Analysis of Long Period Variable Stars with Nonparametric Tests for Trend Detection

(Running title: Analysis of Long Period Variable Stars)

Cheolwoo Park and Jeongyoun Ahn
Department of Statistics
University of Georgia
GA 30602, USA

Martin Hendry
SUPA, School of Physics and Astronomy
University of Glasgow
G12 8QQ, UK

Woncheol Jang*
Department of Epidemiology and Biostatistics
University of Georgia
GA 30602, USA

*Address for correspondence: 129B Coverdell Center, Department of Epidemiology and Biostatistics, University of Georgia, GA 30602 Email: jang@uga.edu

Abstract

In astronomy the study of variable stars – i.e., stars characterized by showing significant variation in their brightness over time – has made crucial contributions to our understanding of many fields, from stellar birth and evolution to the calibration of the extragalactic distance scale. In this paper, we develop a method for analyzing multiple, (pseudo)-periodic time series with the goal of detecting temporal trends in their periods. We allow for non-stationary noise and for clustering among the various time series. We apply this method to the long-standing astronomical problem of identifying variables stars whose regular brightness fluctuations have periods that change over time. The results of our analysis show that such changes can be substantial, raising the possibility that astronomers' estimates of galactic distances can be refined. Two significant contributions of our approach, relative to existing methods for this problem, are as follows: 1. The method is nonparametric, making minimal assumptions about *both* the temporal trends themselves but also the covariance structure of the non-stationary noise. 2. Our proposed test has higher power than existing methods. The test is based on inference for a high-dimensional Normal mean, with control of the False Discovery Rate to account for multiplicity. We present theory and simulations to demonstrate the performance of our method relative. We also analyze data from the American Association of Variable Star Observers and find a monotone relationship between mean period and strength of trend similar to that identified by Hart et al. (2007).

Keywords and Phrases: Adaptive Neyman test; Functional clustering; False discovery rate; Long period variable stars; Multiple testing; Pivotal test; Principal components; SiZer analysis; Trend detection; Wavelets.

1 Introduction

The study of variable stars has a long and illustrious history in astronomy, making crucial contributions to our understanding of many fields, from stellar birth and evolution to the calibration of the extragalactic distance scale. Variable stars are characterized by significant, and often periodic, variation in their brightness over time. These variations may be due, for example, to periodic changes in the amount of light received from a binary star system as one of the stars is regularly eclipsed by its binary companion (Popper, 1967). Equally, the variations may be due to changes in the star's intrinsic luminosity – caused for example by a regular pulsation mechanism resulting from physical processes within the star. The main focus of this paper is a class of variable stars, called *long-period variables*, whose brightness variations are fairly regular over time scales of months or years. Long-period variables are also known as *Mira* variables – named after the archetypical star Mira A in the constellation Cetus. Miras are red giant stars, in the latter stages of their evolution, which show strong variations in brightness and color due to changes in their radius, luminosity and temperature induced by pulsation. Their period of pulsation varies significantly from star to star but generally lies in the range 100 to 700 days. Whitelock (1999) review a number of reasons for the astrophysical importance of these long period variables, highlighting in particular their suitability as distance indicators – a fact which makes them useful tracers (i.e., covariates) of e.g. the size, thickness and structure of the disk and spiral arms of the Milky Way galaxy. The use of Miras as distance estimators relies upon established relations between their mean luminosity, color and pulsation period. Although strongly motivated by astrophysical theory, these relations are calibrated empirically – using nearby stars whose distance is otherwise known – and may then be applied to more remote objects to estimate their distance. (See, for example Kanbur et al. (1997) for a detailed discussion).

Application of the period-luminosity and period-luminosity-color relations for long period variables clearly requires first that the period of each variable be reliably determined, and moreover that the variables be analyzed for evidence of secular (i.e., long-term non-periodic) changes and other long-term trends in their period. Hall et al. (2000); Oh et al. (2004) address the former issue in detail.

We investigate the latter issue in this paper; specifically, we address the problem of detecting temporal trends in the period of Mira type variables. We will do so using a dataset consisting of the measured times of maximum and minimum brightness, distilled from approximately 75 years of brightness observations, for 378 long period variables. These data have been compiled and collated in Mattei et al. (1990), on behalf of the American Association of Variable Star Observers (AAVSO). See Mattei et al. (1990) for further details on how the data are collected.

Following e.g. Kanbur et al. (1997), we use the time interval between successive maxima as a measure of the instantaneous pulsation period of each variable star. As noted by previous authors, the time interval between successive minima could also be used for this purpose, but the minima are generally less well constrained because of sparse and noisy observations. Also, are the minima less reliable statistically. The number of period estimates for each variable star ranges between 32 and 212.

Analysis of long period variables has an extensive literature – with frequency domain methods, based on the Schuster periodogram, generally favored although a number of time domain methods can also be found. (See e.g. Genton and Hall (2007) for a recent review). In particular several analyses of the evidence for evolutionary trends in the period and/or amplitude of these variable stars have been conducted. These analyses are important for testing the efficacy and accuracy of long period variables as distance indicators, insofar as systematic changes in their period could lead directly to a bias in their inferred distance. Moreover, identifying and constraining evolutionary trends in the light curves of long period variables could provide valuable insight into the physical mechanisms - such as e.g. variations in the rate of mass loss from their outer envelopes - that determine their luminosity and light curve shape.

Previously, Percy et al. (1990) and Percy and Colivas (1999) use a data set similar to that considered in this paper in an attempt to detect evidence that the period of Miras could change on long time scales. Both papers use the method of Eddington and Plakidis (1929), which is suitable for detecting monotonic changes in period but not more general trends. Koen and Lombard (2004) develop a frequency domain test which is not limited to monotonic changes and use a test statistic that depends on two parameters: the intrinsic cycle-to-cycle variation in period and the measurement error. Their method is a refinement of Koen and Lombard (2001), which simultaneously use the time intervals between both successive maxima and successive minima – modeling these as separate first-order moving average processes. Their test statistic measure the agreement between the theoretical frequency spectrum of these time intervals, under the null hypothesis of a constant pulsation period, and the observe spectrum. The authors describe two methods for dealing with missing observations: either missing values are simply ignored in the calculation of the test statistic or missing values are estimated via linear interpolation. They also use the t-test suggested by Muirhead (1986) to identify outliers.

More recently Hart et al. (2007) present a time domain analysis of period changes. Their approach is also not restricted to monotonic trends; however, unlike Koen and Lombard (2004), it includes an explicit model for heteroscedasticity in the measured times of maximum brightness. Moreover they adopt the method of Storey (2002) for controlling positive false discovery rate (pFDR). Their approach is designed to be nonparametric insofar as potential trends in the tested data are estimated using local linear smoothers – the smoothing parameters of which are chosen using one-sided cross validation to account for its dependence. They propose to use a generalized profile likelihood ratio for testing the null hypothesis of no trend. The distribution of their test statistic under the null hypothesis is estimated using a bootstrap algorithm to avoid specifying a non-standard limit sampling distribution for a given star. On the other hand, their approach *does* assume a specific parametric form for the covariance structure for each star. They report that the strength of trend in the time interval between maxima is an increasing function of the mean period length.

A central objective of this paper is to propose a nonparametric test to detect any significant trend in time series. As explained in Section 2, our approach is different from Hart et al. (2007); while both approaches are based on smoothing techniques, the proposed method is more flexible in that we do *not* require an explicit parametric form for the covariance structure of the data, unlike in Hart et al. (2007). Our test statistic is similar to the adaptive Neyman test proposed by Fan (1996), but more effective due to faster convergence rate.

The remainder of the paper is organized as follows. Section 2 describes our nonparametric approach to testing for a trend in a mean function when data are curves. In order to account for possible non-stationary error structure before the test statistic is calculated, we employ functional clustering and functional principle component analysis. Then we develop the testing procedure for the mean function based on high dimensional normal mean inference. A simulation study is presented in Section 3. The variable star data analysis is integrated into Section 4. We conclude the paper with discussion in Section 5.

2 Nonparametric Test for Trend Detection

We introduce the statistical model for the variable star data in Section 2.1. Since the model has possibly non-stationary errors, a decorrelation method that yields stationary errors is presented in Section 2.2. Section 2.3 discusses some existing testing procedures for the trend in the mean function and proposes our new adaptive pivotal thresholding test.

2.1 Statistical Model

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$ denote a vector of the observed time intervals between successive occurrences of maximum brightness, at the chronologically ordered epochs $1, \ldots, n_i$ for ith star, $i = 1, \ldots, 378$. Here an epoch refers to one complete cycle of the periodic variation of a star. Then we can use the following model;

$$Y_{ij} \equiv Y_i(t_{ij}) = \mu_{i0} + f_i(t_{ij}) + \epsilon_i(t_{ij}), \tag{1}$$

where $t_{ij} = (j - 0.5)/n_i$ are standardized epochs, $j = 1, \ldots, n_i$, μ_{i0} is the grand mean of $\mathbf{Y}_i$, f_i is an unknown function and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})$ are possibly non-stationary error terms with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_i$.

We are interested in testing whether there is any trend in a curve against the null hypothesis $H_{i0} : f_i = 0$. The key idea of our method is to convert this problem into the testing of high-dimensional normal mean after decorrelation of the non-stationary error. To achieve this goal we estimate the covariance structure first. As Fan et al. (2007) point out, a nonparametric estimator of the covariance matrix is not guaranteed to be positive definite. Hart et al. (2007) assume a specific parametric covariance structure and used a profile likelihood approach for inference. While their approach is reasonable, their relatively simple parametric model may miss certain types of the true covariance structure. In order to avoid both a strict parametric approach and also a possibility of having non-positive definite covariance, we assume that $\mathbf{Y}_i$'s can be clustered based on their covariance structures, i.e., within a cluster the covariance functions are the same while the mean functions are allowed to be different.

