of $H_T(\phi_T)$ and $H(\phi)$. In Lemmas 3.3 and 3.4, I will assume, without loss of generality (WOLOG), that the time series to be classified is from Π_1; however, to ease the notation, I will not explicitly display this dependence. Thus, all expectations in Lemmas 3.3 and 3.4 are with respect to Π_1, and I use f_T and f generically for $f_{1,T}$ and f_1. The following lemma shows that $H_T(\phi_T)$ is an asymptotically unbiased estimator of $H(\phi)$.

Lemma 3.3 *Under Assumption 2.1, 2.2, and 2.3 in Section 2.5.5*

$$E\left(H_T(\phi_T)\right) = H(\phi) + O\left\{ \left(\frac{\sup |S_i|}{T} \right)^{\mu} \right\} + O\left(\frac{2^{J_T}}{T} \right)^{\mu} + O(b_T^s) \tag{3-10}$$

where $b_t = M_{J_T}/T = 2^{-J_T}$, $s := \inf_v s_v$, *and* μ *is the order of Hölder condition in Assumption 2.1.*

Proof.

$$
\begin{aligned}
E(H_T(\phi_T)) &= \frac{1}{T}\sum_{i:\cup_i|S_i|=B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\phi_T(u_i,\omega_{k_i})f_T(u_i,\omega_{k_i}) \\
&= \frac{1}{T}\sum_{i:\cup_i|S_i|=B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\left[\phi(u_i,\omega_{k_i})+O(b_T^s)\right]\left[f(u_i,\omega_{k_i})+O(b_T^s)\right] \\
&= \frac{1}{T}\sum_{i:\cup_i|S_i|=B_T}\left\{\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\phi(u_i,\omega_{k_i})f(u_i,\omega_{k_i})+O(b_T^s)\right\} \\
&= \sum_{i:\cup_i|S_i|=B_T}\frac{|S_i|}{T}\left[\int_{-\pi}^{\pi}\phi(u_i,\omega)f(u_i,\omega)\,d\omega+O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\}+O(b_T^s)\right] \\
&= \sum_{i:\cup_i|S_i|=B_T}\frac{|S_i|}{T}\int_{-\pi}^{\pi}\phi(u_i,\omega)f(u_i,\omega)\,d\omega+O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\}+O(b_T^s) \\
&= \int_0^1\int_{-\pi}^{\pi}\phi(u_i,\omega)f(u_i,\omega)\,d\omega d\mu+O\left\{\left(\frac{\sup|S_i|}{T}\right)^{\mu}\right\} \\
&\quad +O\left\{\left(\frac{2^{J_T}}{T}\right)^{\mu}\right\}+O(b_T^s)
\end{aligned}
$$

∎

I also need the following lemma to prove the consistency.

Lemma 3.4 *Suppose that the fourth moment of the orthonormal increment process* $z_{S_i,\omega_{k_i}}$ *in the SLEX model is finite and equal to a constant* C. *Under Assumption 2.1, 2.2, and 2.3,*

$$\text{Var}\left\{H_T(\phi_T)\right\}=O(T^{-1})+O(b_T^s/T)+O(2^{J_T}/T) \tag{3-11}$$

Proof.

Let

$$P_i=\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\phi_T(u_i,\omega_{k_i})\hat{\alpha}_{S_i,k_i,T},$$

then

$$
\begin{aligned}
\mathrm{Var}(H_T(\phi_T)) &= \mathrm{Var}\left[\sum_{i:\cup_i|S_i|=B_T} \frac{|S_i|}{T}\frac{1}{|S_i|}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|} \phi_T(u_i,\omega_{k_i})\,\hat{\alpha}_{S_i,k_i,T}\right] \\
&= \frac{1}{T^2}\sum_{i:\cup_i|S_i|=B_T} \mathrm{Var}(P_i) + \frac{1}{T^2}\sum\sum_{i\neq j}\mathrm{Cov}(P_i,P_j)
\end{aligned}
$$

in obvious notation for the summations. Now,

$$
\begin{aligned}
\frac{1}{T^2}\sum_{i:\cup_i|S_i|=B_T}\mathrm{Var}(P_i) &= \frac{1}{T^2}\sum_{i:\cup_i|S_i|=B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\phi_T^2(u_i,\omega_{k_i})\mathrm{Var}(\hat{\alpha}_{S_i,k_i,T}) \\
&= \frac{1}{T^2}\sum_{i:\cup_i|S_i|=B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\phi_T^2(u_i,\omega_{k_i})f_T^2(u_i,\omega_{k_i}) \\
&= \frac{1}{T^2}\sum_{i:\cup_i|S_i|=B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\left[\phi^2(u_i,\omega_{k_i})+O(b_T^s)\right]\left[f^2(u_i,\omega_{k_i})+O(b_T^s)\right] \\
&= \frac{1}{T^2}\sum_{i:\cup_i|S_i|=B_T}\sum_{k_i=-|S_i|/2+1}^{|S_i/2|}\phi^2(u_i,\omega_{k_i})f^2(u_i,\omega_{k_i})+O(b_T^s/T) \\
&= O(T^{-1})+O(b_T^s/T)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{T^2}\sum\sum_{i\neq j}\mathrm{Cov}(P_i,P_j) &= \frac{1}{T^2}\sum\sum_{i\neq j}\mathrm{Cov}\left[\sum_{k_i}\phi_T(u_i,\omega_{k_i})\hat{\alpha}_{S_i,k_i,T},\sum_{k_j}\phi_T(u_j,\omega_{k_j})\hat{\alpha}_{S_j,k_j,T}\right] \\
&= \frac{1}{T^2}\sum\sum_{i\neq j}\sum_{k_i}\sum_{k_j}\mathrm{Cov}\left[\phi_T(u_i,\omega_{k_i})\hat{\alpha}_{S_i,k_i,T},\phi_T(u_j,\omega_{k_j})\hat{\alpha}_{S_j,k_j,T}\right] \\
&= \frac{1}{T^2}\sum\sum_{i\neq j}\sum_{k_i}\sum_{k_j}\phi_T(u_i,\omega_{k_i})\phi_T(u_j,\omega_{k_j})\mathrm{Cov}\left[\hat{\alpha}_{S_i,k_i,T},\hat{\alpha}_{S_j,k_j,T}\right] \\
&= \frac{1}{T^2}\sum\sum_{i\neq j}\sum_{k_i}\sum_{k_j}\left[\phi(u_i,\omega_{k_i})\phi(u_j,\omega_{k_j})+O(b_T^s)\right] \\
&\quad \mathrm{Cov}\left(\left|\frac{1}{\sqrt{|S_i|}}\sum_t x_t\phi_{S_i,\omega_{k_i}}(t)\right|^2,\left|\frac{1}{\sqrt{|S_j|}}\sum_t x_t\phi_{S_j,\omega_{k_j}}(t)\right|^2\right) \\
&= \frac{1}{T^2}\sum\sum_{i\neq j}\sum_{k_i}\sum_{k_j}\phi(u_i,\omega_{k_i})\phi(u_j,\omega_{k_j})|\theta_{S_i,\omega_{k_i},T}|^2|\theta_{S_j,\omega_{k_j},T}|^2 \\
&\quad \mathrm{Cov}(|z_{S_i,\omega_{k_i}}|^2,|z_{S_j,\omega_{k_j}}|^2)+O(2^{(1-s)J_T}/T).
\end{aligned}
$$

