RC23520 (W0502-053) February 8, 2005
Mathematics

IBM Research Report

Online Change-point Detection Methods for Non-stationary Time Series

Hyunyoung Choi, Hernando Ombao
University of Illinois at Urbana-Champaign

Bonnie Ray
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Online Change-point Detection Methods for Nonstationary Time Series

Hyunyoung Choi[1], Hernando Ombao[2], and Bonnie Ray[3]

January 11, 2005

Abstract

We present two new spectral-based methods for real-time detection of changes in autocorrelation structure in a continuous-valued time series. Our methods are built on the idea that changes in the autocorrelation are reflected by changes in the Fourier or wavelet-based spectrum and can be detected by comparing estimated spectra of adjacent blocks of the series. To be effective for slowly changing spectral structure, the methods are extended to allow information from more than one past block to be used in determining if a change has occurred. Through simulation, we evaluate the performance of our methods and find that they can provide reliable and quick detection of changes in covariance structure in an online monitoring framework. We illustrate the methods using electroencephalogram traces (brain waves) and run-time computer performance metrics.

Keywords: Change point, Maximal Overlap Discrete Wavelet Transform (MODWT), Online Monitoring, Periodogram, Scalogram.

[1]H. Choi is Ph.D. student at the Department of Statistics, University of Illinois at Urbana-Champaign. Email: hchoi9@uiuc.edu.

[2]H. Ombao is Assistant Professor, Department of Statistics and Department of Psychology, University of Illinois at Urbana-Champaign. Email: ombao@uiuc.edu. His research is supported, in part, by NSF DMS 0405243.

[3]Bonnie Ray is Research Staff Member, Mathematical Sciences Department, IBM T.J. Watson Research Center. Email: bonnier@us.ibm.com

1 Introduction

In monitoring a sequence of correlated observations over time, a key issue is the quick detection of any change in the underlying process structure, while controlling the rate of false alarms at a fixed level. Quick detection of change(s) enables immediate decision making and dynamic control of a system because it allows users to react quickly and make necessary adjustments in a timely manner.

Much prior work on this problem has focused on methods for detecting changes in the mean and/or variance structure of a time series, most often using a time domain approach. For example, Inclan (1993) and Inclan and Tiao (1994) discuss time domain methods for detecting multiple changes in the variance of a time series, with application to series of stock returns. Online methods for detecting changes in a process have received less attention. For retrospective change point detection, Kluppelberg and Mikosch (1996) present a test based on the cumulative integrated periodogram. Ombao, Heo and Stoffer (2004) present a parametric approach to online change point detection for time series, based on fitting autoregressive (AR) models to blocks of a series and comparing adjacent blocks using likelihood-based methods. Although AR models can capture a wide-range of process behavior, the results of the methodology are likely to be poor if the structure differs significantly from the assumed AR process. Nonparametric methods, such as those based on spectral properties of the process, have not been examined in the online setting.

The main contribution of this paper are the online change point detection methods that are based on the Fourier and wavelet spectra, which are decompositions of variance across frequency and scale, respectively. By partitioning the series into short blocks, so that stationarity within a block can be safely assumed, we develop spectral-based methods for comparing the process behavior across blocks, to determine whether a change has occurred. As mentioned above, an advantage of the spectral-based methods is that they do not assume a particular process structure, such as AR, and thus can be effective for series with non-ARMA spectral structure. For improved power in detecting slowly changing second-order properties, we also show how our procedure can be extended to use spectral information from more than one previous block without undue additional computational burden.

The remainder of the paper is organized as follows: In Section 2, we discuss Fourier and wavelet analyses, which are the foundation of our proposed methods, and provide a summary of relevant prior work. In Section 3, we outline our proposed methodology, including a procedure for sequential testing relevant for the online framework, and provide illustrations to illuminate the ideas. Section 4 summarizes the results of a simulation study to compare the power of the various tests under different change scenarios. Section 5 presents applications to monitoring of computer performance metrics and the behavior of earthquake and explosion seismic waves records. Section 6 concludes. Some simulation details are included in an appendix.

2 Statistical Background

2.1 Fourier-based spectral method

Let $\{X_t\}$ be a weakly stationary random process that has an absolutely summable autocovariance function $\gamma(\cdot)$. The spectrum of the process is defined to be $f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp(-i2\pi\omega\tau)$, where $\omega \in (-\frac{1}{2}, \frac{1}{2}]$. The spectrum provides information concerning the second-order properties of the process as represented in the frequency domain. The periodogram, an estimate of the spectrum computed from a stretch of time series $X_1, \ldots, X_T$ (T is assumed to be even, for simplicity) is defined at the Fourier frequencies $\omega_k = \frac{k}{T}, k = -\left(\frac{T}{2} - 1\right), \ldots, \frac{T}{2}$ to be

$$I(\omega_k) = (1/T) \left| \sum_{t=1}^{T} X_t \exp(-i2\pi\omega_k t) \right|^2 .$$

To test whether an observed time series was generated from a process having a specified spectrum $f_0(\omega)$, Jenkins and Watts (1968) introduced the "cumulative periodogram test", based on the observation that the cumulative sum of the ratio of the periodogram value at frequency ω_k relative to the (true) spectral value at frequency ω_k follows a uniform distribution. Coates and Diggle (1986) extended this idea to compare the spectra of two independent time series under the assumption that their underlying second-order properties are the same.

Consider two independent time series $\{X_{g,t} : t = 1, \ldots, T; g = 1, 2\}$ from which we compute the periodograms $I_1(\omega_k)$, $I_2(\omega_k)$, respectively. We concentrate on the periodogram only at the principal Fourier frequencies $\omega_k, k = 1, \ldots, m$ (where $m = \frac{T}{2} - 1$) and consider only the non-negative frequencies, since the spectrum is symmetric around zero. We also ignore the periodogram at $k =$

$0, \frac{T}{2}$, since the periodogram at these indices contains information on the mean level of the process and power at the Nyquist frequency, which are typically not of interest. At the principal Fourier frequencies, it is well known that the ratios $\frac{I_i(\omega_k)}{f_i(\omega_k)}$'s are asymptotically independently and identically distributed observations having distribution $\frac{\chi^2_2}{2}$. See, e.g., Brillinger (1981) and Shumway and Stoffer (2000). Thus the ratios of the periodograms are distributed approximately independently as

$$R_k = \frac{I_1(\omega_k)}{I_2(\omega_k)} \frac{1}{\theta_k} \sim F_{2,2}, \text{ where } \theta_k = \frac{f_1(\omega_k)}{f_2(\omega_k)}. \tag{2.1}$$

The null hypothesis that the two time series have the same spectra can be formulated as $H_0 : \theta_k = 1$ for all k.