For clustering, we employ the k-centers functional clustering method of Chiou and Li (2007), which considers L^2 distance measure between curves to make the cluster assignment. Their method takes both the mean and eigen-structure of each cluster into account to declare different clusters. The cluster predictions are based on a nonparametric random-effect model, coupled with an iterative mean and covariance updating.

Since we want to assign a clustering membership only based on the covariance similarity, we apply their clustering method after subtracting a *pilot* mean function estimate from each curve. The pilot mean function estimate is obtained after preliminary (global) decorrelation via PACE (principal component analysis through conditional expectation) by Yao et al. (2005) for more effective estimation. The PACE procedure, which will be used to estimate the covariance function once the clusters are formed, is explained in more detail in the following subsection.

2.2 Decorrelation with Clustering

Suppose that we observe $Y_i(t)$ which is a realization of a random curve $X_i(t)$ from a mixture of stochastic processes with measurement error $\delta_i(t)$. We assume that the mixture process consists of subprocess and each subprocess is associated with a cluster. Furthermore, we assume that the covariance functions of each curve are the same within clusters.

Consider a Karhunen-Loève expansion of the random function such that

$$X_i(t) = \mu_{i0} + f_i(t) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t),$$

where the eigenfunctions ϕ_k are an orthonormal basis and the principal score coefficients $\xi_{ik} = \langle X_i - \mu_{i0} - f_i, \phi_k \rangle$ are independent random variables with $\mathrm{E}(\xi_{ik}) = 0$ and $\mathrm{Var}(\xi_{ik}) = \lambda_k$. Here λ_k are eigenvalues of the covariance of $X_i(t)$ and we assume that $\{\lambda_k\}_{k=1}^{\infty}$ is a nonincreasing sequence and $\sum_k \lambda_k < \infty$. Then our model (1) for non-stationary data can be rewritten as follows:

$$Y_i(t_{ij}) = \mu_i(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{ij}) + \delta_{ij}, \tag{2}$$

where $\mu_i(t_{ij}) = \mu_{i0} + f_i(t_{ij})$ and δ_{ij} are additional measurement errors that are assumed to be iid with $\mathrm{E}(\delta_{ij}) = 0$ and $\mathrm{Var}(\delta_{ij}) = \sigma_\delta^2$ and independent of the random coefficients ξ_{ik}.

Yao et al. (2005) suggest to use principal component analysis through conditional expectation (PACE) to estimate ξ_{ik}:

$$\widetilde{\xi}_{ik} = \mathrm{E}(\xi_{ik}|\mathbf{Y}_i) = \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})$, $\boldsymbol{\phi}_{ik} = (\phi_k(t_{i1}), \ldots, \phi_k(t_{i,n_i}))$, and $\boldsymbol{\Sigma}_i = \mathrm{cov}(\mathbf{Y}_i, \mathbf{Y}_i)$. They further assumed that $X_i(t)$ can be well approximated by the projection on the functional space spanned by the first K eigenfunctions and recommended to use the following predictor of $X_i(t)$ in practice:

$$\widehat{X}_i^K(t) = \widehat{\mu}_i(t) + \sum_{k=1}^{K} \widehat{\xi}_{ik}\widehat{\phi}_k(t)$$

where

$$\widehat{\xi}_{ik} = \widehat{\mathrm{E}(\xi_{ik}|\mathbf{Y}_i)} = \widehat{\lambda}_k \widehat{\boldsymbol{\phi}}_{ik}^T \widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i).$$

Here $\widehat{\boldsymbol{\mu}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$ are estimates of mean and covariance functions of $\mathbf{Y}_i$ using local smoothing and $\widehat{\lambda}_{ik}$ are the sample eigenvalues, $\widehat{\boldsymbol{\phi}}_{ik}$ are the sample eigenfunctions that are obtained by solving the following eigen-equations

$$\int \widehat{G}(s,t)\widehat{\phi}_k(s)d(s) = \widehat{\lambda}_k\widehat{\phi}_k(t),$$

where $(\widehat{\mathbf{\Sigma}}_i)_{j,l} = \widehat{G}(t_j, t_l) + \widehat{\sigma}^2_\delta \widetilde{\delta}_{jl}$. Here, $\widetilde{\delta}_{jl}$ is 1 if $j = l$ and 0 otherwise. For the examples in this paper, we choose K based on the proportion of the explained variance, with the threshold .85. The local linear smoother (Fan and Gijbels, 1996) is used since mean, covariance, and eigenfunctions are assumed to be smooth. See Yao et al. (2005) for more details.

An important assumption of the PACE procedure, which does not hold for our model (1), is that all the curves in the data have the identical mean and covariance functions so that they can be estimated from the pooled data. In this work we assume that the curves can be divided into a few clusters according to their covariance structure. We propose the following steps for clustering:

C1. Run PACE to estimate $(\xi_{ik}, \phi_k(t))$ based on *all* the curves in the data and denote it by $(\widehat{\xi}^{(A)}_{ik}, \widehat{\phi}^{(A)}_k(t))$.

C2. For each curve, estimate a pilot mean curve $\mu_i(t)$ with $Y_i(t) - \sum_{k=1}^{K} \widehat{\xi}^{(A)}_{ik}\widehat{\phi}^{(A)}_k(t)$ using local polynomial and obtain $\widehat{\mu}_i(t)$.

C3. Apply the clustering procedure proposed by Chiou and Li (2007) to $Y_i(t) - \widehat{\mu}_i(t)$.

Step C1 is required to obtain more accurate $\widehat{\mu}_i(t)$ in Step C2. Once $\widehat{\mu}_i(t)$ is subtracted from the original data $Y_i(t)$, each curve roughly has a zero mean function, and thus the clustering procedure can be done based on the covariance structure. In Step C2, we choose the bandwidth using the plug-in method proposed by Ruppert et al. (1995).

Once the clusters of the curves are found, decorrelation with PACE is applied within each cluster to the original data $Y_i(t)$. In the following we list the detailed steps of the iterative decorrelation procedure. For each cluster,

E1. (Preliminary PACE) Run PACE to estimate an initial value of $(\xi_{ik}, \phi_k(t))$ and denote it by $(\widehat{\xi}_{ik}^{(0)}, \widehat{\phi}_k^{(0)}(t))$.

E2. For each curve, estimate the mean curve $\mu_i(t)$ with $Y_i(t) - \sum_{k=1}^{K} \widehat{\xi}_{ik}^{(0)} \widehat{\phi}_k^{(0)}(t)$ using local polynomial and obtain $\widehat{\mu}_i(t)$.

E3. Run PACE again with $Y_i(t) - \widehat{\mu}_i(t)$ to estimate $(\xi_{ik}, \phi_k(t))$ and denote it by $(\widehat{\xi}_{ik}, \widehat{\phi}_k(t))$.

E4. Obtain $Y_i^*(t) = Y_i(t) - \sum_{k=1}^{K} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$, which is subject to the testing procedure which will be explained in the next subsection.

Again, Step E1 is required to obtain more accurate $\widehat{\mu}_i(t)$ in Step E2. Then, we remove a trend $\widehat{\mu}_i(t)$ from each curve so that the covariance structure can be estimated by PACE in Step E3. In this way we allow each curve to have a different mean function and still use PACE for estimating the common covariance structure within a cluster. In Step E2, we choose the bandwidth using the plug-in method proposed by Ruppert et al. (1995). We visualize the decorrelation procedure in Figures 3 and 4 in Section 3 using one of the simulated examples. Further iterations can be implemented until convergence and we stop the iteration if the sum of the ratio $\|\widehat{\boldsymbol{\mu}}_i^{(j)} - \widehat{\boldsymbol{\mu}}_i^{(j+1)}\|$ and $\|\widehat{\boldsymbol{\mu}}_i^{(j)}\|$ is less than a small constant, where j is the iteration index. In our simulation examples, $\sum_i \|\widehat{\boldsymbol{\mu}}_i^{(1)} - \widehat{\boldsymbol{\mu}}_i^{(2)}\|/\|\widehat{\boldsymbol{\mu}}_i^{(1)}\|$ is close to zero, which suggests that one iteration is sufficient.

2.3 Testing a Trend after Decorrelation

In this section, we explain the connection between the nonparametric trend detection and high dimensional normal mean inference problem using the approach by Fan

(1996); Beran and Dümbgen (1998). Without loss of generality, we assume $n_i = n$ and drop the subscript i from here on for the sake of simplicity. After the decorrelation process in the previous section, we can assume

$$Y_j^* \equiv Y^*(t_j) = \mu_0 + f(t_j) + \epsilon_j, \quad j = 1, \ldots, n, \tag{3}$$

where μ_0 is the grand mean of $\mathbf{Y} = (Y_1, \ldots, Y_n)$, f is an unknown function and ϵ_j are stationary errors with $\mathrm{E}(\epsilon_j) = 0$ and $\mathrm{Var}(\epsilon_j) = \sigma^2$. We are interested in testing whether there is any trend in a curve against the null hypothesis $H_0 : f = 0$, which we can convert into testing the high-dimensional normal mean after whitening error terms using a wavelet basis and eigenfunctions. Fan (1996) introduces nonparametric tests for high dimensional normal mean. We propose a new nonparametric test for high-dimensional normal mean by converting nonparametric confidence sets given by Beran and Dümbgen (1998) and Genovese and Wasserman (2005).