Without loss of generality, assume the set of frequencies at block S_i is coarser than that at block S_j. Since $z_{S_i,\omega_{k_i}}$ is orthonormal,

$$\begin{aligned}
\frac{1}{T^2}\sum\sum_{i\neq j}\mathrm{Cov}(P_i,P_j) &= \frac{1}{T^2}\sum\sum_{i\neq j}\sum_{k_i}\phi(u_i,\omega_{k_i})\phi(u_j,\omega_{k_i})|\theta_{S_i,\omega_{k_i},T}|^2|\theta_{S_j,\omega_{k_i},T}|^2\mathrm{C} \\
&\quad +O(2^{(1-s)J_T}/T) \\
&= O(2^{J_T}/T)+O(2^{(1-s)J_T}/T) \\
&= O(2^{J_T}/T).
\end{aligned}$$

So,

$$\mathrm{Var}(H_T(\phi_T)) = O(T^{-1}) + O(\frac{b_T^s}{T}) + O(2^{J_T}/T).$$

∎

With these lemmas, I can start approaching the proof of consistency.

Lemma 3.5 *Let $D(f_{1,T},f_{2,T};x)$ be the discriminant statistic defined in (3.1) where $f_{1,T}$ and $f_{2,T}$ denote the spectrum of the category Π_1 and Π_2. Then, under Π_1, $D(f_{1,T},f_{2,T};x)$ converges to $E[D(f_{1,T},f_{2,T};x)|\Pi_1]$ in probability. Under Π_2, $D(f_{1,T},f_{2,T};x)$ converges to $E[D(f_{1,T},f_{2,T};x)|\Pi_2]$ in probability.That is, under Π_1,*

$$D(f_{1,T},f_{2,T};x) \xrightarrow{P} E[D(f_{1,T},f_{2,T};x)|\Pi_1]. \tag{3-12}$$

Under Π_2,

$$D(f_{1,T},f_{2,T};x) \xrightarrow{P} E[D(f_{1,T},f_{2,T};x)|\Pi_2]. \tag{3-13}$$

Proof.

I only prove the convergence under Π_1. The proof of the convergence under Π_2 is similar. To prove the convergence, it is sufficient to show the conditional variance of $D(f_{1,T},f_{2,T};x)$ converge to 0 as $T\to\infty$.

Set

$$R_T = 1/T \sum_{i:\cup_i S_i = B_T} \sum_{k_i=-|S_i|/2+1}^{|S_i|/2} \log \frac{|f_{2,T}(u_i, \omega_{k_i})|}{|f_{1,T}(u_i, \omega_{k_i})|}. \tag{3-14}$$

Then,

$$D(f_{1,T}, f_{2,T}; x) = R_T + H_T(\phi_T), \tag{3-15}$$

and

$$E[D(f_{1,T}, f_{2,T}; x)|\Pi_1] \;=\; R_T + E[H_T(\phi_T)|\Pi_1].$$

Hence,

$$E\left\{[D(f_{1,T}, f_{2,T}; x) - E\{D(f_{1,T}, f_{2,T}; x)\}]^2 |\Pi_1\right\} \;=\; Var(H_T(\phi_T)|\Pi_1) \to 0$$

.

as $T \to \infty$ by Lemma 3.4. ∎

The consistency of the discriminant statistic $D(f_{1,T}, f_{2,T}; x)$ is stated and proven as follows:

Theorem 3.2 *Under the established notation and conditions,*

$$\lim_{T\to\infty} P(2|1) \;=\; \lim_{T\to\infty} P(D(f_{1,T}, f_{2,T}; x) < 0|\Pi_1) = 0.$$

$$\lim_{T\to\infty} P(2|1) \;=\; \lim_{T\to\infty} P(D(f_{1,T}, f_{2,T}; x) \geq 0|\Pi_2) = 0.$$

Proof.

I prove the first result, the second result follows in a similar manner. Since

$$D(f_{1,T}, f_{2,T}; x) \;=\; \frac{1}{T} \sum_{i:\cup_i S_i = B_T} \sum_{k_i=-|S_i|/2+1}^{|S_i|/2} \left\{ \log \frac{|f_{2,T}(u_i, \omega_{k_i})|}{|f_{1,T}(u_i, \omega_{k_i})|} \right.$$
$$\left. + \hat{\alpha}_{S_i,k_i,T}\left[f_{2,T}^{-1}(u_i, \omega_{k_i}) - f_{1,T}^{-1}(u_i, \omega_{k_i})\right]\right\},$$

$$E[D(f_{1,T}, f_{2,T}; x)|\Pi_1] = \int_0^1 \int_{-\pi}^{\pi} \left[\log \frac{|f_{2,T}(u,\omega)|}{|f_{1,T}(u,\omega)|} + f_{1,T}(u,\omega) f_{2,T}^{-1}(u,\omega) - 1\right] d\omega du$$
$$+O((sup|S_i|/T)^{\mu}) + O((2^{J_T}/T)^{\mu}) + O(b_T^s).$$

Note that

$$\log \frac{|f_{2,T}(u,\omega)|}{|f_{1,T}(u,\omega)|} + f_{1,T}(u,\omega) f_{2,T}^{-1}(u,\omega) - 1$$

is always nonnegative. So,

$$E(D(f_{1,T}, f_{2,T}; x)|\Pi_1) \to C \geq 0. \tag{3-16}$$

as $T \to \infty$ By (3-16) and Lemma 3.5, the result follows immediately. ∎

In practice, the true spectra, $f_{1,T}$ and $f_{2,T}$ are unknown. I replace these quantities in $D(f_{1,T}, f_{2,T}; x)$ 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 $l = 1, 2$. Denote the periodograms computed from the m-th time series in the training dataset from class Π_l to be $\hat{\alpha}_{i,k}^{l,m}$. Then the estimate of $f_{l,T}(u_i, \omega_k)$ is

$$\hat{f}_{l,T}(u_i, \omega_k) = \frac{1}{N_l} \sum_{m=1}^{N_l} \hat{\alpha}_{i,k}^{l,m} \qquad l = 1, 2. \tag{3-17}$$

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

$$\tilde{f}_{l,T}(u_i, \omega_k) = \sum_{s=-M}^{M} W_s \hat{f}_{l,T}(u_i, \omega_{k-s})$$

where $2M + 1$ is the span of the weights; $W_s \geq 0$ and $\sum_{s=-M}^{M} W_s = 1$. The estimate $\tilde{f}_{2,T}$ can be obtained in a similar manner if N_2 is small.

Corollary 3.1 *Let $\hat{f}_{l,T}(u_i, \omega_k)$, for $l = 1, 2$ be the estimates given in (3-17), and let $N = \min\{N_1, N_2\}$. Then, under the established notation and conditions,*

$$\lim_{T\to\infty} \lim_{N\to\infty} P\left\{D(\hat{f}_{1,T}, \hat{f}_{2,T}; x) < 0 | \Pi_1\right\} = 0 \tag{3-18}$$

$$\lim_{T\to\infty} \lim_{N\to\infty} P\left\{D(\hat{f}_{1,T}, \hat{f}_{2,T}; x) \geq 0 | \Pi_2\right\} = 0 \tag{3-19}$$

PROOF.