Coates and Diggle (1986) proposed nonparametric methods to examine the above hypothesis. Of these, the CUSUM method, motivated by the work of Jenkins and Watts (1968), was shown to have highest power in the situations examined. Under H_0, the set of values $z_{1,k} = \log(1+1/R_k)$ and $z_{2,k} = \log(1+R_k)$ are asymptotically independent random samples from an exponential distribution with mean one. Then $U_j = \sum_{k=1}^{j} z_{g,k} / \sum_{k=1}^{m} z_{g,k},\ g = 1, 2, j = 1, \cdots, m-1$ is the order statistic from a uniform distribution. The CUSUM method uses the Kolmogorov Smirnov (KS) statistic to test for departure of the U_j from a uniform distribution. A limitation of this method is that the achieved significance level depends on the labeling of the two time series, which define $z_{1,k}$ and $z_{2,k}$. To address this issue, Coates and Diggle (1986) apply their test twice, i.e., on each of the ratios R_k and $1/R_k$. They do not discuss adjustment of the level of their test for multiple testing.

If interest is centered only on differences in spectral shape, with no concern about changes purely in the magnitude of the spectra (changes in variance), the normalized periodogram can be used for testing instead, as the sum of the periodogram values is a consistent estimator of the process variance.

In a follow-up paper to Coates and Diggle (1986), Diggle and Fisher (1991) suggested an alternative method for comparing the spectra of two time series, by examining a quantile-quantile plot of the two normalized cumulative sample periodograms or, more formally, using a two-sample KS test with critical values obtained using a randomization method. The latter method, while addressing the issue of arbitrary labeling, is not computationally feasible in a online monitoring framework due to the reliance on randomization for obtaining critical values. One major contribution of this paper

is a method that takes into account ratios R_k and $1/R_k$ simultaneously and hence addresses the arbitrary labeling issue in a natural manner, while lending itself feasible in the online monitoring setting.

2.2 Wavelet-based spectral method

As an alternative to frequency-based decomposition of time series variance, a more recent approach has been to use wavelets to decompose the variance across different time scales. Consider a time series which is a union of disjoint stationary blocks. Percival (1995) outlined the connection between the variance of the stationary block, the spectrum and the wavelet variance via the following:

$$\mathrm{Var}(X_t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega)d\omega = 2\sum_{j=1}^{\infty}\int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} f(\omega)d\omega = \sum_{j=1}^{\infty} v^2(2^j)$$

where the wavelet variance associated with the scale $\lambda_j = 2^j$ is

$$v^2(\lambda_j) \approx 2\int_{1/(2\lambda_j)}^{1/\lambda_j} f(\omega)d\omega.$$

Our goal is to develop a method that can detect a change in the wavelet variance across time blocks.

One may view the wavelet variance as a decomposition of variance across a partitioning of the frequency interval $(0, \frac{1}{2}]$, which is $\bigcup_{j=1}^{\infty}(1/(2\lambda_j), 1/\lambda_j]$. This suggests that a reasonable estimator for $v^2(\lambda_j)$ would involve wavelet coefficients that correspond to the scale λ_j. In this paper, we estimate the wavelet variance $v^2(\lambda_j)$ using the maximum overlap discrete wavelet transform (MODWT), described in Percival and Walden (2000, Chapter 5) as the coefficients of scaled and shifted mother wavelets that share maximal overlap or support in time. In other words, the time shifts used to construct MODWT coefficients are smaller than those used in the discrete wavelet transform (DWT). Percival (1995) demonstrated that the estimator that is based on MODWT has a smaller asymptotic variance than that based on DWT.

We now set some notation. Let $X_t, t = 1, \ldots, T$ be the time series of interest; $\tilde{h}_{j,\tau} = h_{j,\tau}/2^{j/2}$ is the j^{th} level MODWT scaling filter, where $h_{j,\tau}, j = 1, \ldots, J = \log_2(T)$ and $\tau = 0, \ldots, N_j - 1$ be the wavelet basis function corresponding to level j and shift τ; $N_j = (2^j - 1)(L_j - 1) + 1$ is the total number of coefficients at scale j; and L_j is the length of the wavelet filter for level j. The

MODWT wavelet coefficients at level j and shift τ are defined to be

$$\tilde{w}_{j,\tau} = \sum_{\ell=0}^{L_j-1} \tilde{h}_{j,\ell} X_{(t-\tau)_{\text{mod } T}}.$$

The wavelet coefficients at scale λ_j are nominally associated with frequencies in the interval $(\frac{1}{2^{j+1}}, \frac{1}{2^j}]$.

An unbiased estimator of $v^2(\lambda_j)$ based on MODWT coefficients is given by the scalogram

$$\widehat{v^2}(\lambda_j) = \frac{1}{\widetilde{N}_j} \sum_{k=L_j-1}^{N-1} \tilde{w}_{j,k}^2, \tag{2.2}$$

where $\widetilde{N}_j = T - L_j + 1$. The asymptotic distribution of the scalogram is given by

$$\widehat{v^2}(\lambda_j) \sim \frac{v^2(\lambda_j)}{\eta_j} \chi^2_{\eta_j}, \tag{2.3}$$

where η_j is the empirical degrees of freedom (Percival and Walden, 2000). For small sample sizes, setting $\eta_j = \max\{\widetilde{N}_j/2^j, 1\}$ provides reasonable, if slightly optimistic, confidence bounds on the estimated scalogram. We refer the reader to Percival and Walden (2000, Chapter 8) for a detailed derivation of this result. We also note the close parallel to results on the asymptotic didstribution of the (Fourier) periodograms. See, e.g., Shumway and Stoffer (2000).

Scalograms have received wide attention in the statistical literature due to their wide applicability and the solid theoretical foundation provided by Percival and Walden (2000). We highlight some of the applications relevant to this paper. Whitcher, Byers, Guttorp and Percival (2002, Chapter 7) proposed a method for retrospectively testing homogeneity of variance in a long memory time series. Their method detects the time of unknown variance changes using the MODWT. Maharaj (2002) clusters non-stationary time series based on similarity of their scalograms by testing the ratio of the sum of squared wavelet coefficients of the two time series. Determination of similarity is made using randomization tests applied to scalograms, analogous to the work of Diggle and Fisher (1991) with periodograms. However, these methods have not been adapted to the online monitoring setting. Thus, another major contribution of this work is the development of a method for detecting, in an online setting, a change in the wavelet variance structure of the time series, including a comparison of the developed method to that proposed in the Fourier framework.

3 Testing for Change in the Autocorrelation and Autocovariance Structure

For online monitoring of changes in correlation structure of a time series, we propose two different methods. The first is based on ratios of periodograms but overcomes the problem of labeling two time series in Coates and Diggle (1986). The second is a wavelet scalogram-based method. Each of the methods use test statistics whose asymptotic distributions assume that the adjacent blocks are close to being independent. Here, we describe and illustrate the methods, while Section 4 summarizes simulation results showing the efficacy of the two methods for a variety of different change scenarios.

3.1 Gamma Distribution-based Symmetric Ratio method

For detecting changes in the variance decomposition using Fourier-based methods, we propose using a symmetric ratio (SR) statistic, S_k, which is based on the ratio R_k, as defined in (2.1), and its inverse $1/R_k$.