2.3.1 Testing Multivariate Normal Mean in High Dimension

Suppose, without loss of generality, $\mu_0 = 0$. Suppose further that $f \in L_2$. Then one can expand this function with an orthogonal basis $\{\psi_k\}_{k=1}^{\infty}$,

$$f(t) = \sum_{k=1}^{\infty} \theta_k \psi_k(t),$$

where $\theta_k = \langle f, \psi_k \rangle$. Because the number of estimable parameters cannot be bigger than the size of data, in practice we often use an approximation f^n for f:

$$f^n(t) = \sum_{k=1}^{n} \theta_k \psi_k(t).$$

By assuming a certain functional space such as Besov space, one can easily show that $\|f - f^n\|^2 = \sum_{k=n+1}^{\infty} \theta_k^2$ is negligible. Hence testing $f = 0$ is (approximately) equivalent to testing $H_0 : \boldsymbol{\theta} \equiv (\theta_1, \theta_2, \ldots, \theta_n) = \mathbf{0}$.

Our next task is then to derive a test statistic for testing H_0. Let

$$Z_k = \frac{1}{n} \sum_{j=1}^{n} Y_j^* \psi_k(t_j).$$

If ϵ_j are independent and $\mathrm{Var}(\epsilon_j) = \sigma^2$, then Z_k are independent and, by the Central Limit Theorem, $Z_k \sim N(\theta_k, \sigma_n^2)$. where $\sigma_n^2 = \sigma^2/n$.

If the error terms are stationary in the original function space, one can use wavelet transform as an orthogonal basis to whiten the error terms and make an inference in the wavelet domain (Fan, 1996). Whitening refers to the ability of the discrete wavelet transform to decorrelate time series data. The decorrelation properties have been established for wavelets with compact support and are stronger for wavelets with a higher number of vanishing moments. See Wornell (1996) for more details.

As a result, the significance test of no trend is equivalent to testing a multivariate normal mean in high dimension:

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\theta} \neq \mathbf{0}.$$

Given a specific alternative $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the optimal test based on the Neyman-Pearson lemma is to reject H_0 if

$$\boldsymbol{\theta}_0^T \mathbf{Z} > \sigma_n \|\boldsymbol{\theta}_0\| z_\alpha,$$

where $\mathbf{Z} = (Z_1, \ldots, Z_n)$, z_α is the $100(1-\alpha)$th percentile of the standard normal distribution and α is the significance level. Since the optimal test statistic depends on the specific alternative $\boldsymbol{\theta}_0$, a naïve test statistic would be $T_n = \|\mathbf{Z}\|^2 = \sum_{k=1}^{n} Z_k^2$. Specifically, reject H_0 if

$$T_n \geq \sigma_n^2 \chi_{n,1-\alpha}^2 \approx \sigma_n^2 (n + \sqrt{2n} z_\alpha),$$

since $T_n/\sigma_n^2 \sim \chi_n^2$ under H_0. Here $\chi_{n,1-\alpha}^2$ is the $100(1-\alpha)$th percentile of χ^2 distribution with degrees of freedom n. The above test statistic, however, suffers from extremely low power under the alternative $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ when n is large. Furthermore, the power converges to no more than α even though $\|\boldsymbol{\theta}_0\| \to \infty$ when $\|\boldsymbol{\theta}_0\| \geq \sqrt{2\sqrt{2} z_\alpha} n^{1/4} \sigma_n$ (Wasserman, 2005).

2.3.2 The Adaptive Neyman and Thresholding Tests

Fan (1996) suggests testing only part of components of $\boldsymbol{\theta}$ so one can increase the power of tests. If most of large coefficients of $\boldsymbol{\theta}$ are placed in the first m components, using test statistics based on $\sum_{k=1}^{m} Z_k^2$ is reasonable. Let $T_k = Z_k/\widehat{\sigma}_n$ where $\widehat{\sigma}_n^2$ is an unbiased estimator of σ_n^2. Wasserman (2005) suggests to use

$$\widehat{\sigma}^2 = n\sigma_n^2 = \frac{n}{J_n} \sum_{j=n-J_n+1}^{n} Z_j^2,$$

where $J_n = n/4$ as a default value. The following $\widehat{m}$ is suggested for m

$$\widehat{m} = \underset{m:1 \leq m \leq n}{\operatorname{argmax}} \left\{ \frac{1}{\sqrt{m}} \sum_{k=1}^{m} (T_k^2 - 1) \right\}.$$

Then the adaptive Neyman test rejects H_0 if

$$T_{AN}^* = \frac{1}{\sqrt{2\widehat{m}}} \sum_{k=1}^{\widehat{m}} (T_k^2 - 1) \tag{4}$$

is too large.

Fan (1996) notes that the adaptive Neyman test is not effective when large coefficients of $\boldsymbol{\theta}_0$ are located at large indices. Based on the fact that the coefficients at low resolution level can be reasonably large, he applies the hard thresholding technique only to coefficients at high resolution level and proposes another test statistic:

$$T_H^* = \sum_{k=1}^{J_0} T_k^2 + \sum_{k=J_0+1}^{n} T_k^2 I(|T_k| > \delta_H), \tag{5}$$

which we shall refer to Fan's thresholding test. For the choice of J_0 and δ_H, see Fan (1996). Fan (1996) derives the asymptotic distributions of the above test statistics, however, the convergence rates are rather slow. Fan and Lin (1998) provide finite sampling distributions of test statistics based on simulation studies and recommend to use them for hypothesis testing. However, the table in Fan and Lin (1998) does not provide p-values but ranges of them. This can be problematic when one wants to adjust multiplicity of simultaneous testing with a p-value adjustment method such as the False Discovery Rate (FDR) controlling method (Benjamini and Hochberg, 1995).

2.3.3 Adaptive Pivotal Thresholding Test

We propose a new nonparametric testing procedure based on the connection between testing and confidence interval. For univariate hypothesis testing, one can always use a confidence interval as a testing procedure by checking whether the confidence interval contains the null. We extend this idea to high dimensional hypothesis testing based on the uniform confidence sets approach developed by Beran and Dümbgen (1998) and Genovese and Wasserman (2005).

Based on $\|\mathbf{Z} - \boldsymbol{\theta}\|^2/\sigma_n^2 \sim \chi_n^2$, a naïve confidence set for $\boldsymbol{\theta}$ is

$$\mathcal{B}_n = \left\{ \boldsymbol{\theta} \in R^n \middle| \|\mathbf{Z} - \boldsymbol{\theta}\|^2 \leq \sigma_n^2 \chi^2_{n,1-\alpha} \right\}.$$

Then,

$$\Pr(\boldsymbol{\theta} \in \mathcal{B}_n) = 1 - \alpha, \quad \forall \boldsymbol{\theta} \in R^n.$$

The expected radius of this confidence set is $\sqrt{n\sigma_n^2} = \sigma$, which is disappointing since the radius will remain a constant even when $n \to \infty$. This confidence set is indeed the L^2 loss function centered at the Maximum Likelihood Estimators (MLE). Hence, if one uses a shrinkage estimator such as the James-Stein estimator (James and Stein, 1960) as the center of the ball, the loss function which is equivalent to the square of the radius of the confidence ball will be smaller.

Beran and Dümbgen (1998) extend this idea by introducing a generalized shrinkage estimator, called a modulator. A modulator is a vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ with $0 \leq \lambda_j \leq 1$, and a modulation estimator is a componentwise linear estimator of the form:

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^T \mathbf{Z} = (\lambda_1 Z_1, \lambda_2 Z_2, \ldots, \lambda_n Z_n).$$

For example, a nested subset selection modulator is of the form of $\boldsymbol{\lambda} = (1, \ldots, 1, 0, \ldots, 0)$.

Let $L_n(\boldsymbol{\lambda}) \equiv \|\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}) - \boldsymbol{\theta}\|^2$ and $\widehat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \widehat{R}_n(\boldsymbol{\lambda})$. Here $\widehat{R}_n(\boldsymbol{\lambda})$ be the Stein's Unbiased Risk Estimator (SURE) of $R_n(\boldsymbol{\lambda}) = \mathrm{E}(L_n(\boldsymbol{\lambda}))$ (Stein, 1981):

$$\widehat{R}_n(\boldsymbol{\lambda}) = \sum_{j=1}^{n} (Z_j^2 - \widehat{\sigma}_n^2)(1 - \lambda_j)^2 + \widehat{\sigma}_n^2 \sum_{j=1}^{n} \lambda_j^2.$$

If we define the pivot process:

$$B_n(\boldsymbol{\lambda}) \equiv \sqrt{n}(L_n(\boldsymbol{\lambda}) - \widehat{R}_n(\boldsymbol{\lambda})),$$