By the SLLN, $D(\hat{f}_{1,T}, \hat{f}_{2,T}; x) \to_{as} D(f_{1,T}, f_{2,T}; x)$ as $N \to \infty$, and the result follows immediately from Theorem 3.2.

3.4 The Algorithm of the Discriminant Scheme

After discussing the details of the proposed method, I will combine all the steps as an algorithm in this section. I call this algorithm the SLEX Nonstationary Discrimination Algorithm. With this algorithm, the entire picture of how to fulfill the proposed method will be clear.

Let $A_{j,b}$ represent the desired SLEX best basis, $B_{j,b}$ denote a set of basis vectors at level j and block b of the SLEX bases library, $\mathbf{I}(\cdot, \cdot)$ [see (2-2)] denotes the Kullback-Leibler divergence, $f_{1,T}$ and $f_{2,T}$ be the spectrum of the category Π_1 and Π_2, and $D(f_{1,T}, f_{2,T}; x)$ is the discriminant statistic in Definition 3.1. Then, the algorithm is as follows:

SLEX Nonstationary Discrimination Algorithm :

Suppose the training dataset consists of N_1 time series from Π_1 and N_2 time series from Π_2. Let $\{\mathbf{x}_i^{(1)}\}_{i=1}^{N_1}$ and $\{\mathbf{x}_i^{(2)}\}_{i=1}^{N_2}$ denote the training datatsets for Π_1 and Π_2 respectively.

Step 0: Specify the maximum depth of decomposition J.

Step 1: Construct *time-frequency energy maps* Γ_1^S and Γ_2^S [see (3-1) for definition of Γ_1^S and Γ_2^S].

Step 2: Set $A_{J,b} = B_{J,b}$ and $\triangle_{J,b} = \mathbf{I}(\Gamma_1^S(j,b,\cdot),\Gamma_2^S(j,b,\cdot))$ for $b = 0,\ldots,2^J - 1$.

Step 3: Determine the $A_{j,k}$ for $j = J-1,\ldots,0$, $b = 0,\ldots,2^J - 1$ by the following rule:

Set $\triangle_{j,b} = \mathbf{I}(\Gamma_1^S(j,b,\cdot),\Gamma_2^S(j,b,\cdot))$.

If $\triangle_{j,b} \geq \triangle_{j+1,2b} + \triangle_{j+1,2b+1}$,

then $A_{j,b} = B_{j,b}$,

else $A_{j,b} = A_{j+1,2b} \oplus A_{j+1,2b+1}$ and set $\triangle_{j,b} = \triangle_{j+1,2b} + \triangle_{j+1,2b+1}$.

Step 4: Calculate the SLEX periodogram of all $\{\mathbf{x}_i^{(1)}\}_{i=1}^{N_1}$ and all $\{\mathbf{x}_i^{(2)}\}_{i=1}^{N_2}$.

Step 5: Estimate $f_{1,T}$ and $f_{2,T}$ by averaging the SLEX periodogram over each category of the training dataset.

Step 6: Calculate the SLEX periodogram of data X_t.

Step 7 Classify X_t to Π_1 if $D(f_{1,T}, f_{2,T}; x) \geq 0$.

Remarks:

(1.) Based on practical experience, $J = \log_2 T - \log_2(256) + 1$ is an ideal initial choice for the maximum depth of decomposition J in Step 0, that is, there are 128 observations in the blocks of level J. Users can then make further adjustments of J according to the data at hand.

(2.) The selection procedure in Step 3 is fast [O(T)]. **(3.)** The selected basis maximizes the discriminant measure among all the bases in the SLEX bases library. For details, refer to Saito (1994).

(3.) Steps 0, 1, 2, and 3 are so-called the feature extraction step. Steps 4, 5, 6, and 7 are so-called classification step.

Chapter 4

Simulation Results and Data Analysis

To evaluate the proposed method, I performed three simulation studies and one analysis of a real seismic dateset constructed by Blandford (1993). Throughout this section, I restricted the number of categories to two. In the first simulation study, I evaluated my method on some stationary data. In the second simulation study, I applied the discriminant method to some piecewise locally stationary processes. A zero-mean stochastic process is said to be piecewise locally stationary if it is locally stationary at all time points, except possibly at finitely many jump points (see Adak, 1998, and Ombao et al, 2001, for details). For example, a process Y_t defined by

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

where Y_t^1 is white noise and Y_t^2 is AR(1) with coefficient 0.1 is a piecewise locally stationary process. The performance of the proposed discriminant method were tested in various situations. In the third simulation study, I evaluated our method on the data with time-varying spectrum. Two slowly changing AR(2) processes were used in this simulation study.

To simplify the depiction of the results, I define the following: Let T denote the length of process, let n denote the number of training data sets in each group, let

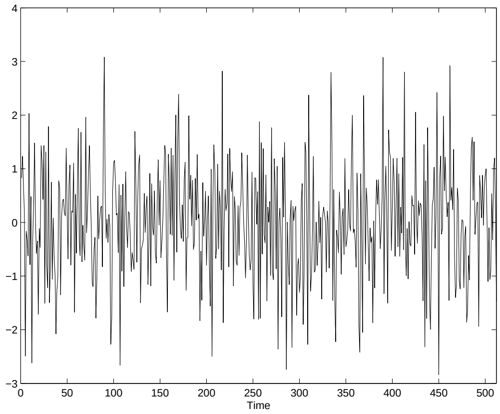

Figure 4.1 White Noise

J denote the level of the SLEX transform, $J = \log_2 T - \log_2(256) + 1$, let m denote the total number of time series datasets used in evaluating the error rate, and let $\mathrm{ER}(\phi)$ denote the error rate corresponding to AR(1) with coefficient ϕ.

4.1 Simulation on Stationary Processes

In this section, I applied the proposed method to the data set with two different categories of stationary processes. One category is white noise and the other category is AR(1) with coefficient ϕ. I ran separate simulations for different AR coefficients, $phi = -0.5, -0.3, -0.1, 0.1, 0.3, 0.5$. I generated n values of the white noise sequences and n values of an AR(1) as the training data set. From the training data set, I obtained the best basis and the estimated spectrum for each category. Then, I generated m values of a new white noise and EN values of a new AR(1) for evaluating the error rate. Plots of simulated white noise and AR(1) with coefficient 0.1 are shown in Figure 4.1 and Figure 4.2.

The simulation results are shown in Table 4.1. We can see that the proposed method can easily discriminate between the white noise and the AR(1) with coefficient $\phi = -0.5, -0.3, -0.3, 0.5$. There were some errors when the method was

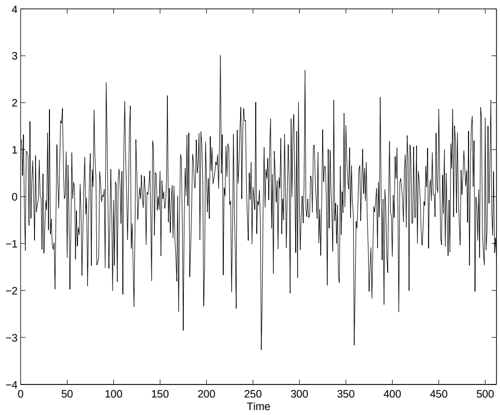

Figure 4.2 AR(1) with coefficient 0.1

Table 4.1 Simulation results for stationary data

T	n	m	ER(−.5)	ER(−.3)	ER(−.1)	ER(.1)	ER(.3)	ER(.5)
512	8	50	0%	0%	8%	6%	0%	0%
512	16	50	0%	0%	6%	6%	0%	0%
512	32	50	0%	0%	4%	6%	0%	0%
1024	8	50	0%	0%	2%	2%	0%	0%
1024	16	50	0%	0%	2%	2%	0%	0%
1024	32	50	0%	0%	0%	0%	0%	0%
2048	8	50	0%	0%	0%	0%	0%	0%
2048	16	50	0%	0%	0%	0%	0%	0%

applied to distinguish the white noise and AR(1) with coefficient 0.1 and −0.1. This result is not unreasonable since the spectrum of a white noise process and AR(1) with ϕ close to zero are very similar. However, the error rate decreased when the T (the size of the data sets) or n (the number of training data sets) increased. When $T = 1024$ with $n = 32$ or when $T \geq 2048$, the error rates were zero.