Theorem 3.1. *Define* $S_k = max\ \{\log(1+R_k), \log(1+1/R_k)\} - \log 2$. *Under the assumption that* R_k *is distributed as an* $F(2,2)$ *random variable, it follows that* S_k *is distributed as an Exponential*(1) *random variable.*

Proof. Under the null of no difference in spectral structure, as in (2.1), R_k follows an $F_{2,2}$ distribution, which guarantees $\log(1+R_k)$ and $\log(1+1/R_k)$ each have exponential distributions with mean of one. To derive the distribution function of S_k, we note that $\log(1+R_k) \geq \log(1+1/R_k)$ if and only if $R_k \geq 1$ if and only if $\log(1+R_k) > \log 2$. We now proceed with the proof. First, define $S_k = \max\{\log(1+R_k), \log(1+1/R_k)\} - \log 2$. Next, denote the cdf of S_k to be $F(r) = P(S \leq r)$. It is obvious that when $r \leq 0$, $F(r) = 0$. For $r > 0$

$$\begin{aligned} F_S(r) &= P(\log(1+R_k) \geq \log 2) P(\log(1+R_k) - \log 2 < r | \log(1+R_k) \geq \log 2) + \\ &\quad P(\log(1+1/R_k) \geq \log 2) P(\log(1+1/R_k) - \log 2 < r | \log(1+1/R_k) \geq \log 2) \\ &= P(0 \leq \log(1+R_k) - \log 2 < r) + P(0 < \log(1+1/R_k) - \log 2 < r) \\ &= 1 - \exp(-r). \end{aligned}$$

□

The symmetric ratio statistic is attractive because it is asymptotically distributed as an Exponential(1) random variable. This allows the user to objectively choose a threshold by specifying the probability of type I error of any test.

Using the approximate independence of periodogram ordinates at the Fourier frequencies, the $\{S_k\}$ form a set of $(m-1)$ independent exponential random variables, each with mean one. Then $T_S = \sum_{k=1}^{m-1} S_k$ follows a Gamma distribution with shape parameter $\alpha = m-1$ and scale parameter $\beta = 1$. Large values of T_s indicate departures from H_0. Note that the SR statistic avoids the arbitrary labeling issue. Changes in time series behavior in a *particular* frequency band can be identified by aggregating S_k only over the frequency band of interest, with appropriate degrees of freedom adjustment.

The distinction between the CUSUM and the SR methods can be understood in this way. Differences between the two time series are manifested in different amplitudes at different frequency bands. The CUSUM method derives its power from the aggregation of small deviations of the transformed periodogram values from a uniform distribution. As such, the technique will be most effective for those processes having power concentrated in a particular frequency band. However, as noted above, the power of the test will depend on which series is used to form the numerator of R_k. The CUSUM method will emphasize deviations either from the right- or left-hand left tail behavior of an Exp(1) distribution, depending on this labeling. The use of a CUSUM method to look for deviations of S_k from a uniform distribution will not be as effective; taking the maximum at each k does not preserve the ordering of the deviations, which tends to decrease the departure of the statistic, $U_j = \sum_{i=1}^{j} S_i / \sum_{i=1}^{m} S_i$, from a uniform distribution. Hence, we propose using the distributional properties of $T_S = \sum_{k=1}^{m-1} S_k$ as the basis of our test statistic.

For time series having power spread over different frequency bands, such as for a series with a discrete spectrum, the CUSUM method provides little benefit. By considering the maximum of the transformed ratio and its inverse in the form of the SR statistic, we always capture the maximum deviation from the null hypothesis. By not using a Kolmogorov-Smirnov statistic, we lose the cumulative effect when the power is concentrated in a particular frequency band, but gain

power in situations where the time series has power spread over different frequency bands. The simulation results discussed in Section 4 bear out this behavior.

3.2 Scalogram Method

As an alternative to Fourier-based change-point detection methods, we also consider a method based on the wavelet decomposition of variance components, as captured by the scalogram at each level $j = 1, \ldots, J = \log_2(T)$, where the scalogram is defined in (2.2).

The steps in the scalogram approach are as follows:

1. For each level j, construct scalograms $\widehat{\sigma_1^2}(\lambda_j)$ and $\widehat{\sigma_2^2}(\lambda_j)$ from two independent time series $\{X_{g,t}\}, t = 1, \cdots, T; g = 1, 2$

2. Take the ratio $r(\lambda_j) = \frac{\widehat{\sigma_1^2}(\lambda_j)}{\widehat{\sigma_2^2}(\lambda_j)}$. Then, under the null hypothesis of equal scalograms for each block, $r(\lambda_j)$ follows an F_{η_j,η_j} distribution, where η_j can be approximated as $\eta_j = \max\{\tilde{N}_j/2^j, 1\}$, with $\tilde{N}_j$ as in Section 2.2.

3. For each scale j, use $r(\lambda_j)$ to test for departure from the null of equal scalograms across the two blocks. Note that a two-tailed test is used.

4. Adjust the p-values for multiple comparisons across scales.

Note that each level of the scalogram provides information on the variance components in a different frequency band and at different times. If we are interested in the behavior of the series over all frequency levels, all of the scalograms should be tested simultaneously. A standard Bonferroni method can be used to correct for multiple testing. However, we use the correction method of Benjamini and Hochberg (1995), which is less stringent compared to the Bonferroni method, and therefore tolerates more false positives. In Benjamini and Hochberg's method, the p-value of each test is ranked from largest to smallest. The largest p-value is left unchanged and each successively smaller p-value is multiplied by the total number of tests and divided by its rank, i.e. the adjusted p-value is $p(\frac{n}{n-i}), i = 1, \ldots, n-1$. If the resulting value is less than the significance level of the test, the null of equal autocorrelation structure is rejected. For detecting changes in scalograms, we found that the Benjamini and Hochberg correction method gave a test that was more powerful without an unduly large false positive rate as compared to the method of Bonferroni.

Scalograms are typically not normalized. Thus, unless each block is normalized to have unit variance before the wavelet filter is applied, the scalogram-based change-point detection method will detect changes in variance as well as changes in correlation structure.

3.3 Sequential Test Procedure for Online Change Point Detection

Ideally, information from more than one previous block should be used to compare to the current block, for improved power. However, in an online setting, it is important to avoid unnecessary computation, such as recomputation of the periodogram (scalograms) for the entire series consisting of $(n-1)$ prior blocks for which a difference in autocorrelation structure cannot be rejected. This suggests using a multiple testing approach, where the problem is to determine if the n^{th} time block is different from the previous $(n-1)$ blocks given that all previous $(n-1)$ blocks share common spectrum.