Beran and Dümbgen (1998) show that $B_n(\widehat{\boldsymbol{\lambda}})/\widehat{\tau}(\widehat{\boldsymbol{\lambda}})$ converges to $N(0,1)$ where $\widehat{\tau}^2(\boldsymbol{\lambda})$ is an unbiased estimator of $\tau^2(\boldsymbol{\lambda}) = \mathrm{Var}(B_n(\boldsymbol{\lambda}))$. For example, consider a nested subset selection modulator $\boldsymbol{\lambda}_m = (\underbrace{1, \ldots, 1}_{m}, 0, \ldots, 0)$. Then, the pivot process is

$$B_n(\boldsymbol{\lambda}_m) = \sqrt{n}(L_n(\boldsymbol{\lambda}_m) - \widehat{R}_n(\boldsymbol{\lambda}_m)),$$

where $\widehat{R}_n(\boldsymbol{\lambda}_m) = m\widehat{\sigma}_n^2 + \sum_{j=m+1}^n (Z_j^2 - \widehat{\sigma}_n^2)$. If we denote $\widetilde{m} = \underset{m}{\operatorname{argmin}} \widehat{R}_n(\boldsymbol{\lambda}_m)$, then the corresponding modulator estimator is of the form of $\boldsymbol{\theta}(\boldsymbol{\lambda}_{\widetilde{m}}) = (\theta_1, \ldots, \theta_{\widetilde{m}}, 0, \ldots, 0)$ and

$$\begin{aligned} \tau^2(\boldsymbol{\lambda}_{\widetilde{m}}) &= 2n\sigma_n^2 \left(n\sigma_n^2 + 2 \sum_{j=\widetilde{m}+1}^{n} \theta_j^2 \right), \\ \widehat{\tau}^2(\boldsymbol{\lambda}_{\widetilde{m}}) &= 2n\widehat{\sigma}_n^2 \left(n\widehat{\sigma}_n^2 + 2 \sum_{j=\widetilde{m}+1}^{n} (Z_j^2 - \widehat{\sigma}_n^2) \right). \end{aligned}$$

Following Stein (1981); Beran and Dümbgen (1998), one can invert the pivot $B_n(\boldsymbol{\lambda}_m)$ in terms of the loss function to construct a $100(1-\alpha)\%$ confidence ball $\mathcal{D}_n$ for $\boldsymbol{\theta}$:

$$\mathcal{D}_n \equiv \left\{ \boldsymbol{\theta} \middle| \|\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}_{\widetilde{m}}) - \boldsymbol{\theta}\|^2 \leq \frac{z_\alpha \widehat{\tau}(\boldsymbol{\lambda}_{\widetilde{m}})}{\sqrt{n}} + \widehat{R}_n(\boldsymbol{\lambda}_{\widetilde{m}}) \right\}.$$

One can reject $H_0 : \boldsymbol{\theta} = \mathbf{0}$ if $\mathcal{D}_n$ does not contain the origin. In other words, reject H_0 when

$$\|\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}_{\widetilde{m}})\|^2 = \sum_{k=1}^{\widetilde{m}} Z_k^2 > s_n^2(\boldsymbol{\lambda}_{\widetilde{m}}) \equiv \frac{z_\alpha \widehat{\tau}(\boldsymbol{\lambda}_{\widetilde{m}})}{\sqrt{n}} + \widehat{R}_n(\boldsymbol{\lambda}_{\widetilde{m}}),$$

which we shall refer to the Adaptive Pivotal Test. This is similar to the adaptive Neyman test, but different in choosing m.

One also can extend Fan's thresholding test (5) based on Genovese and Wasserman (2005). Suppose that

$$\widetilde{Y}_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where $x_i = i/n$ and $\epsilon_i \sim N(0, \sigma^2)$.

Let ϕ and ψ denote a father and mother wavelet respectively:

$$\phi_{J_0,k}(x) = 2^{J_0/2}\phi(2^{J_0}x - k), \quad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Assuming $f \in L^2[0,1]$, one can expand f with wavelet basis:

$$f(x) = \sum_{k=0}^{2^{J_0}-1} \alpha_k \phi_{J_0,k}(x) + \sum_{j=J_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x),$$

where $\alpha_k = \langle f, \phi_{J_0,k}\rangle$ and $\beta_{j,k} = \langle f, \psi_{j,k}\rangle$. Define f_n, the projection of f on the span of the first $n = 2^{J_1}$ basis as follows:

$$f_n(x) = \sum_{k=0}^{2^{J_0}-1} \alpha_k \phi_{J_0,k}(x) + \sum_{j=J_0}^{J_1-1} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x).$$

Also, define naïve estimates of wavelet coefficients as follows:

$$\begin{aligned}
\widetilde{\alpha}_k &= \sum_{i=1}^{n} \widetilde{Y}_i \int_{(i-1)/n}^{i/n} \phi_{J_0,k}(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} \phi_{J_0,k}(x_i)\widetilde{Y}_i \approx \alpha_k + \frac{\sigma}{\sqrt{n}} Z_k, \\
\widetilde{\beta}_{j,k} &= \sum_{i=1}^{n} \widetilde{Y}_i \int_{(i-1)/n}^{i/n} \psi_{j,k}(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k}(x_i)\widetilde{Y}_i \approx \beta_{j,k} + \frac{\sigma}{\sqrt{n}} Z_{j,k},
\end{aligned}$$

where Z_k and $Z_{j,k}$ are standard normals. Then we consider the following test procedure for H_0: all wavelet coefficients are zero versus H_1 : at least one wavelet coefficient is not zero.

Denote $\boldsymbol{\mu}^n = (\mu_1, \ldots, \mu_n) = (\alpha_0, \ldots, \alpha_{2^{J_0}-1}, \beta_{J_0,0}, \ldots, \beta_{J_0,2^{J_0}-1}, \ldots, \beta_{J_1-1,2^{J_1-1}-1})$, and define

$$L_n = \sum_{\ell=1}^{n} (\mu_\ell - \widehat{\mu}_\ell)^2, \quad \widehat{\sigma}^2 = 2\sum_{\ell=(n/2)+1}^{n} \widetilde{\mu}_\ell^2,$$

where $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1, \ldots, \widetilde{\mu}_n) = (\widetilde{\alpha}_0, \ldots, \widetilde{\beta}_{J_1-1,2^{J_1-1}-1})$ and $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_n) = (\widehat{\alpha}_0, \ldots, \widehat{\beta}_{J_1-1,2^{J_1-1}-1})$, which is defined below. We consider the test procedure based on soft thresholding estimators:

$$\widehat{\alpha}_k = \widetilde{\alpha}_k, \quad \widehat{\beta}_{j,k} = sgn(\widetilde{\beta}_{j,k})(|\widetilde{\beta}_{j,k}| - \lambda_j)_+.$$

Here one can choose $\lambda_j = \widehat{\sigma}\sqrt{2\log n}/\sqrt{n}$, the universal threshold or the levelwise SureShrink rule which minimizes

$$S_n(\lambda_j) = \frac{\widehat{\sigma}^2}{n} 2^{J_0} + \sum_{j=J_0}^{J_1-1} S_j(\lambda_j),$$

where

$$S_j(\lambda_j) = \sum_{k=0}^{2^j-1} \left[\frac{\widehat{\sigma}^2}{n} - 2\frac{\widehat{\sigma}^2}{n} I_{\{|\widetilde{\beta}_{j,k}| \le \lambda_j\}} + \min(\widetilde{\beta}_{j,k}^2, \lambda_j^2) \right], \quad J_0 \le j \le J_1 - 1.$$

Genovese and Wasserman (2005) show that

$$\frac{\sqrt{n}(L_n - S_n(\widehat{\boldsymbol{\lambda}}))}{\sqrt{2}\widehat{\sigma}^2} \sim N(0,1).$$

In other words, reject $H_0 : \boldsymbol{\mu}^n = \mathbf{0}$ if

$$\sum_{\ell=1}^{n} \widehat{\mu}_\ell^2 > s_n^2(\widehat{\boldsymbol{\lambda}}) = \widehat{\sigma}^2 \frac{z_\alpha}{\sqrt{n/2}} + S_n(\widehat{\boldsymbol{\lambda}}), \tag{6}$$

which we shall refer to the Adaptive Pivotal Thresholding Test.

Although the computation of the power of the adaptive pivotal test is not straightforward, one can use the radius of confidence sets as a measure of power comparison. Consider a specific alternative $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Then the power of the pivot test is $\Pr_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}(\|\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}_{\widetilde{m}})\|^2 > s_n^2(\boldsymbol{\lambda}_{\widetilde{m}}))$. Therefore, a smaller radius of the confidence set indicates the higher power of the test based on that confidence set given $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$.

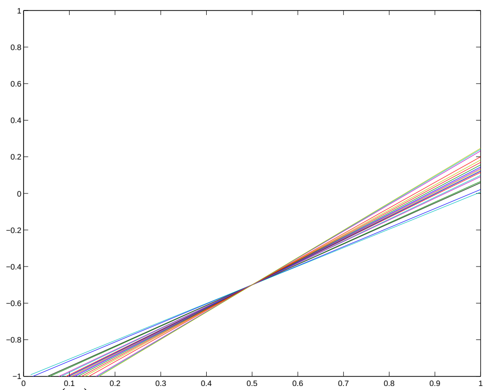
Summary

1. Clustering: Assign each curve to one of C clusters using Chiou and Li's functional clustering method.
2. Decorrelation: Subtract the estimated covariance structures based on PACE.
3. Test: Compute a p-value of H_{0i} : f_i =constant with the Adaptive Pivotal Thresholding Test.
4. Multiple comparison: Use the FDR controlling method for multiplicity adjustment.

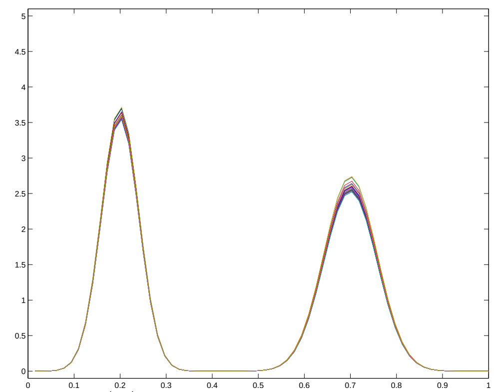
Figure 1: Algorithm of the proposed method

Beran and Dümbgen (1998) show that $s_n(\boldsymbol{\lambda}_{\tilde{m}})$ is of order $O(n^{-1/4})$ which is much faster than the convergence rate of the radius of the χ^2 ball. It is not straightforward how to compare $s_n(\boldsymbol{\lambda}_{\tilde{m}})$ with the convergence rates of the adaptive Neyman test, but Li (1989) shows that $O(\sigma_n n^{1/4}) = O(n^{-1/4})$ is the fastest possible convergence rate and also Ingster and Suslina (2003) show that $O(\sigma_n n^{1/4})$ is a critical rate that determines if the alternative is detectable. Hence we expect the adaptive pivotal test to outperform or at least be comparable with the adaptive Neyman test. Similar arguments can be applied to the comparison between Fan's thresholding test and our adaptive pivotal thresholding test. In our numerical analysis we use the adaptive pivotal thresholding test in (6) that can take advantage of wavelet representation.