4.2 Simulation on Piecewise Stationary Processes

Next, I applied the method to the data set with two different categories of piecewise stationary processes. One category, say Π_1, is composed of processes defined by

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

where Y_t^1 is white noise and Y_t^2 is AR(1) with coefficient 0.1. The other category, say Π_2, is composed of processes defined by

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

where X_t^1 is white noise and X_t^2 is AR(1) with coefficient $\phi = -0.5, -0.3, -0.1, 0.3, 0.5$. I generated n time series from each category as the training data set. From the training data set, I obtained the best basis and the estimated spectrum for each category. Then, I generated m new data for each category to evaluate the error rate. A simulated process of the category with white noise and AR(1) [coefficient 0.1] is shown in Figure 4.3. A simulated process from the category with white noise and AR(1) [coefficient 0.3] is shown in Figure 4.5. Figure 4.4 and Figure 4.6 are their corresponding SLEX periodograms.

The simulation results are shown in the Table 4.2. We can see that the proposed method can easily tell the difference between these two categories when the AR(1) in the Π_2 with coefficient $\phi = -0.5, -0.3, 0.5$. There were some errors when $\phi = -0.1$ and 0.1 in the Π_2 . But the error rate decreased when the T (the size of the data sets) or n (the number of training data sets) increased. When $T = 1024$ with $n = 16$ or when $T \geq 2048$, the error rates were zero.

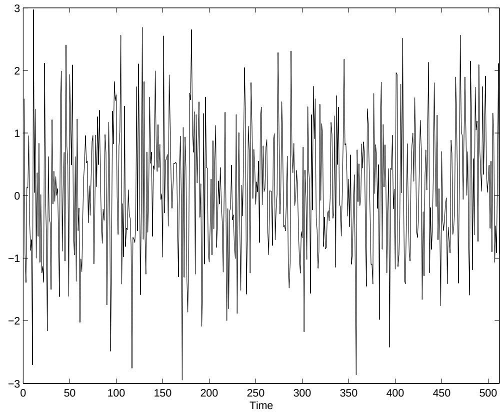

Figure 4.3 White Noise and AR(1) with coefficient 0.1

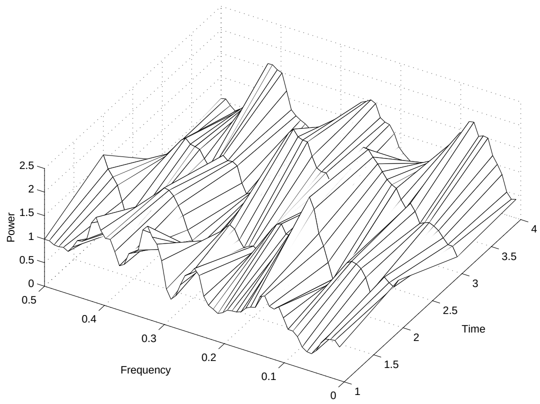

Figure 4.4 A SLEX periodogram of the White noise and AR(1) with coefficient 0.1

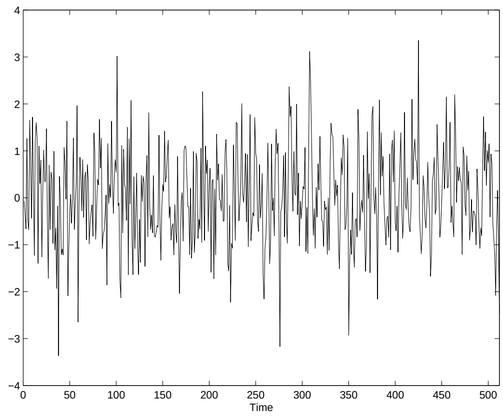

Figure 4.5 White Noise and AR(1) with coefficient 0.3

4.3 Simulation on Slowly Varying AR(2) Processes

In the final simulation, I applied the proposed method to the data set with two different categories of slowly varying AR(2) processes. (The slowly varying AR(2)

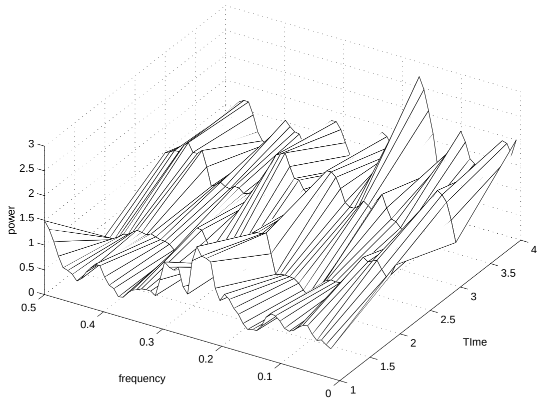

Figure 4.6 A SLEX periodogram of the White noise and AR(1) with coefficient 0.3

Table 4.2 The Simulation Results for Piecewise Stationary Data

T	n	m	ER($-.5$)	ER($-.3$)	ER($-.1$)	ER(.3)	ER(.5)
512	8	50	0%	0%	6%	4%	0%
512	16	50	0%	0%	4%	4%	0%
1024	8	50	0%	0%	4%	2%	0%
1024	16	50	0%	0%	0%	0%	0%
2048	8	50	0%	0%	0%	0%	0%
2048	16	50	0%	0%	0%	0%	0%

processes used in this simulation are from Ombao, Raz, von Sachs and Malow, 2001.) The first category is $Y_t = a_t Y_{t-1} - 0.81 Y_{t-2} + \xi_t$, $(t = 0, \ldots, 1023)$ where $a_t = 0.8[1 - \Delta \cos(\pi t/1024)]$, $\Delta = 0.5$; and ξ_t are iid standard normal. The second category is $Y_t = a_t Y_{t-1} - 0.81 Y_{t-2} + \xi_t$, $(t = 0, \ldots, 1023)$ where $a_t = 0.8[1 - \Delta \cos(\pi t/1024)]$, $\Delta = 0.4, 0.3, 0.2$. The plot of a_t is shown in Figure 4.7.