A formulation of the testing problems is as follows: let $H_1, \cdots, H_{n-1}$ be a collection of $(n-1)$ null hypotheses, where H_i is the hypothesis that block $(n-i)$ and block n have the same autocorrelation structure. This formulation constitutes a multiple testing situation and an adjustment is needed to obtain correct significance levels for the test. For this situation, we consider that recent blocks contain more relevant information concerning changes in process structure, and use a Bonferroni correction with exponentially decreasing significance levels to adjust for multiple testing. More specifically, we use $\alpha_i^* = \alpha w_i$ for testing H_i, where $w_i = \frac{1/2^i}{1-1/2^{n-1}}$, such that the $\sum_{i=1}^{n-1} \alpha_i^* = \alpha$. We first test H_{n-1} at significance level α_{n-1}^*. If H_{n-1} is rejected, a change in correlation structure is said to have occurred. Otherwise, we proceed to test H_{n-2} at significance level α_{n-2}^*. Testing continues until a rejection is obtained or all $(n-1)$ tests have been performed, in which case the current block cannot be said to be different from the previous blocks in terms of autocorrelation structure. The sequential testing procedure should have power similar to that of the method based only on comparing most recent block to previous block, as α_{n-1}^* for testing H_{n-1} is only slightly smaller than an unadjusted α value. An advantage of the sequential testing procedure is the potentially enhanced ability to (eventually) detect changes for process with slowly evolving autocorrelation structures, which might be otherwise hard to detect if the differences between successive blocks is small. We explore this capability in more detail in the next section.

3.4 Illustrations

To illustrate the proposed methods, we generated a time series of length $T = 128$ from a piecewise stationary process x_t having change point at $t = 65$, with x_t defined as

$$x_t = \begin{cases} -0.9x_{t-1} + \epsilon_t, & 1 \leq t \leq 64 \\ 0.9x_{t-1} + \epsilon_t, & 65 \leq t \leq 128 \end{cases}$$

Thus, the first block of $n = 64$ consists of a process with power concentrated at high frequencies, while the second block has power concentrated at low frequencies. The top plot of Figure 1 shows the time series with a vertical line indicating the change point. The middle plot shows the normalized periodogram for each block, while the bottom plot shows the scalogram for each block. There is a clear visual difference between the periodograms and the scalograms for each block. The simulations and data analysis reported in this paper are based on a Haar wavelet basis to compute the scalogram although in theory one may use other wavelet filters.

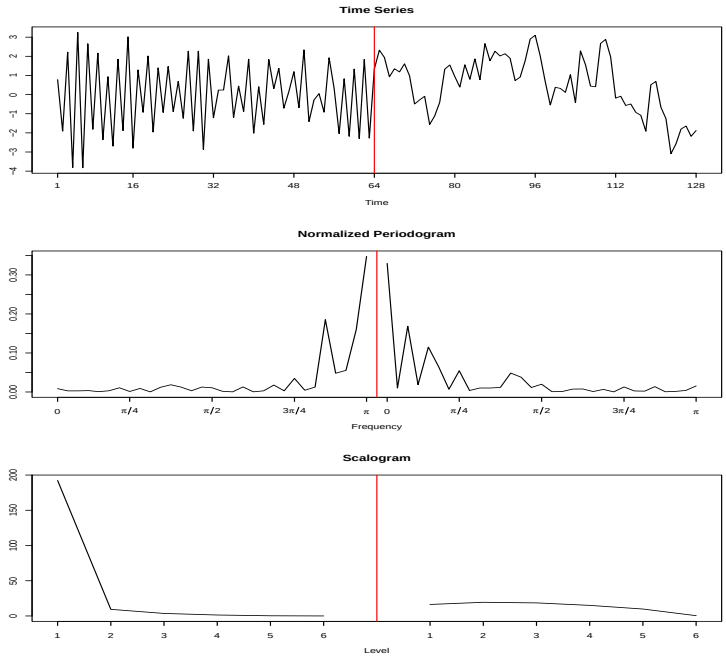

Figure 1: Time series, periodogram, and scalogram for a simulated AR(1) process with change in AR coefficient at $t = 64$

The top and middle left plots of Figure 2 show the transformed ratios of the normalized periodograms for the two blocks, $z_{1,k} = \log(1 + R_k)$ and $z_{2,k} = \log(1 + 1/R_k)$, while the top and middle right plots show the corresponding CUSUM statistics based on these $z_{i,k}$, relative to a uniform

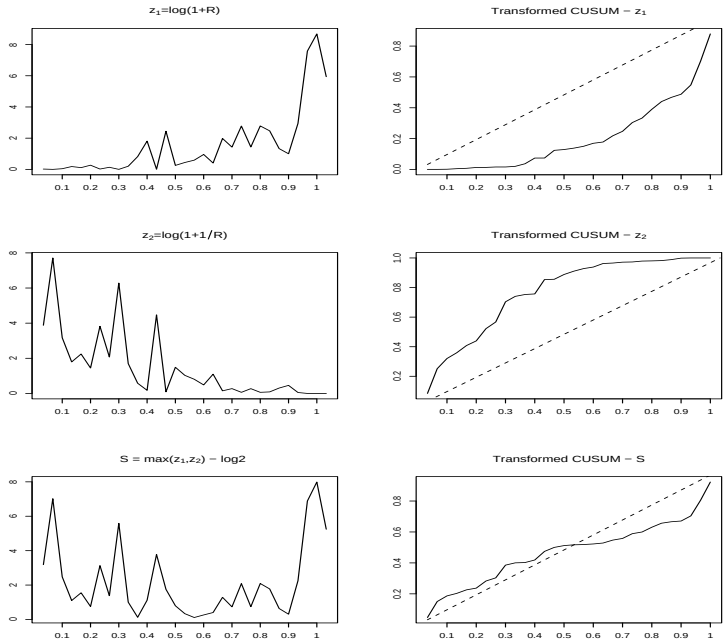

Figure 2: Left hand plots show transformed ratio statistics and maximum of transformed ratio statistics based on two blocks of size $n = 64$ for series of Figure 1, while right hand plots show corresponding CUSUM statistics based on the transformed ratios, relative to that for uniformly distributed random variables.

distribution. The departure from a uniform distribution is larger using $z_{1,k}$ in this example. The bottom left plot shows $S_k = \max\{\log(1 + R_k), \log(1 + 1/R_k)\} - \log 2$, the basis of the SR statistic, while the bottom right plot shows what happens if the CUSUM statistic is used with S_k. We see that taking the maximum of the two transformed ratios results in smaller overall departure from a uniform distribution in this case, as discussed in Section 2, due to the increase in the denominator term. Note that our method does not use the CUSUM approach, rather, we use a test statistic that has an (asymptotic) gamma distribution. Using the SR statistic, the null hypothesis of same spectral structure is rejected with $p = 0.00356$. The scalogram method rejects with $p \sim 0.00000$.

Alternatively, consider a process of length $T = 128$ having in first block of size $n = 64$ a discrete spectrum that takes value one at all frequencies except for a randomly selected 5% of them (3 frequencies in the case illustrated, $\omega_k = 5, 20, 30$), where it takes value 32. At time $t = 65$, the process switches to follow a white noise process. See Appendix A for detailed discussion on simulating a process with a discrete spectrum. Figures 3 and 4 show plots analogous to those

in Figures 1 and 2 for this situation. We see that the CUSUM method is much less powerful in detecting deviations from uniform distribution when the differences in spectral behavior are spread among different frequency bands. The SR statistic gives a p-value of 0.037 for this case. The scalogram method rejects the null with $p = 0.002$, while the CUSUM method can't reject the null, having p-value $= 0.521$.