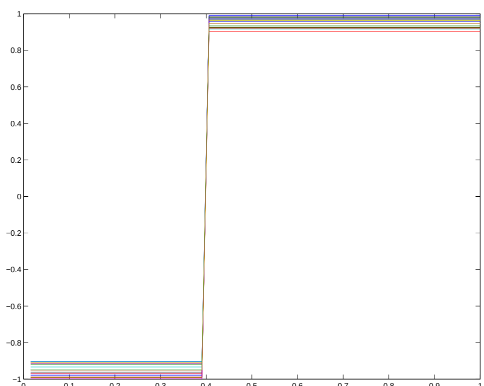
When a large number of tests are performed as in our case study, one must adjust the multiplicity of tests. Standard procedures such as Bonferonni method are very conservative in large scale testing problems to account for the multiplicity. Benjamini and Hochberg (1995) propose to control FDR which is defined as the mean of the number of false rejections divided by the number of total rejections. In this paper we use FDR since it provides a reasonable criterion especially when there are many tests



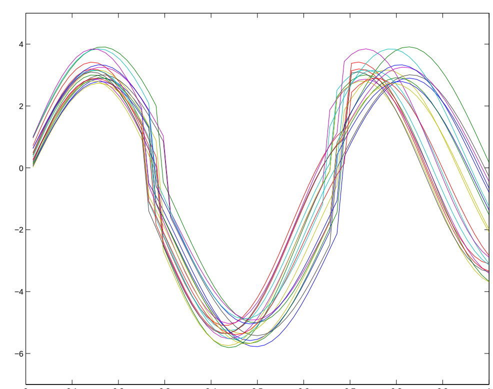
(a) Example 1: Exponential



(b) Example 2: Linear



(c) Example 3: Mean shift



(d) Example 4: Heavisine

Figure 2: Four simulated curves: (a) exponential, (b) linear, (c) mean shift and (d) heavisine trends. The curves are jittered so that each curve has a slightly different mean trend. The explicit forms are given in the text.

compared to traditional methods. Figure 1 summarizes our data analysis strategy.

3 Simulated Examples

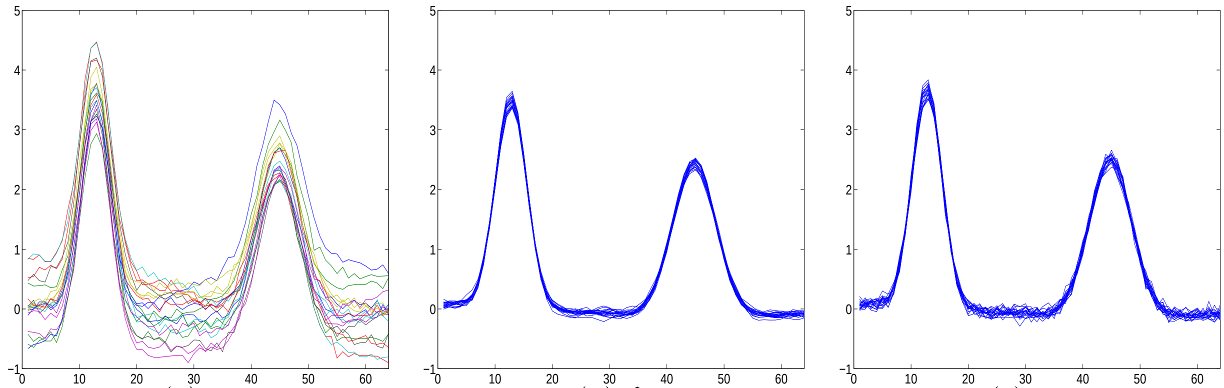
We consider four curve examples shown in Figure 2, and three error structures. The curves are given as

$$
\begin{aligned}
f_1(t) &= (1+0.5U)(t-0.5)-0.5, \\
f_2(t) &= (3.5+0.3U)\exp(-300(t-0.2)^2)+(2.5+0.3U)\exp(-150(t-0.7)^2), \\
f_3(t) &= -(0.9+0.1U)I(t\leq 0.4)+(0.9+0.1U)I(t>0.4), \\
f_4(t) &= U-0.5+(3+0.5U)\sin((3+0.5U)\pi t) \\
&\quad -sgn(t-(.25+.05U))-sgn((.65+.05U)-t),
\end{aligned}
$$

where U follows a uniform distribution $(0,1)$.

For each example, we generate 100 curves for each of 100 iterations. The number of grid points in each function is $n=64$. At each iteration, 20 curves have nonzero mean ("signal group") and 80 with zero mean function ("no signal group"). Note that the signal curves are jittered so that each curve has a slightly different mean trend. We add three different errors to the functions: independent normal errors, AR(1) with coefficient 0.5, and non-stationary errors. In the case of non-stationary error, we follow the simulation setting in Chiou and Li (2007); In (2), we consider eigenfunctions $\phi_1(t)=\sqrt{2}\sin(\pi t), \phi_2(t)=\sqrt{2}\cos(\pi t)$ with $\lambda_1=.8, \lambda_2=.6$ for the signal group, and $\phi_1(t)=\sqrt{2}\sin(2\pi t), \phi_2(t)=\sqrt{2}\cos(2\pi t)$ with $\lambda_1=.4, \lambda_2=.3$ for the no signal group. We note that the eigenfunctions from the variable stars exhibit sinusoidal shapes as shown in Figure 6. For each simulation setting, the signal to noise ratio is set to be five.

For the clustering step we find that the correct clusters (signal vs. no signal) are always identified by our procedure for the simulated examples. To remove the possible non-stationary error structure we apply the PACE procedure described in



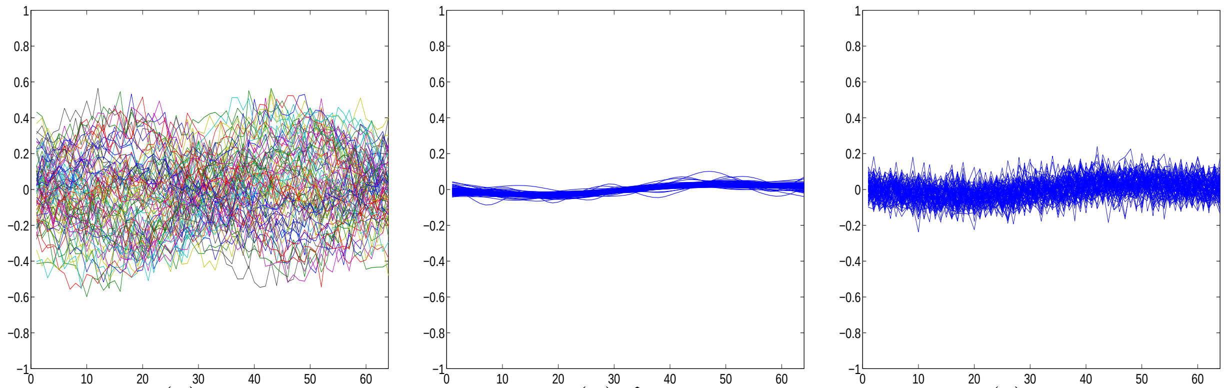
(a) Y_i (b) $\widehat{\mu}_i$ (c) Y_i^*

Figure 3: The decorrelation procedure for the signal group in Example 2. In the figures, (a) shows the observed data with the non-stationary error, (b) estimated mean curves and (c) the decorrelated curves subject to the adaptive pivotal thresholding test.

Section 2.2. For illustration purposes, we take Example 2 and show the observed data with the non-stationary error, estimated mean curves, and the decorrelated curves subject to the adaptive pivotal thresholding test in Figures 3 (signal group) and 4 (no signal group). Both figures show that the PACE procedure successfully removes the non-stationary error.

For the trend detection we consider three different methods; 'AN' stands for Fan's adaptive Neyman test in (4), 'WAV' for Fan's thresholding test in (5) , and 'PIV' for the adaptive pivotal thresholding method in (6). To summarize the comparisons of the three methods we use the following measures:

- False Discovery Rate (FDR) $= P(H_0$ is true $\mid$ Reject $H_0)$
- False Negative Rate (FNR) $= P(H_0$ is not true $\mid$ Accept $H_0)$
- Specificity $= P($Accept $H_0 \mid H_0$ is true$)$
- Sensitivity $= P($Reject $H_0 \mid H_0$ is false$)$.



(a) Y_i (b) $\widehat{\mu}_i$ (c) Y_i^*

Figure 4: The decorrelation procedure for the no signal group in Example 2. In the figures, (a) shows the observed data with non-stationary error, (b) estimated mean curves and (c) the decorrelated curves subject to the adaptive pivotal thresholding test.

One may consider FNR as an extension of type II error in multiple comparison. For more details about the justification of using FNR, see Genovese and Wasserman (2002). In our simulation study, we set the FDR level to 0.15.