I generated 10 data sets from each category as the training data. From the training data set, I can get the best basis and the estimated spectrum for each

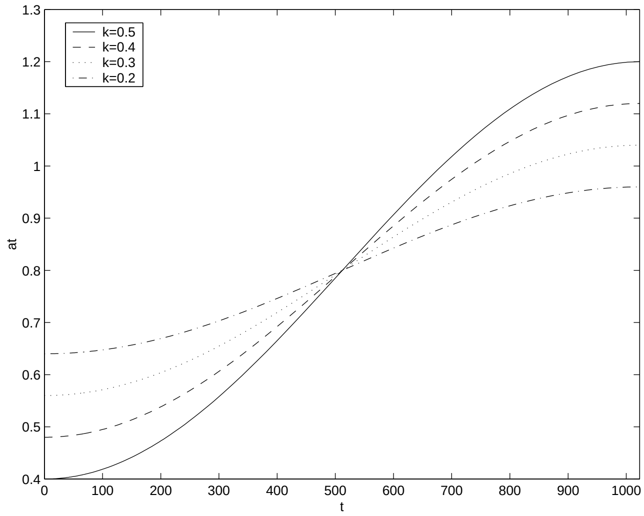

Figure 4.7 The coefficient $a(t)$ when Δ=0.5,0.4,0.3, and 0.2

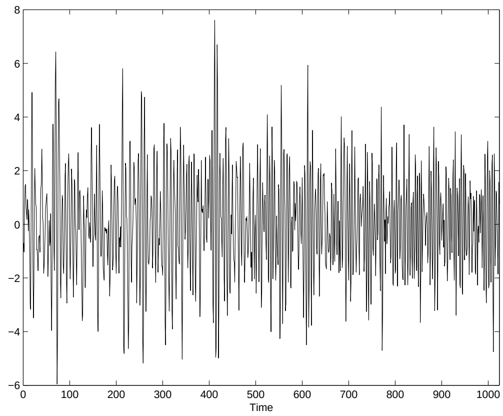

Figure 4.8 A simulated slowing changing AR(2) dataset from the first category

category. Then, I generated 10 new data for each category to evaluate the error rate. A simulated data set from the first category is shown in Figure 4.8. A simulated data set from the second category with $a_t = 0.8[1 - 0.3\cos(\pi t/1024)]$ is shown in Figure 4.9. Figure 4.10 and Figure 4.11 are their corresponding SLEX periodograms.

The simulation results are shown in Table 4.3. We can see that the error rates are zero when Δ in the second category equals to 3 and 2. Note that k is just one of the components in a_t, and a_t is just one of the coefficients of AR(2).

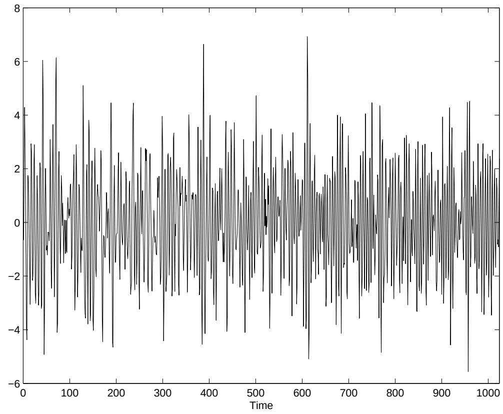

Figure 4.9 A simulated slowing changing AR(2) dataset from the second category

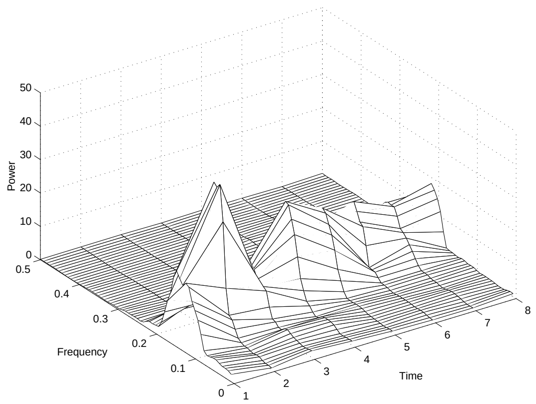

Figure 4.10 A Time-Varying SLEX periodogram of the first category

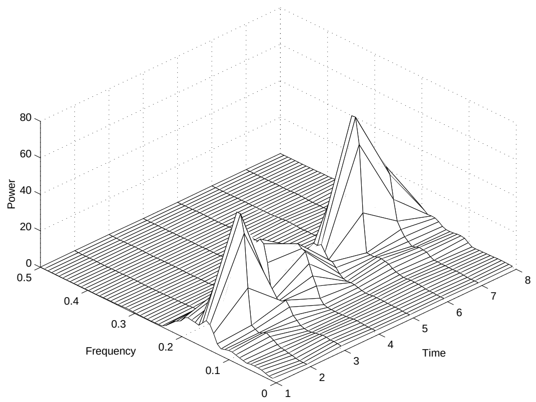

Figure 4.11 A Time-Varying SLEX periodogram of the second category

4.4 Data Analysis

Discriminating between nuclear explosions and earthquakes is a problem of critical importance for monitoring a comprehensive test ban treaty. A data set con-

Table 4.3 The simulation results for slowly varying AR(2)

Δ	Error rate
0.4	8%
0.3	0%
0.2	0%

structed by Blandford (1993) which are regional (100-2000 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 NZ event) are given in Kakizawa et al (1998). A typical earthquake, a typical explosion, and the unknown NZ event are shown in Figure 4.12. Figure 4.13, Figure 4.14, and Figure 4.15 displayed the corresponding SLEX periodograms.

To evaluate our method, I used the holdout procedure, that is, I used all the data at hand as the training data set (not including the NZ event) except the one which I was going to classify. The maximum depth J used in this analysis was $\log_2 T - \log_2(256) + 2$ where $T = 2048$. I smoothed the periodogram of the training dataset with a span of 2 points. That is, I averaged every point with its following point by a filter (see the program in Appendix I). I also smoothed the periodogram of the held out time series with a span of 26 points. Note that different span may result in different discriminant result. Let f_T and g_T denote the respective spectrum of earthquake and explosion. The result is given in Table 4.4. Based on the result in the table, there is no misclassification. The value of the test statistic $D(f_T, g_T; x)$ is -3.1552 for the unknown NZ event. That is, the NZ event is classified as an

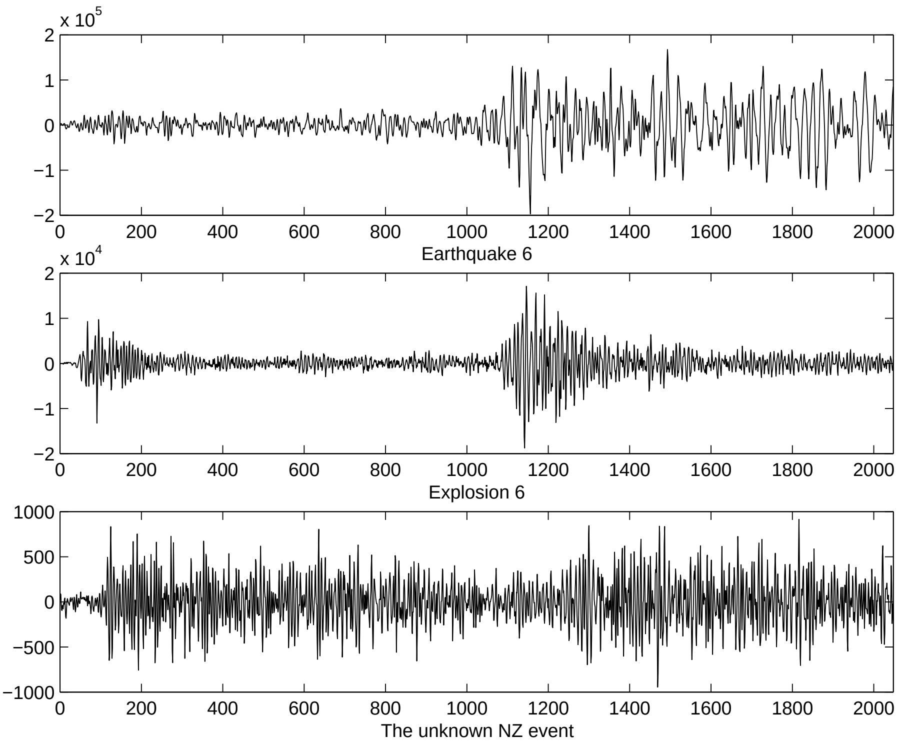

Figure 4.12 The top panel is a typical earthquake. The middle panel is a typical explosion. The bottom panel is the unknown NZ event