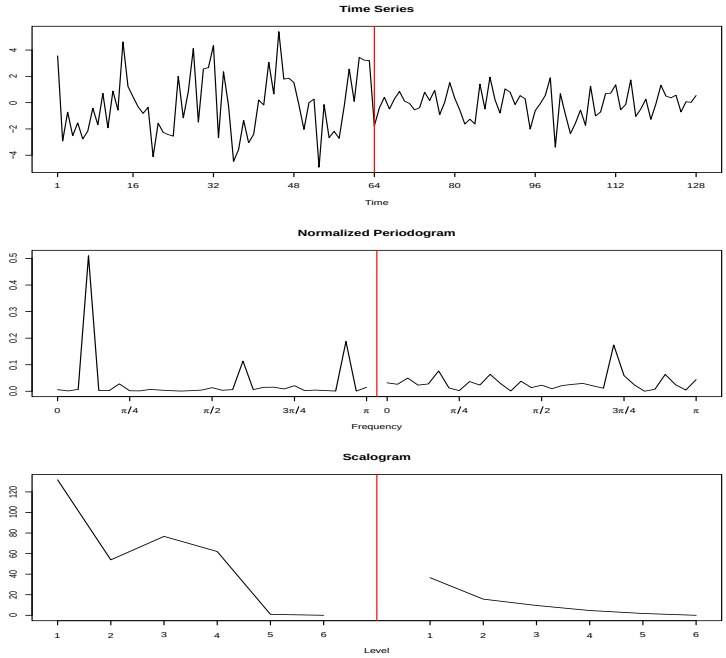

Figure 3: Time series, periodogram, and scalogram plots for simulated process having peak in variance components at discrete frequencies 5, 20, 30, changing to a white noise process at $t = 65$.

Now consider a slowly evolving AR(1) process $x_t = a_t x_{t-1} + \epsilon_t,\ t = 1, \cdots, 576$, where ϵ_t are *i.i.d.* standard normal and $a_t = -0.8 + 0.2 \times [t/64]$. The top, middle, and bottom plots of Figure 5 show the gradual change in the process as reflected in the series values, the normalized periodogram of each block, and the scalogram of each block.

In order to detect changes in a process as quickly as possible, we focus on small blocks, making it potentially harder to detect small changes due to a small number of observations. The sequential testing procedure discussed in Section 3.3 can increase the power of the test since more information is essentially aggregated across time blocks. For the example of Figure 5, applying the CUSUM or SR test methods only to adjacent blocks results in no detection of a change in autocorrelation structure. Applying the sequential test procedure, a change point is detected between blocks 8 and

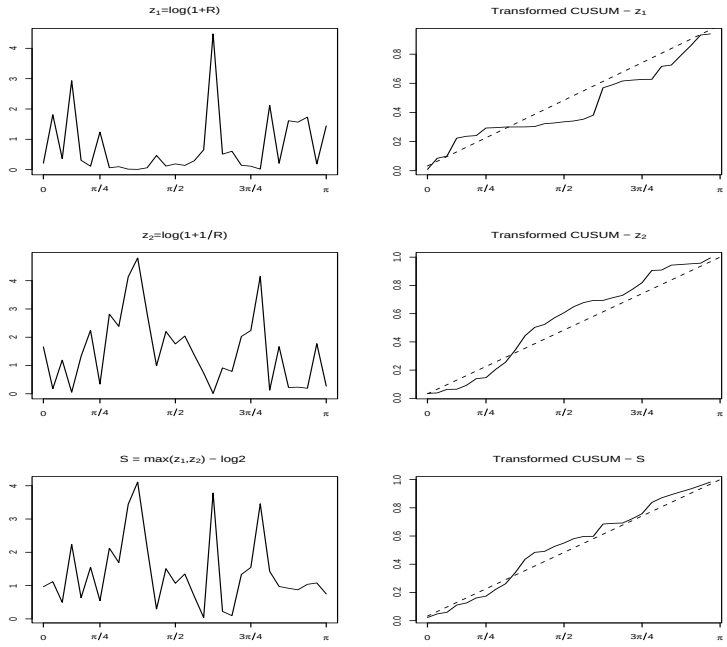

Figure 4: Left hand plots show transformed ratio statistics and maximum of transformed ratio statistics based on two blocks of size $n = 64$ for series of Figure 3, while right hand plots show corresponding CUSUM statistics based on the transformed ratios, relative to that for uniformly distributed random variables.

9 using the SR statistic, and between blocks 6 and 7 using the scalogram method.

The next section summarizes results of more broad scale simulation studies to assess the capability of the proposed change point detection methods under scenarios similar to the three discussed above.

4 Simulation Study

We conducted three simulation experiments, each with 1000 replicated time series of length $T = 1024$. Each series was divided into blocks of length $n = 64$, for a total of 16 blocks (15 tests). This block size was chosen so as to balance the trade-off between the ability to more accurately identify the change point (small block size) and higher power (larger block size). In practice, the choice of block size will depend on the particular application. The CUSUM, SR, and scalogram methods, with and without sequential testing, were applied to each set of blocks. In each case, we report the percentage of times (out of 1000) that a change was detected at a block boundary using

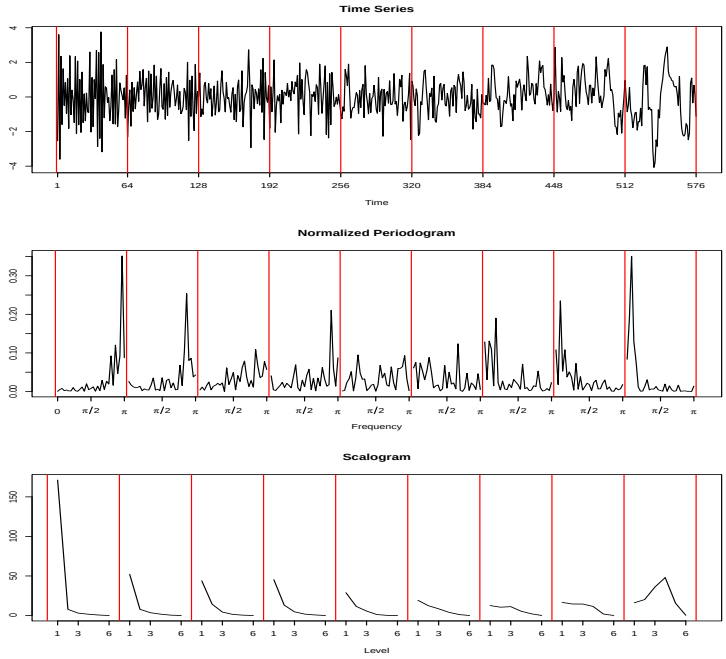

Figure 5: Time series, blocks periodograms, and block scalograms for a single realization from a slowly changing AR(1) process

a significance level of 5%. The CUSUM and SR statistics are based on normalized periodograms in all cases.