Table 1 reports the average numbers out of 100 repetitions of these measures for the first curve example. We compare the three methods with and without the PACE procedure to assess the effect of decorrelation. In the table it can be seen that only the proposed PIV method controls FDR under 0.15 and it produces the highest Specificity for all the three error types when the proper procedure is conducted. The WAV method performs better than AN, but it fails to control FDR. The advantage of the PACE procedure is most apparent for the non-stationary errors. As expected, its contributions for AR(1) and independent errors are not so drastic. Since we discover the similar patterns for the rest of the curve examples, in the following tables we report only the result with the PACE procedure for the non-stationary errors and without the PACE procedure for the independent and AR(1) errors. As for FNR and Sensitivity, all three methods yield nearly perfect values, 0 and 1 respectively, and we

Table 1: Summary of Measures (Example 1)

Error	Method	PACE	FDR	FNR	Specificity	Sensitivity
Indep.	AN	No	0.3230	0	0.8754	1
		Yes	0.3218	0	0.8761	1
	WAV	No	0.2145	0	0.9291	1
		Yes	0.2117	0	0.9303	1
	PIV	No	0.0356	0	0.9903	1
		Yes	0.0348	0	0.9564	1
AR(1)	AN	No	0.3325	0	0.8690	1
		Yes	0.3148	0	0.8789	1
	WAV	No	0.2153	0	0.9281	1
		Yes	0.1683	0	0.9470	1
	PIV	No	0.0302	0	0.9918	1
		Yes	0.0277	0	0.9926	1
Nonst.	AN	No	0.7982	0	0.0110	1
		Yes	0.4819	0	0.6506	1
	WAV	No	0.7961	0	0.2380	1
		Yes	0.3381	0	0.8218	1
	PIV	No	0.7956	0	0.0266	1
		Yes	0.1412	0.0001	0.9116	0.9995

observe the same phenomenon in the rest of the examples.

Tables 2–4 report the results from the second, third, and fourth examples, respectively. The results show that the proposed PIV method properly controls FDR, at the level of less than 0.15, for each of the three error structures. The AN and WAV yield larger FDRs than 0.2 in all cases, which indicates that they fail to control FDR. As Fan (1996) points out, the convergence rates of those tests are slow, and thus he recommended to compute p-values using finite sampling distributions of test statistics based on simulations. This may explain failures of the Fan's tests in controlling FDR. The use of the simulation tables for our analysis is not ideal since they only provide ranges of p-values while we need *exact* p-values to control FDR. As for Specificity, the AN produces the lowest and the PIV produces the highest values (always > 0.9). The simulation results support the conjecture that the proposed method outperforms AN and WAV under various curve examples and error structures while correctly controlling FDR levels.

Table 2: Summary of Measures (Example 2)

Error	Method	FDR	FNR	Specificity	Sensitivity
Indep.	AN	0.3363	0	0.8679	1
	WAV	0.2298	0	0.9221	1
	PIV	0.0346	0	0.9906	1
AR(1)	AN	0.3299	0	0.8718	1
	WAV	0.2240	0	0.9239	1
	PIV	0.0291	0	0.9921	1
Nonst.	AN	0.4795	0	0.6529	1
	WAV	0.3361	0	0.8244	1
	PIV	0.1379	0	0.9139	1

Table 3: Summary of Measures (Example 3)

Error	Method	FDR	FNR	Specificity	Sensitivity
Indep.	AN	0.3291	0	0.8715	1
	WAV	0.2102	0	0.9303	1
	PIV	0.0254	0	0.9931	1
AR(1)	AN	0.3095	0	0.8829	1
	WAV	0.2242	0	0.9244	1
	PIV	0.0279	0	0.9925	1
Nonst.	AN	0.4818	0	0.6529	1
	WAV	0.3365	0	0.8246	1
	PIV	0.1397	0	0.9128	1

In what follows we investigate how sensitive the three methods are to the proportion of curves with nonzero mean function. Note that this proportion was 0.2 in the previous experiment. We further conduct a simulation with Example 2 using the proportion=0 (i.e., all 100 curves belong to a no signal group) and 0.5 (i.e., 50 curves belong to a no signal group and the other 50 curves belong to a signal group) and report the results in Table 5.

When the proportion is 0, we examine type I error of the three methods with

Table 4: Summary of Measures (Example 4)

Error	Method	FDR	FNR	Specificity	Sensitivity
Indep.	AN	0.3210	0	0.8780	1
	WAV	0.2126	0	0.9295	1
	PIV	0.0317	0	0.9914	1
AR(1)	AN	0.3383	0	0.8660	1
	WAV	0.2256	0	0.9229	1
	PIV	0.0283	0	0.9924	1
Nonst.	AN	0.4795	0	0.6529	1
	WAV	0.3361	0	0.8244	1
	PIV	0.1379	0	0.9139	1

$\alpha = 0.05$. While AN and WAV fail to control type I error for all three error structures, PIV succeeds to control it under 0.05 in all cases. When the proportion is 0.5, FDR is set to be 0.15. It can be seen that AN and WAV achieve FDR less than 0.15 for both independent and AR(1) errors, which is an improvement compared to proportion=0.2 where they both fail. PIV also controls FDR and produces smaller values compared to the previous experiment. In terms of Specificity, PIV produces the highest value for both independent and AR(1) errors. For the non-stationary errors, although FDRs of AN and WAV drop significantly compared to proportion=0.2, they still cannot control FDR as their values are greater than 0.15. On the contrary, PIV again successfully controls FDR and also produces the highest Specificity. We obtain similar results for the other three examples and do not report them to save space.

4 Analysis of Long Period Variable Stars

We analyze the variable star data with the proposed method and compare our result with Hart et al. (2007) in Section 4.1. The negative association between the mean period and the strength of trend, previously identified by Hart et al. (2007) is re-confirmed with a SiZer (SIgnificant ZERo crossing of the derivatives, Chaudhuri and

Table 5: Summary of Measures with different proportions of nonzero mean curves (Example 2)

		prop.=0	prop.=0.5			
Error	Method	Type I	FDR	FNR	Specificity	Sensitivity
Indep.	AN	0.1315	0.1477	0	0.8238	1
	WAV	0.0843	0.0953	0	0.8930	1
	PIV	0.0113	0.0141	0	0.9854	1
AR(1)	AN	0.1297	0.1478	0	0.8230	1
	WAV	0.0814	0.0940	0	0.8939	1
	PIV	0.0115	0.0149	0	0.9846	1
Nonst.	AN	0.2759	0.2854	0	0.5272	1
	WAV	0.1426	0.2070	0	0.6840	1
	PIV	0.0422	0.1320	0	0.7780	1

Marron (1999)) analysis in Section 4.2. Section 4.3 gives a extension of the previous analysis by providing a list of variable stars with "non-linear" trend.

4.1 Variable Stars with Trend

The 378 variable star curves in the data are divided into clusters and decorrelated by the procedures presented in Section 2. Here we present the result with the number of clusters $C = 3$, since when $C = 2$, the algorithm assigns all data to one cluster. When $C = 4$, the result is similar to $C = 3$ and the size of one cluster is too small (38) to properly estimate the covariance structure.

In order to verify the existence of clusters in the star data, we study between and within cluster variability. For each curve, we convert $\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i$ to vectors of length 300 by observing values at 300 regularly spaced grid points between 0 and 1. Then we calculate all pairwise distance between 378 vectors. We define the between-clusters variability as the sum of pairwise distances between clusters and the within-cluster variability as the sum of pairwise distances within a cluster. Then, for the current clustering solution, we calculate the ratio of between and within variability

Table 6: Summary of variable star data analysis

	Hart et al		
PIV	Trend	No trend	Total
Trend	16	8	24
No trend	37	317	354
Total	53	325	378

and compare it with the ratios from 10,000 random clustering solutions with the same size allocation as the current one. From this we obtain an empirical significance probability .0001 by calculating the proportion of clustering solutions with a higher ratio than the current one.

Since parts of the data are irregularly spaced due to missing epochs, we interpolate the missing values by applying the local linear smoothing to $Y_i^*(t)$ in Step E4. Also, since the numbers of epochs are not the power of two, we use the adaptive lifting scheme in nonparametric wavelet regression proposed by Nunes et al. (2006).

Hart et al. (2007) find that 56 out of the 378 stars have significant trends in the times between maximum brightness, controlling the pFDR level at 0.05 for multiplicity adjustment. For a fair comparison, however, we apply the FDR method at the level of 0.05 to their p-values and find 53 significant trends. In our analysis, on the other hand, we find that 24 stars have significant trends. Table 6 summarizes the difference between our results and those found by Hart et al. (2007).