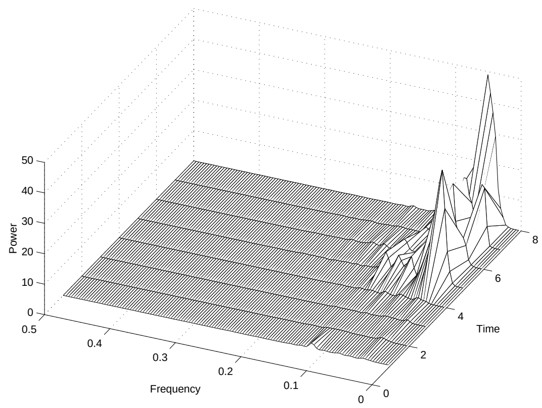

Figure 4.13 The SLEX periodogram of the earthquake 6

explosion which agrees with the result in Kakizawa et al (1998).

The matlab programs used in this data analysis are attached in the Appendixes.

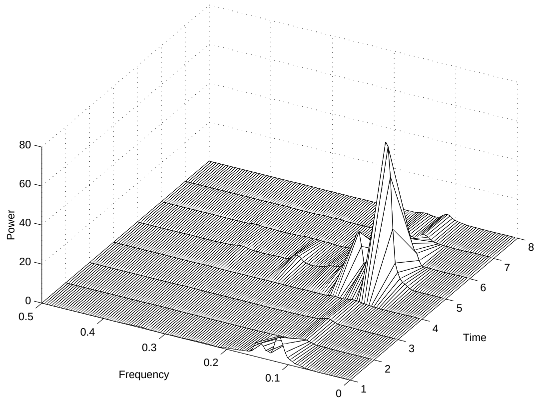

Figure 4.14 The SLEX periodogram of the explosion 6

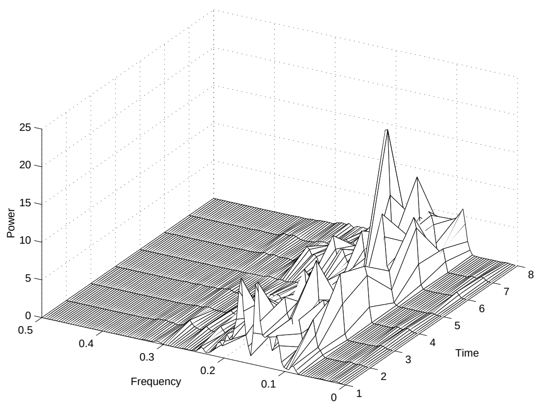

Figure 4.15 The SLEX periodogram of the NZ event

Table 4.4 The results for real data analysis

Holdout process	$D(f_T, g_T; x)$ value	Holdout process	$D(f_T, g_T; x)$ value
Eq1	0.1526	Exp1	-0.5096
Eq2	0.6271	Exp2	-18.6391
Eq3	0.8132	Exp3	-1.4676
Eq4	0.0131	Exp4	-17.5373
Eq5	0.5283	Exp5	-12.5065
Eq6	0.7189	Exp6	-10.3111
Eq7	0.6116	Exp7	-2.1453
Eq8	0.6469	Exp8	-0.1044

Chapter 5

Conclusion and Future Work

In this study, I proposed a method for the discrimination of the nonstationary processes that is based on the SLEX transform. The SLEX transform is localized in both time and frequency domains thus the proposed method is able to extract local features of the data. Moreover, the proposed method is computationally efficient and hence is able to handle large datasets. Using the Kullback-Leibler distance of the *time-frequency energy maps* (Saito, 1994), my procedure selects the basis from the SLEX library that can best discriminate between classes of nonstationary time series. My classification rule, which is based on the Kullback-Leibler distance between the estimated spectra or the groups, is shown to be consistent, i.e., 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.

As for future work, I can broaden the method to multivariate nonstationary time series. The SLEX transform is complex-valued, and hence the SLEX model will allow a natural way to model the time-lag and inter-relationships between components of multivariate time series. Based on these properties of the SLEX model, the extension of the method to the multivariate situation is of great interest.

I have developed the divergence measure between two categories of the SLEX model in Theorem 3.1. Actually, this divergence measure can also be considered as the divergence between two parametric models. Based on this idea, I could develop a method for fitting parametric models to nonstationary processes in the future.

APPENDIX A

APPENDIX A The Main Matlab Progrom for Section 4.4

```
function result1=classifyeqexp01(tnum)
%This file is the main program
%for operating the holdout procedure
%Input tnum, the holdout event
%Output the value of the discriminant statistic
targetdata=tnum;
datanum=17;%There are 17 events in the dataset. 1-8 are earthquakes,
%9-16 are explosions,
%and the 17-th is the unknown NZ event.
T=2048;%The length of the data
%The finest level of the SLEX bases library.
J = nextpow2(T)-nextpow2(256)+2;
e = T/2^(J+1); %The overlap of the SLEX bases function.

%get the training dataset from the eq+exp.dat.
%there are 8 eqs and 8 exps and 1 NZ in eq+exp.dat.
fid=fopen('c:\project\eq+exp.dat');
data=fscanf(fid,'%g',[T,datanum]);
fclose(fid);

%Standardize the data
for i=1:datanum,
    stddata(:,i)=(data(:,i)-mean(data(:,i)))./std(data(:,i));
end
```

```
eq=stddata(:,(1:8));%The category of earthquake
exp=stddata(:,(9:16));%The category of explosion

%Cut one of the data outof the training dataset
%(The holdout procedure)
if targetdata <= 8,
eq(:,targetdata)=[];
else
exp(:,targetdata-8)=[];
end
data1=eq;
data2=exp;

nct1=size(data1);
%The total number of data in the earthquake category
nc1=nct1(1,2);
nct2=size(data2);
%The total number of data in the explosion category
nc2=nct2(1,2);

%Get the relative entropy
c_entropy=classify_entropy06(T,J,e,nc1,nc2,data1,data2);

%get the cost for each level, bolck and frequency
c_c=classify_cost(T,J,c_entropy);
% a matrix with J+1raws and 2.^J column
```

```
%By the Local Discriminant Basis Selection Algorithm
%to get the best basis
O = classify_bba(c_c);
% a matrix with J+1raws and 2.^J column

%Get the estimated spectrum of each category
[s_priodo1,s_priodo2]=sl_priodogram6(T,J,e,nc1,nc2,data1,data2,O);
%The estimated spectrum of earthquake category
fhead=(sum(s_priodo1'))'./nc1;
%The estimated spectrum of explosion category
ghead=(sum(s_priodo2'))'./nc2;

nct=1;%The number of target data (It can be more than one)
%Get the data which had been hold out
dataT=stddata(:,targetdata);
%The SLEX periodogram of the hold out data
s_priodo11=sl_priodogram6s(T,J,e,nct,dataT,O);

%The classification step.
%Substitute the feature into the discriminant statistic
for i=1:nct,
Dstat11(i,1)=D_stat1(T,fhead,ghead,s_priodo11(:,i));
end

result1=Dstat11%The value of the discriminant statistic.
%(The target data is classified to earthquake if result1 > 0)
```