4.1 Abruptly Changing Process with Discrete Spectra

Consider a time series, $\{X_t : t = 1, \cdots, 1024\}$, which is a piecewise stationary process with change points at $t = 250$ and 510

$$X_t = \begin{cases} X_{1,t}, & 1 \leq t \leq 250, \\ X_{2,t}, & 1 \leq t \leq 510, \\ X_{3,t}, & 511 \leq t \leq 704, \end{cases} \tag{4.1}$$

In this example, we assume that the spectrum of each block is discrete. A discrete process, Y_t, has a discrete Cramèr representation

$$Y_t = \sum_{k=T/2-1}^{T/2} A_k \exp(i2\pi kt/T)$$

where the random coefficient is decomposed into a product of a fixed and random part $A_k = \alpha_k z_k$. The fixed quantity, α_k, is restricted to be real-valued and symmetric about $k = 0$ whereas z_k is complex-valued random variable with mean 0 and unit variance (for purposes of identifiability).

They z_k's are uncorrelated over k despite the special relationship $\alpha_k = \alpha_k^*$ for negative $k < 0$ (where b^* denotes the complex-conjugate of b). See Appendix A for a description of generating periodogram data from such processes.

In the first block of the model 4.1, $\alpha_k = 1$ for all k. That is, the spectrum for the first block is flat (without any peaks). In the second block, there are 3 peaks which are selected randomly and $\alpha_k = 2^5$ at these peaks. In the third block, $\alpha_k = 2^8$ also at the 3 frequencies that are selected randomly which may differ from the peaks at the second block. Table 1 shows change detection results for series generated to have such structure. The results indicate that the CUSUM method has little power in detecting such changes. Both the SR and the scalogram methods are more sensitive to the changes than the CUSUM method. We observe that at the methods have a higher rate of detection at the second change-point than the first change-point. This is because in the first change point there are really only 3 frequencies at which the discrete spectra differ. On the other hand, at the second change point, there could be as much as 6 frequencies at which the spectra differ. Moreover, we note that the updating procedure improves the power to detect the changes to some degree in all cases.

Table 1: Results for discrete spectrum process having an abrupt change

	CUSUM		SR		Scalogram	
Update	w/o	with	w/o	with	w/o	with
1	2.0	2.0	5.3	5.3	1.3	1.3
2	3.4	2.5	4.3	4.4	0.8	0.6
3	4.3	3.7	4.4	4.9	1.1	0.6
4	10.8	9.6	26.2	26.3	94.1	94.7
5	2.8	4.9	1.2	5.5	0.4	5.9
6	2.9	3.3	1.6	3.0	0.7	2.3
7	2.1	2.3	1.7	2.1	0.7	1.1
8	3.3	2.8	2.2	2.4	0.5	0.5
9	2.2	2.3	1.7	1.7	0.4	0.6
10	3.0	2.3	1.7	1.3	0.5	0.5
11	8.4	7.5	3.9	3.0	12.9	17.1
12	47.3	56.4	93.3	95.5	100	100
13	2.9	4.4	0.3	9.2	1.1	1.6
14	2.5	2.9	0.3	2.2	1.2	1.3
15	2.6	1.9	0.0	0.7	0.8	0.7

4.2 Slowly changing AR(1) process

In practice, a random process may change slowly over time. Consider a slowly changing AR(1) process, $y_t = \alpha_t y_{t-1} + \epsilon_t$, where $\alpha_t = 1.8\left[\frac{e^{\beta(t-512)/1024}}{1+e^{\beta(t-512)/1024}} - \frac{1}{2}\right]$. As α_t changes from -0.9 to 0.9, the process changes from a high frequency one to a low frequency one. As β increases, the changes around $t = 512$ become sharper; the asymptote of α_t is 0.9 as t increases and -0.9 as t decreases. A similar process was investigated in Ombao et al (2001) although their method was a retrospective segmentation of a non-stationary time series based on the SLEX (localized Fourier) transform. Table 2 gives the results of the change point detection methods applied to this process when $\beta = 50$. The coefficient α_t changes very quickly around $t = 512$. As expected, there is a higher probability of detecting a change for blocks around the $t = 512$. The scalogram method has moderate power in detecting the slowly changing *high* frequency structure, but little power in detecting the slowly changeing *low* frequency structure. The SR method performs slightly better than the CUSUM method for this process.

Table 2: Results for Smoothly Changing AR(1) Process

	CUSUM		SR		Scalogram	
Update	w/o	with	w/o	with	w/o	with
1	7.4	7.4	11.1	11.1	21.4	21.4
2	6.5	6.2	10.7	11.0	20.6	23.8
3	6.3	7.5	11.1	12.9	20.3	24.9
4	5.9	7.1	11.8	13.1	20.9	24.8
5	6.1	7.8	10.3	11.1	21.6	26.0
6	7.0	8.4	10.3	12.3	19.2	24.3
7	30.0	40.6	28.6	42.4	54.7	66.2
8	97.0	97.6	72.9	86.8	84.9	86.3
9	29.7	30.4	28.5	35.3	2.0	9.0
10	6.5	12.3	10.5	19.3	0.9	2.5
11	7.5	9.6	10.0	13.5	0.9	1.0
12	5.3	6.6	9.9	12.7	0.9	0.8
13	5.9	7.7	8.2	9.7	1.5	0.7
14	6.6	8.0	9.7	10.4	1.4	0.9
15	6.7	7.7	10.5	12.6	1.1	0.8

In summary, we see that both the SR and the scalogram methods improve upon the CUSUM method. The updating procedure results in additional gains in power, particularly in the case of a process having slowly changing correlation structure. The SR method is preferred if only changes in

spectral shape are of interest. The power of the scalogram method appears to be highly dependent on the underlying spectral structure.

5 Data Illustrations

5.1 Analysis of Brain Waves

We analyze an electroencephalogram (EEG) signal recorded at the left temporal channel during an epileptic seizure. The signal was recorded at a sampling rate of 100 Hz (100 observations per second). This signal is a subset of that used in Ombao et al (2001, 2005) who developed the Auto-SLEX method which retrospectively determines changes in the electrical activity of the brain during an epileptic seizure. To illustrate our methods, we analyze only a short segment ($T = 1024$) of the original EEG series around the time of seizure onset using blocks of size $n = 64$, for a total of 15 blocks. Our interest lies in detecting changes in the micro-structure of the EEG that might be predictive of seizure onset, rather than detecting only the seizure onset. The dotted lines in the top plots of Figure (6) indicate detected changes using the SR (left) and scalogram (right) methods, with sequential testing. The time of seizure onset (as determined by the study neurologist) is indicated by the heavy dashed line. We see that in the left temporal channel, the SR method detects changes before the onset of seizure as well as onset of the seizure itself. The scalogram method detects the onset of seizure, as well as another change after seizure onset. The bottom plots of Figure (6) show that power seems to be concentrated at slightly higher frequency bands after seizure onset. These results are comparable to those of Ombao et al (2001, 2005) in that similar change points are detected around seizure onset. However, it is worth noting that the SR method was able to detect changes in the brain electrical activity even before seizure onset. Indeed, it is typically the case that changes in the brain occur even before one observes physical manifestations of an episode such as loss of consciousness, loss of balance, etc. These results suggest that our methods can be potentially useful for monitoring EEGs in an online setting.