Figure 5 shows some curves for which Hart et al. (2007) and the proposed method do not agree. Figures 5(a) and (b) are examples where Hart et al. (2007) declare as displaying significant trends but we do not. Note that the curves for these two stars both display sinusoidal trends, and note further that these types of trend are flagged as significant in Hart et al. (2007). The proposed method, on the other hand, does not flag them as significant and explains the sinusoidal appearance as part of the non-stationary covariance error. This claim can be supported by Figure 6 where

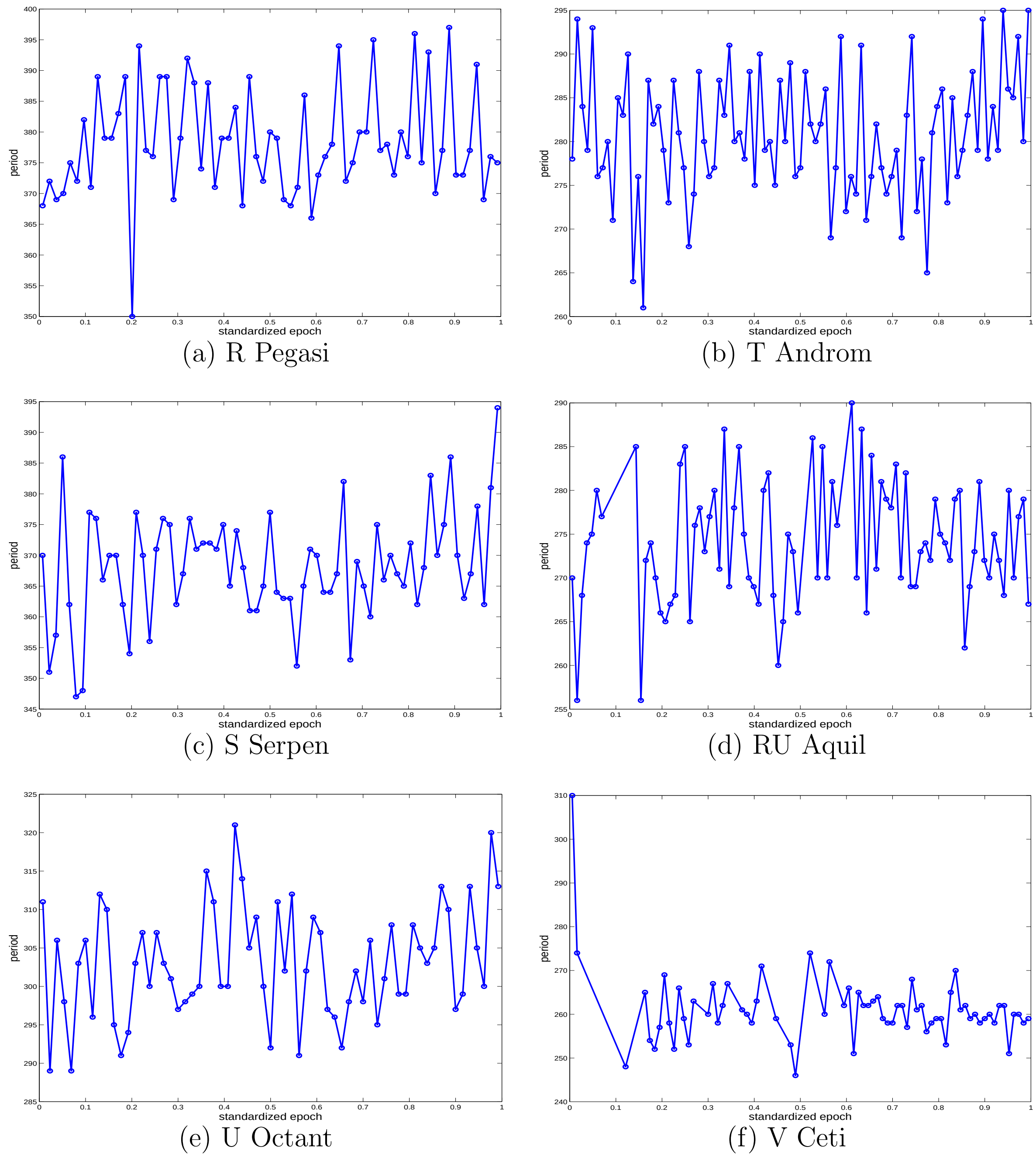
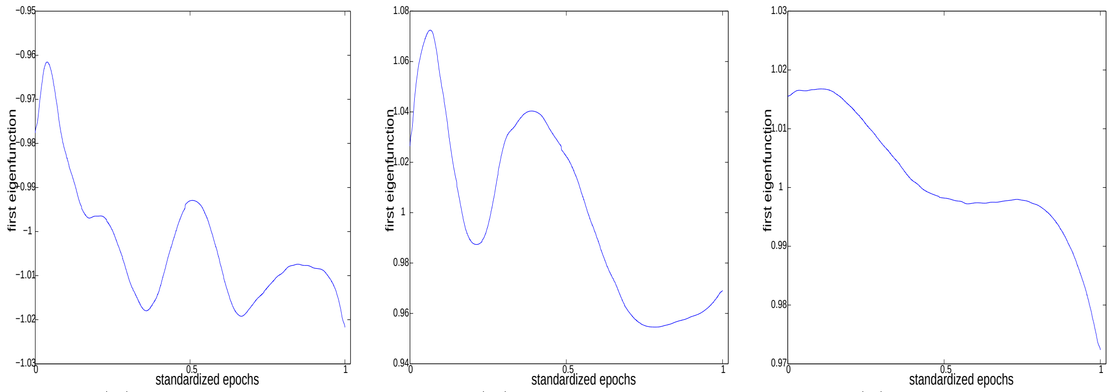


Figure 5: Examples where the two methods do not coincide. (a) and (b): examples where Hart et al. (2007) declare as displaying significant trends but the proposed method does not. (c)–(f): examples where the proposed method declares as displaying significant trends but Hart et al. (2007) do not.



(a) cluster 1 (b) cluster 2 (c) cluster 3

Figure 6: The first eigenfunction of each cluster of the proposed method for the variable stars data. They all show the sinusoidal shape.

the first eigenfunctions of the covariance estimate of three clusters are displayed. All three functions show the sinusoidal shape. Since Hart et al. (2007) use a parametric approach for estimating the non-stationary covariance structure, there is a possibility that their error model misses some particular forms of the dependent structure.

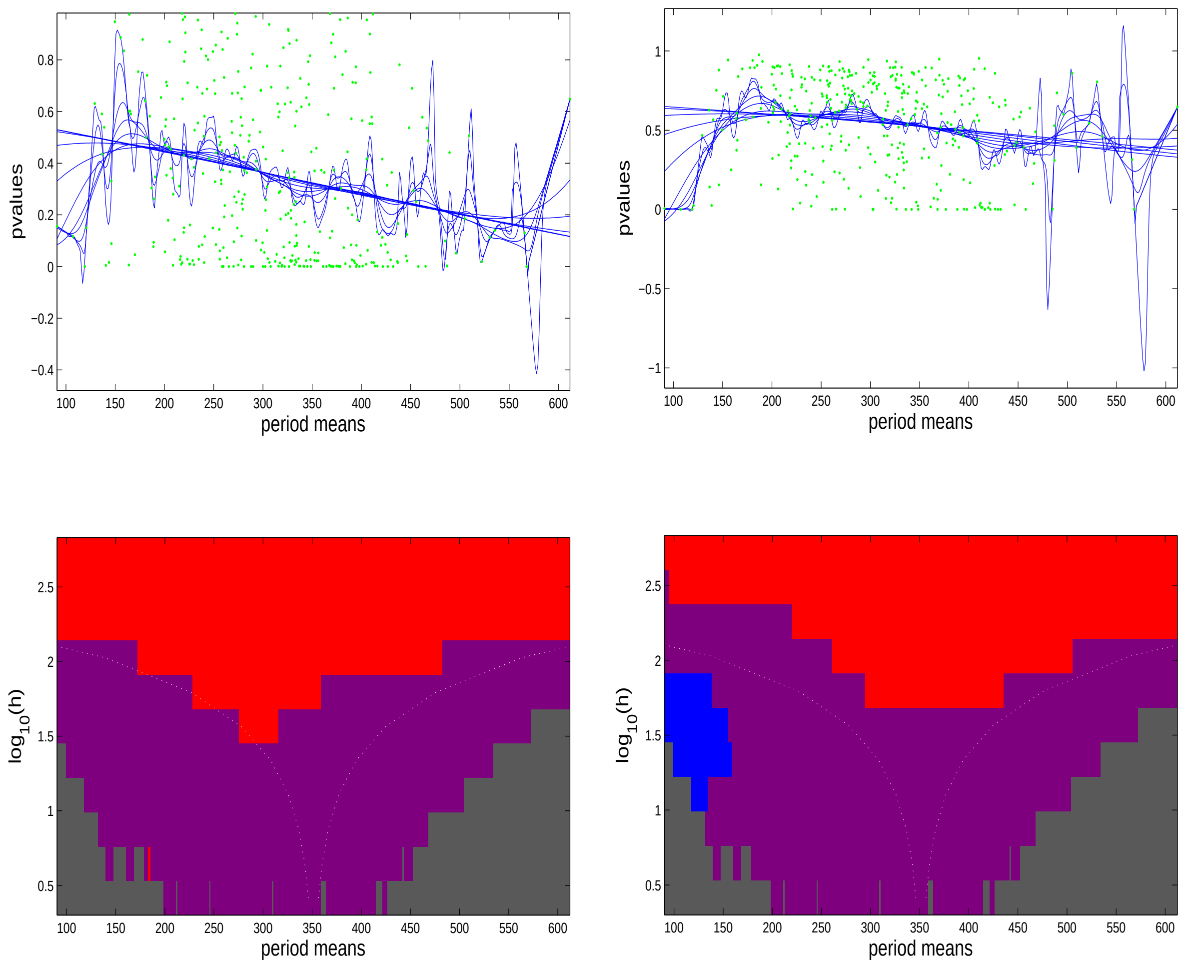
Conversely, Figures 5(c)–(f) show exemplary curves where the proposed method declares as displaying significant trends but Hart et al. (2007) do not. Note that Figures 5(c) and (e) show increasing trends and Figure 5(d) displays a mean shift – both of which do not seem to be explicable in terms of non-stationary error structure. Moreover, Figure 5(f) contains a large outlier at the first standardized epoch and this seems to drive the proposed method to declare the trend as significant. One should further investigate whether or not this first data point is an important observation.

4.2 Revisited Association of Mean Period and the Strength of Trend

Hart et al. (2007) investigate the relationship between mean period and the strength of trend indicated by p-values using the order selection test, and find that longer period stars tend to have smaller p-values. We perform a similar test using a SiZer analysis.