APPENDIX B

APPENDIX B The Matlab Subroutine for Calculating the Entropy

```
function classify_entropy=classify_entropy
(T,J,e,nc1,nc2,data1,data2)
p=5;
% calculate the time-frequency energy map
%(Naoki, S. (1994),p67) of each class
nominator1=zeros(T,J+1);
nominator2=zeros(T,J+1);
denominator1=0;
denominator2=0;

for i=1:nc1;
   for j = 0:J;
 blksz = T/2^j;
 D1 = slexgram(data1(:,i),[],blksz,e);
 %SLEX coefficients of x1 for all blocks and levels
     raw1(:,j+1) =smoothwin1(D1(:),p);
     raw1t(j+1)=norm(abs(raw1(:,j+1))).^2;
   end
   nominator1=nominator1+abs(raw1).^2;
   denominator1=denominator1+raw1t;
end

for i=1:nc2;
```

```
   for j = 0:J;
 blksz = T/2^j;
     D2 = slexgram(data2(:,i),[],blksz,e);
     %SLEX coefficients of x2 for all blocks and levels
     raw2(:,j+1) = smoothwin1(D2(:));
     raw2t(j+1)=norm(abs(raw2(:,j+1))).^2;
   end
   nominator2=nominator2+abs(raw2).^2;
   denominator2=denominator2+raw2t;
end


for j=0:J
%the time-frequency energy map of class 1
    gamma1(:,j+1)=nominator1(:,j+1)./denominator1(j+1);
%the time-frequency energy map of class 2
    gamma2(:,j+1)=nominator2(:,j+1)./denominator2(j+1);
end

%Calculate the cost
classify_entropy=gamma1.*log(gamma1./gamma2)
+gamma2.*log(gamma2./gamma1);%(the J(p,q) in Naoki)
```

APPENDIX C

APPENDIX C The Matlab Subroutine for Calculating the Cost

```
function classify_c=classify_cost(T,J,c_entropy)


classify_c1 = zeros(J+1,2^J); %matrix to hold costs


for j = 0:J,
   blksz = T/2^j; %block size for this level
   for blk = 1:2^j,
      ind1 = (blk-1)*blksz+1;
      ind2 = blk*blksz;
      tmp = c_entropy(ind1:ind2,j+1);
 % get the KL distance for each levels and blocks
      classify_c1(j+1,blk) = sum(tmp);
   end
end


classify_c=classify_c1;
```

APPENDIX D

APPENDIX D The Matlab Subroutine for Getting the Best Basis

```
function seg = BBA(C,h)

%BBA - Best Basis Algorithm
% seg = BBA(Costs)
%
%  C - cost matrix
%  seg - segmentation matrix
%
%  reference: Coifman & Wickerhauser,
% "Entropy based algorithms for best basis selection",
%IEEE Transactions on Information Theory, 32: 712-718,
% 1992

J = size(C,1)-1; %maximum level
seg = zeros(size(C)); %mark the segments
seg(size(C,1),:) = 1;

%start with second highest level of segmentation
for j = J:-1:1,
   numblks = 2^j; %number of blocks in level
   for k = 1:2:numblks,
    %compute cost of children
      cc = C(j+1,k)+C(j+1,k+1);
```

```
      %the cost is less than the children
      if cc > C(j,(k+1)/2),
      %change block cost if children are bigger
         C(j,(k+1)/2) = cc;
      else
         seg(j,(k+1)/2) = 1;
    %if block cost bigger than children, mark
    %unmark all children and grandchildren of this block
         l=j; kk=[k k+1];
         while l<=J,
            seg(l+1,kk)=0;
            l = l+1;
            kk = [kk*2-1 kk*2];
         end
      end
   end
   if nargin > 1, wait((J-j+1)/J,h); end
end
```

APPENDIX E

APPENDIX E The Matlab Subroutine for Getting the SLEX Periodogram

```
function [s_priodo1,s_priodo2]=
sl_priodogram(T,J,e,nc1,nc2,data1,data2,O)
%s_amount=round(T.*percent);
p=5;
priodogram1=zeros(T,nc1);
priodogram2=zeros(T,nc2);
for j=0:J,
    blksz=T./2^j;
    for blk=1:2^j,
      ind1 = (blk-1)*blksz+1;
      ind2 = blk*blksz;
        if O(j+1,blk)==1,
            for k1=1:nc1
                  x1=data1(:,k1);
                      slexc1=slexgram(x1,[],blksz,e);
                      perod1=abs(slexc1(:,blk)).^2;
      priodogram1(ind1:ind2,k1)=smoothwin1(perod1,p);
            end

            for k2=1:nc2
                  x2=data2(:,k2);
                    slexc2=slexgram(x2,[],blksz,e);
                      perod2=abs(slexc2(:,blk)).^2;
```

```
      priodogram2(ind1:ind2,k2)=smoothwin1(perod2,p);
            end
        end
    end
end

s_priodo1=priodogram1;
s_priodo2=priodogram2;
%[temp,index1]=sort(s_entropy);
%index2=index1(T-s_amount+1:T);
%index=sort(index2);

%i_length=length(index);
%for h=1:i_length,
%     s_priodo1(h,:)=priodogram1(index(h),:);
%     s_priodo2(h,:)=priodogram2(index(h),:);
%end
```

APPENDIX F

APPENDIX F The Matlab Subroutine for Getting the SLEX Periodogram of the Target Data

```
function s_priodo1=sl_priodogram(T,J,e,nct,data1,O)
%s_amount=round(T.*percent);
p=5;
priodogram1=zeros(T,nct);
for j=0:J,
    blksz=T./2^j;
    for blk=1:2^j,
      ind1 = (blk-1)*blksz+1;
      ind2 = blk*blksz;
        if O(j+1,blk)==1,
            for k=1:nct
               x1=data1(:,k);
              slexc=slexgram(x1,[],blksz,e);
              perod=abs(slexc(:,blk)).^2;
          priodogram1(ind1:ind2,k)=smoothwin1(perod,p);
            end


        end
    end
end


s_priodo1=priodogram1;
```

APPENDIX G

APPENDIX G The Matlab Subroutine for Calculating the Discriminant Statistic

```
function D_stat=D_stat(T,fhead,ghead,IN)
L=length(fhead);
sum_D=0;
for i=1:L,
    sum_D=sum_D+log(abs(ghead(i)./fhead(i)))
    +IN(i)*(inv(ghead(i))-inv(fhead(i)));
end
D_stat=sum_D./T;
```