5.2 Analysis of Computer Performance Metrics

The ability to understand the behavior of a computer program and dynamically adapt to changing patterns of program behavior at run-time is increasingly important for improving performance.

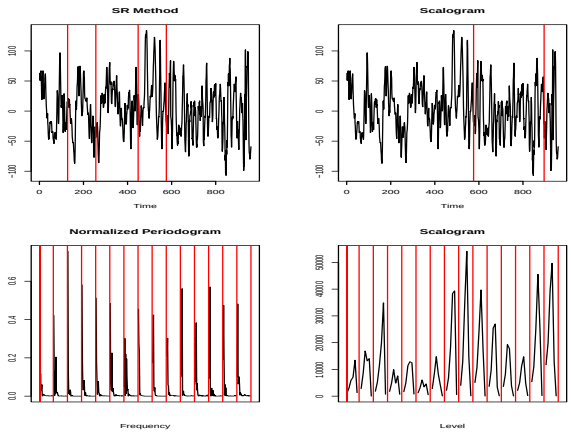

Figure 6: Top plots: Observed EEG for the Left Temporal Channel, with detected change points using SR (left) and scalogram (right) methods. Bottom plots: Periodogram (left) and scalogram (right) for each block of size $n = 64$.

For improving memory performance, one approach is to predict different segments of a program that have recurring data locality during a program's execution. Code-based methods mark loops and functions as phases and estimate their behavior through profiling (see, for example, Hsu and Kemer, 2003). However, this approach cannot react when the same procedure or loop accesses different data during a single execution. One would like to determine changes in program behavior dynamically based on metrics generated at run-time, such as Instructions per Cycle, a measure of overall performance, cache miss rate, a measure of memory performance, or branch prediction accuracy. When a change in phase is detected, elements of the program can be dynamically adapted to meet the changing needs based on observed past behavior for the same phase. For instance, data layout can be reorganized to reduce cache miss rates or improve prefetching efficiency. Existing run-time approaches typically divide program execution into fixed-length intervals and use behavior in past intervals to predict the behavior in future intervals. The behavior is characterized using simple metrics such as cache miss rates, which can be determined at run-time with minimal computational overhead (Balasubramonian, Albonesi, and Dwarkadas, 2000). A method that uses a mix of the two approaches is based on one-time profiling of a program using data collected during runs with training inputs. The program is instrumented by inserting phase markers into the code. When the instrumented program runs, it uses the first few executions of a phase to predict all later executions (Shen, Zhong, and Ding, 2004). In both approaches, the determined phases are dependent on the

particular training inputs used, although this does not seem to be a limiting factor in many practical applications. The harder problem is predicting behavior in programs in which later occurrences of a phase may have dramatically different behavior from the first few occurrences.

Here, we apply the methods of Section 3 to evaluate their usefulness for detecting changes in behavior (phase transitions) of computer processor metrics generated during run-time of a program. Changes in periodic structure, as captured by spectral peaks, are especially of interest, motivating our spectral-based approach. Additionally, Fourier and wavelet transforms on block sizes that are powers of two are very computationally inexpensive.

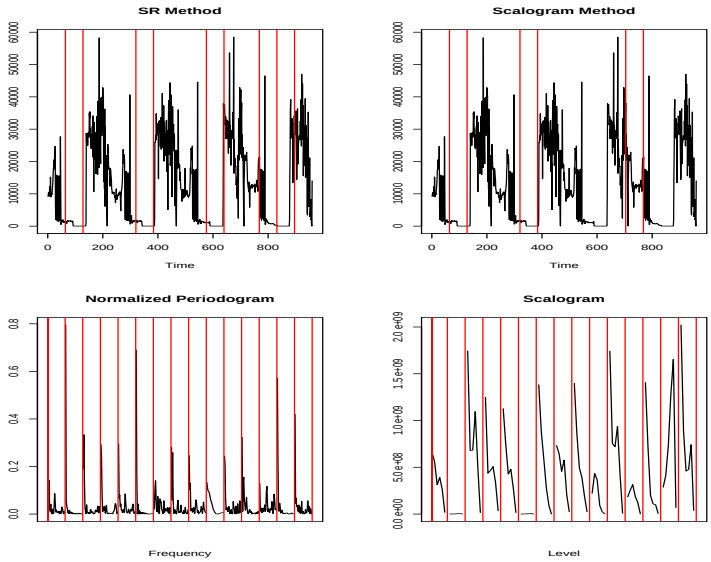

Figure 7: Top plots: Portion of trace ($n = 1024$) for branch mispredict rates (recorded every 10ms) during run-time execution of a specific program using a particular processor architecture

The top row of plots in Figure 7 shows an approximately one second portion of the series of branch mispredict rates measured every 10ms for a particular program running on Power3 processor architecture. The solid lines indicate the detected change points based on the SR method with normalized periodograms (top left) and the scalogram method (top right), using blocks of length $n = 64$ with $\alpha = 0.05$. The corresponding block periodograms and scalograms are shown in the bottom row of plots. In most cases, both methods appear to detect those phase transitions that match with visual assessment by systems researchers, although the scalogram method appears to miss a couple of the changes detected by the periodogram method. Note, however, that the series

appears to have changes in mean level within a block, violating the assumption of stationarity. A change in mean within a block impacts the second-order properties of the process, which may then manifest as a change in estimated spectrum compared to a previous block with the same underlying correlation structure but a constant mean. This suggests that the methods introduced here can be used to detect changes in mean as well as changes in correlation structure. On the other hand, changes in level may be of no inherent interest and thus our methods would indicate spurious change points in this case. One approach to handle this might be to use a simple, wavelet-based detrending method for each block before the spectral test is computed. We will investigate this further in future research.

Of course, to completely address the goal of phase *prediction*, *i.e.* predicting phase behavior, it is necessary to go beyond simply detecting changes in phase as characterized by changing spectral behavior. One approach to the prediction problem would be to match the spectral behavior of the block at which a change is detected to the behavior of past segments determined to be stationary using, for example, the spectral classification methods of Shumway (1982), Shumway (2003) or Huang, Ombao and Stoffer (2004). The program elements would then be dynamically adapted based on observed behavior in the most closely matching past phase. This adaptation would remain in place until the next change point was detected. Additional investigation of the phase prediction problem is beyond the scope of this paper. Here, we have simply used computer processor run-time metrics to illustrate the change-point detection approaches of Section 3.

6 Conclusion

We developed two non-parametric methods for detecting changes in autocorrelation structure of a time series in an online monitoring framework. The methods are spectral-based, detecting changes in the decomposition of variance over frequency or scale, without explicit reference to a parametric model for the spectral behavior. Thus they are more flexible than parametric approaches, which can be quite sensitive to deviations from the parametric assumptions.

The first method, based on Fourier decomposition of spectral variance, overcomes the limitation of Coates and Diggle (1986) on time series labeling. This reduces computational effort in an online framework. The second method is based on the wavelet variance. A multiple testing approach is

used to test for changes in behavior across multiple scales. Both methods are easy to implement, using computationally efficient FFT and MODWT algorithms, respectively. The introduced sequential testing procedure allows for incorporation of information from prior blocks for improved power, while at the same time controlling the level of the test.

The issue of how to choose an appropriate block size is one that is application dependent, weighing the trade-offs between quick detection and computational efficiency. As mentioned in Section 4, a potential issue with both tests is that when the trend is not removed from each stationary block, the distributional properties of the periodograms and scalograms may be affected. We are currently deriving the theory for the asymptotic distributions and for how the test statistic may be utilized for detecting changes in the trend (in addition to the second order structure). Other extensions to the current work include investigation of methods appropriate for detecting changes in correlation structure for multivariate time series.

Appendix A

First, consider a discretized Cramér representation of a stationary time series to be

$$X_t = \frac{1}{\sqrt{T}} \sum_{k=-(T/2-1)}^{T/2} A_k \exp(i2\pi kt/T)$$

where A_k is the coefficient that corresponds to the sinusoid having k cycles over the entire duration of T observations. Note that A_k is random variable that possesses these properties: (i.) At $k = 0, T/2$, A_k is real-valued with mean 0 and variance f_k; (ii.) for $k = 1, \ldots, T/2 - 1$, $A_{-k} = A_k^*$ where A^* denotes the complex conjugate of A; (iii.) For each k, the real and imaginary parts of A_k are uncorrelated and identically distributed with mean 0 and variance $f_k/2$ and (iv.) for $k = 0, \ldots, T/2$, all the real parts and imaginary parts of A_k are uncorrelated (within k and between k's). Moreover, the spectrum f_k is symmetric about 0, i.e., $f_{-k} = f_k$. For our purposes, f_k will be flat except for a few k's where the peaks occur. Denote $\Omega \subset \{-(T/2-1), \ldots, T/2\}$ to be the set of indices where the peaks occur. Then, the discrete spectrum f_k can now be defined to be

$$f_k = \begin{cases} \delta_k, & k \in \Omega \\ 1, & k \notin \Omega, \end{cases}$$

where $\delta_k \gg 1$. To simulate a time series, we outline the following steps:

1. Generate a white noise time series $\mathbf{Z} = [Z_1, \ldots, Z_T]$ from a distribution with mean 0 and variance 1.

2. Compute the vector of Fourier coefficients of $\mathbf{Z}$ and denote it by $\mathbf{d} = (d_k), k = -(T/2 - 1), \ldots, T/2$.

3. Compute the vector $\mathbf{A} = (A_k)$ where $A_k = \sqrt{f_k}\ d_k$.

4. Compute the inverse Fourier transform of $\mathbf{A}$ to obtain the time series $\mathbf{X} = [X_1, \ldots, X_T]$.

Acknowledgments

The authors acknowledge the extensive discussions with Sandhya Dwarkadas and Xipeng Shen, Computer Science Department, University of Rochester. Also the helpful and insightful comments from Ditlev Monrad, Statistics Department, University of Illinois at Urbana-Champaign are acknowledged. Finally, the authors thank Dr. Beth Malow, neurologist at Vanderbilt University, for providing the EEG data set.

References

Balasubramonian, R., Dwarkadas, S., and Albonesi, D.H. (2003). Dynamically Managing the Communication-Parallelism Trade-off in Future Clustered Processors, *30th International Symposium on Computer Architecture*, 275-286, June 2003.

Basseville, M. and Nikiforov, I. (2000). *Detection of Abrupt Changes: Theory and Application.* Massachusetts:Prentice-Hall.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57, 289-300.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory,* Oakland, CA: Holden-Day.

Coates, D.S. and Diggle, P.J. (1986). Tests for comparing two estimated spectral densities. *Journal of Time series analysis,* 7, 7-20.

Diggle, P.J. and Fisher, N.I. (1991). Nonparametric comparison of cumulative periodograms. *Applied Statistics*, 40(3), 423-434.

Hsu, C.-H. and Kermer, U. (2003). The design, implementation and evaluation of a compiler algorithm for CPU energy reduction. In *Proceedings of ACM SIGPLAN COnfernec eon Programming Language Design and Implementation*, San Diego, CA, June, 2003.

Huang, H-Y., Ombao, H. and Stoffer, D. (2004). Classification and Discrimination of Non-stationary Time Series Using the SLEX Model. *Journal of the American Statistical Association*, 99, 763-774.

Inclan, C. (1993). Detection of multiple changes of variance using posterior odds. *Journal of Business and Economic Statistics*, 11, 289-300.

Inclan, C. and Tiao, G. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89, 913-923.

Jenkins, G. and Watts, O. (1968). Spectral Analysis and Its Applications. *Holden Day, Inc.*

Kakizawa, Y., Shumway, R.H. and Taniguchi, M. (1998). Discrimination and Clustering for Multivariate Time Series *Journal of the American Statistical Association*, Vol. 93, 441, 328-340.

Kluppelberg, C. and Mikosch, T. (1996). Gaussian limit fields for the integrated periodogram. *Annals of Applied Probability*, 6, 969-991.

Maharaj, E.A. (2002). Comparison of non-stationary time series in the frequency Domain. *Computational Statistics and Data Analysis,* 40, 131-141.

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, 543-560.

Ombao, H., Heo, J. and Stoffer, D. (2004). Statistical analysis of seismic signals: An almost real-time approach. *Time Series Analysis and Applications to Geophysical Systems: IMA Series 139*, New York:Wiley, 53-72.

Ombao, H., von Sachs, R. and Guo, W. (2005). The SLEX Analysis of Multivariate Non-Stationary Time Series. *Journal of the American Statistical Association,* to appear.

Percival D.B. and Walden A.T. (2000) Wavelet methods for time series analysis. *Cambridge University Press.*

Percival, D. B. (1995) On estimation of the wavelet variance *Biometrika*, Vol 82, 619-631.

Priestley M. B. (1981). *Spectral Analysis and Time Series.* London:Academic Press.

Shen, X., Zhong, Y. and Ding, C. (2004). Locality Phase Prediction, in *Proceedings of the Eleventh International Conference on Architectural Support for Programming Languages and Operating Systems* (ASPLOS XI) , Boston, MA, October 2004.

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

Shumway, R. H. (2003). Time-frequency clustering and discriminant analysis. *Stat. Prob. Letters*, 63, 307-314.

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

Whitcher, B., Byers,S., Guttorp, P. and Percival, D. (2002). Testing for homogeneity of variance in time series: Long memory, wavelets and the Nile River. *Water Resources Research*, 38, pp. 10.1029/2001WR000509