APPENDIX H

APPENDIX H The Matlab Subroutine for Calculating the SLEX Transform

```
function [S,F,T] = slexgram(x,fs,blksz,n_ov)%(Ombao et al, 2000)

%SLEXGRAM - SLEX spectrogram
%  [S,F,T] = slexgram(x,Fs,blksize,n_overlap)
%     x - input data vector
%     fs - sampling frequency
%     blksize - x is broken into blocks of size blksize
%      for fft (default = 256)
%     n_overlap - number of overlapping points
%     (default = blksz/2)
%
%   The outputs depend on the number of output arguments:
%   if nargout == 1,
%     S = SLEX spectra,
%i.e. S(jw) for all frequencies positive and negative
%   else
%     S = SLEX periodograms
%i.e. |S(jw)|.^2 of the positive frequencies only
%     F - frequency vector
%     T - time vector
%
%     No periodogram smoothing is done in this function.
%  For smoothing,
```

```
%     SMOOTHWIN should be called.
%     If there are no output arguments specified,
% SLEXGRAM will generate
%     a pseudocolor plot of the SLEX periodograms
% logarithmically scaled.
%     20*log10(|S(f)|.^2) vs. F and T
%
%  See also: SLEX, SLEX_MULTI, SLEX_EEG,
% SMOOTHWIN, SPECGRAM
%  Created: 11/15/00 SDC; Modified 11/17/00 SDC

warning backtrace
x = x - mean(x);
if nargin < 2 | isempty(fs),  fs = 2; end
if nargin < 3 | isempty(blksz), blksz = 256; end
if nargin < 4, n_ov = blksz/2; end

nx = length(x);
x = [zeros(n_ov,1); x; zeros(n_ov,1)];
[Bp,Bm] = bp(0,blksz,n_ov,blksz+2*n_ov+1,1);
 %compute the taper function
nw = length(Bp);
ncol = fix(nx/blksz);
segsz = 2^nextpow2(blksz+2*n_ov);
z_pad = zeros((segsz-length(Bp)),ncol);

l = 1:ncol;
```

```
col_ind = (l-1)*blksz+1;
row_ind = (0:(blksz + 2*n_ov -1))';
y = x(col_ind(ones(nw,1),:) + row_ind(:,ones(1,ncol)));
yp = Bp(:,ones(1,ncol)).*y;
ym = Bm(:,ones(1,ncol)).*y;

yp = [yp; z_pad];
ym = [ym; z_pad];
yp = fft(yp);
ym = fft(ym);


S = (yp + conj(ym))/sqrt(blksz);

S = S(1:2:end,:);
 %decimate the coefficients to make it length blksz


if nargout ~= 1,
   S = (abs(S)+eps).^2;
   S = S(1:blksz/2,:);
T = linspace(0,nx/fs,ncol)/60;
F = linspace(0,fs/2,blksz/2);

if nargout == 0,
figure;
imagesc(T,F,20*log10(S));
```

```
axis([min(T) max(T) 0 max(F)]);
axis xy;
colormap((jet(256)));
xlabel('Time (min)');
ylabel('Frequency');
end
end
```

APPENDIX I

APPENDIX I The Matlab Subroutine for Smoothing the SLEX Transform

```
function SX_new = smoothwin(X,p,sym)

%SMOOTHWIN - Daniel smoothing window
%  SX = smoothwin(X,support,symmetry)
%  SX - smoothed X
%  X - input data sequence
%  support - smoothing window span,
%   an integer corresponding to the number of
%            points in the window
%  symmetry - 'odd' or 'even' for odd or even functions

if (nargin < 2 | p == 0), p = auto_smo_par(X, 5, 20);
end
if nargin < 3, sym = 'even'; end
L = length(X);

%create Daniel window
W = (1/(p-1))*ones(p,1);
W(1) = W(1)/2;
W(p) = W(p)/2;

if strncmp(sym,'e',1),
   XX = [X(L/2+1:L); X; X(1:L/2)];
```

```
elseif strncmp(sym,'o',1),
   XX = [-X(L/2+1:L); X; -X(1:L/2)];
end

SX = fftfilt(W,XX);

%compensate for the shift caused by the filter
N2 = floor(p/2);
%SX = SX(L/2+N2:L/2+N2+L-1);
SX_new = SX(L/2+N2+1:L/2+N2+L);
```

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Adak, S. (1998). Time dependent spectral analysis of non-stationary time series. *J. Am. Stat. Asso.,* **93**, 1488-1501.

[2] Alagón, J. (1989). Spectral discrimination for two groups of time series. *J. Time Series Anal.,* **10**, 203-214.

[3] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd ed.*. Wiley, NY.

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

[5] Coifman, R. and Wickerhauser, M. (1991). Entropy based algorithms for best basis selection. *IEEE Transactions on Information Theory,* **32**, 712–718.

[6] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics,* **25**, 1–37.

[7] Dargahi-Noubary, G. R. and Laycock, P. J. (1981). Spectral ratio discriminants and information theory. *J. Time Series Anal.,* **16**, 201-219.

[8] Daubechies, I. (1992). *Ten Lectures on Wavelets.* Society for Applied and Industrial Mathematics, Philadelphia, PA.

[9] Donoho, D., Mallat, S. and von Sachs, R. (1998). Estimating covariances of locally stationary processes: rates of convergence of best basis methods. *Technical Report 517, Statistics Department, Stanford University.*

[10] Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis, 3rd ed.*. Prentice-Hall, Englewood Cliffs, NJ.

[11] 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.

[12] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics,* **22**, 79-86.

[13] Miao, G. J. and Clements, M. A. (2002). *Digital Signal Processing and Statistical Classification.* Artech House, Boston.

[14] Ombao, H., Raz, J., von Sachs, R. and Malow, B. (2001). Automatic statistical analysis of bivariate non-stationary time series. *Journal of the American Statistical Association,* **96**, 1234-1345.

[15] Ombao, H., Raz, J., von Sachs, R. and Guo, W. (2002). The SLEX model of non-stationary processes. *Annals of the Institute of Statistical Mathematics,* **32**, 201-225.

[16] Priestley, M. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Ser. B,* **28**, 228-240.

[17] Priestley, M. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

[18] Pulli, J. J. (1996). Extracting and processing signal parameters for regional seismic event identification. In *Monitoring a Comprehensive Test Ban Treaty,* pp. 791-803. E. S. Husebye and A. M. Dainty eds. Doordrecht: Kluwer.

[19] Saito, N. (1994). Local Feature Extraction and Its Applications. *Ph.D. dissertation, Department of Mathematics, Yale University.*

[20] Sakiyama, K. and Taniguchi, M. (2001). Discriminant analysis for locally stationary processes. Technical report.

[21] Shumway, R. H. (1982). Discriminant analysis for time series. In *Classification, Pattern Recognition and Reduction of Dimensionality, Handbook of Statistics Vol. 2,* pp. 1-46. P. R. Krishnaiah and L. N. Kanal eds. Amsterdam: North Holland.

[22] Shumway, R. H. and Stoffer, D. S. (2000). *Time series analysis and its applications.* Springer, NY.

[23] Taniguchi, M., Puri, M. L., and Kondo, M. (1994). Nonparametric approach for non-Gaussian vector stationary processes. *J. Mult. Anal.,* **56**, 259-283.

[24] Wickerhauser, M. (1994). *Adapted Wavelet Analysis from Theory to Software.* IEEE Press, Wellesley, MA.