Figure 7(a) shows the SiZer plot of the p-values versus period means, using the results in Hart et al. (2007). The thin blue curves in the top panel display the family of smoothed curves obtained from the data points (which are shown in green). These curves are the local linear estimates using different bandwidths. Some of the blue curves are oversmoothed, and others are undersmoothed. The goal of SiZer is to determine which features of the blue curves are important and which are just spurious sampling artifacts. The lower panel shows the SiZer map of the data. The horizontal locations in the SiZer map are the same as in the top panel, and the vertical locations correspond to the logarithm of bandwidths of the family of smoothed curves shown as the blue curves in the top panel. Each pixel shows a color that gives the result of a hypothesis test for the slope of the blue curve, at the point indexed by the horizontal location, and at the bandwidth corresponding to that row. When the slope of the blue curve is significantly positive (negative), the pixel is colored blue (red, respectively). When the slope is not significant – which means the sampling noise is dominant – the pixel has the intermediate color of purple. If there are not sufficient data points for the test to be carried out, the corresponding pixel is colored gray.

Figure 7(a) displays a decreasing trend in the family of smoothed curves, and according to the corresponding SiZer map this global trend is statistically significant at coarse scales. The SiZer plot based on our analysis, shown in Figure 7(b), reveals a consistent result except at the beginning. The family of smoothed curves suggests an increasing trend at small scales but a decreasing trend at large scales. This is confirmed by the SiZer map where it shows blue at small scales and red at large scales. For the rest of the region, a global decreasing trend at large scales appears in the map. Therefore, based on the SiZer analysis, we reconfirm the negative association



(a) Hart et al. (b) PIV

Figure 7: SiZer plots of p-values versus period means. The top panels display the family of local linear estimates with different bandwidths and the bottom panels show the significance of trends of the blue curves in the top panels.

between p-values and mean periods that is suggested by Hart et al. (2007).

Hart et al. (2007) comment that, because of the very short (in astronomical terms) total time span of the AAVSO observations considered in their analysis, the association which they tentatively identify between p-values and mean period will *not* be due to systematic evolutionary changes in the stars, but instead caused by small secular changes in their structure – which they speculate might be related to the effects of mass loss, since the mass loss rate is found to be linearly correlated with mean period for Miras with mean period greater than about 400 days (see e.g. Groenewegen et al. (1998) and references therein).

4.3 Variable Stars with Nonlinear Trend

As a further extension of the previous analyses we take the light curves of the 24 stars which show significant trends in period and carry out a test for non-linearity in the trend after decorrelation. Understanding the nonlinear pulsation of long period variables is an important aspect of their study (see e.g. Wood (2006) and references therein). For example, a long-standing problem has been that nonlinear pulsation models of red giant stars exhibited higher-amplitude behavior than real stars – although the addition of turbulent viscosity has improved the situation, as shown by e.g. Olivier and Wood (2005). A better understanding of the impact of non-linear effects, including non-linear trends in period, in the pulsation of Miras could bring insight into e.g. the role and nature of convection on energy transport within their atmospheres, as well as their evolution and mass loss rates.

To carry out the test for non-linearity we first do decorrelation with PACE for each light curve and then apply the test of non-linearity effect in Generalized Additive Models (Chambers and Hastie, 1991). Table 7 shows the results of applying this non-linearity test: only 8 of the 24 variable stars are identified as showing period trends

Trend	Stars
Linear	R Aql, R Cas, R Cyg, RU Aql, RU Her, S Ser, U Oct, W Dra
Non-linear	Z Aur, R Aur, R Hya, R Lep, RU Tau, RX Tau, S Ori, Ss Her, S Umi, T Cep, T Ser, TY Cyg, V Cet, W Eri, W Vel, Z Vel

Table 7: Summary of analysis of linear vs non-linear period trends

which are consistent with linearity; for the remaining 16 variable stars the linear hypothesis is rejected. Therefore, we see that the role of nonlinear effects on period evolution would appear to be important for the majority of the variable stars under study for which there is evidence of a period trend.

5 Concluding Remarks

In this work we analyze time series of the periods between maximum brightness of a group of 378 long period variable stars. We identify the stars which display certain trends in their period, via multiple testing of a mean for a non-stationary time series model. Specifically we propose the adaptive pivotal test and thresholding pivotal test in order to carry out the significance test of no trend when data are curves. These tests are due to Stein-Beran-Dümbgen's pivotal strategy for construction of nonparametric confidence sets in infinite-dimensional normal mean problems. We also adapt the approach of PACE with functional clustering for decorrelation of non-stationary time series data. Our simulation study indicates that the proposed method outperforms, or at least is comparable to, other well known methods including the adaptive Neyman and wavelet thresholding tests. We apply our method to the variable star data and compare our results with the previous analysis. We revisit the relationship between the p-values and mean period and confirmed their relationship with a SiZer analysis and provide a list of variable starts with nonlinear trend.

As Fan (1996) points out, there is a connection between a multivariate normal model with a goodness of fit test and our pivotal method. The proposed method can be used as omnibus lack-of-fit tests. Future study of comparisons with other methods such as in Hart (2009) is needed.

As we mention at the end of Section 2.3, one can use the radius of confidence sets as a measure of power comparison. Beran and Dümbgen (1998) show that the REACT method achieves the convergence rate of order $O(n^{-1/4})$. Since our method is based on the REACT approach, we believe the radius of our test also achieve the same convergence rate. Furthermore, according to Li (1989), the best convergence rate of the radius is $O(n^{-1/4})$ and our method achieves this rate. However, the radius of Fan's test depends on m_0 whose order is unknown, so it is not straightforward to compare his methods with ours. Hence we expect our method is at least comparable to Fan's. Computing the exact power is beyond the scope of this paper and we leave it as future work.

The methodology developed and applied to the specific astrophysical problem of identifying period trends in the light curves of variable stars in this paper has wider potential applications in many other, diverse areas of astrophysics – including, for example: characterization of the light curves of type Ia supernovae and gamma ray bursts; continuum subtraction in the analysis of galaxy and quasar spectra; investigation of the time evolution of noise in ground-based interferometric gravitational wave detectors. In this last application, in particular, the fact that the proposed method takes a nonparametric approach to characterizing non-stationarity in the covariance structure of the noise could be very important for analysis of e.g detector 'glitches', where a parametric form for the noise is less well understood. We will explore some of these potential applications in future work.

Acknowledgements

We are grateful to Professor Jeffrey Hart for kindly sharing his analysis results for comparisons in this paper. We are also thankful to the reviewers and the Associate Editor for their constructive and helpful comments. Park's work was supported in part by National Security Agency Grant No. H982300810056 and Ahn's work was partly supported by NSF grant DMS-0805758.

References

Abramovich, F., Antoniadis, A., Sapatinas, T. and Vidakovic, B. (2004), "Optimal testing in a fixed-effect functional analysis of variance model," *International Journal of Wavelets, Multiresolution and Information Processing*, **2** 323-349.

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,* Ser. B, **57** 289–300.

Beran, R. and Dümbgen, L. (1998), "Modulation of estimators and confidence sets," *The Annals of Statistics*, **26** 1826–1856.

Chambers, J. and Hastie, T.(1991) *Statistical Models in S* Wadworth, Pacific Grove.

Chaudhuri, P. and Marron, J. S. (1999), "SiZer for exploration of structures in curves," *Journal of the American Statistical Association*, **94** 807–823.

Chiou, J.-M. and Li, P.-L. (2007), "Functional clustering and identifying substructures of longitudinal data," *Journal of the Royal Statistical Society*, Ser. B, **69** 679–699.

Cox, D. D. and Lee, J. S. (2008), "Pointwise testing with functional data using the Westfall-Young randomization method," *Biometrika*, **95** 621–634.

Eddington, A. S. and Plakidis, S. (1929), "Irregularities of period variable starts," *Monthly Notices of the Royal Astronomical Society*, **90** 65–71.

Fan, J. (1996), "Test of significance based on wavelet thresholding and Neyman's truncation," *Journal of the American Statistical Association*, **91** 674–688.

Fan, J., Huang, T. and Li, R. (2007), "Analysis of Longitudinal Data With Semiparametric Estimation of Covariance Function," *Journal of the American Statistical Association*, **102** 632–641.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Fan, J. and Lin, S.-K. (1998), "Test of significance when data are curves," *Journal of the American Statistical Association*, **93** 1007–1021.

Genovese C. R. and Wasserman, L. (2002), "Operating characteristics and extensions of the false discovery rate procedure," *Journal of the Royal Statistical Society*, Ser. B, **64** 499–517.

Genovese C. R. and Wasserman, L. (2005), "Confidence sets for nonparametric wavelet regression," *The Annals of Statistics*, **33** 698–729.

Genton M. G. and Hall, P. (2007), "Statistical inference for evolving periodic functions," *Journal of the Royal Statistical Society*, Ser. B, **69** 643–657.

Groenewegen, M. A. T., Whitelock, P. A. Smith, C. H. and Kerchsbaum, F. (1998), "Dust shells around carbon Mira variables," *Monthly Notices of the Royal Astronomical Society*, **293** 18–42.

Hall, P., Reimann, J. and Rice, J. (2000), "Nonparametric estimation of a periodic function," *Biometrika*, **87** 5435–557.

Hart, J. D. (2009), "Frequentist-Bayes lack of fit test based on the Laplace approximation," *Journal of Statistical Theory and Practice*, **3** 681–704.

Hart, J. D., Koen, C. and Lombard, F. (2007), "An analysis of pulsation periods of long period variable stars," *Journal of the Royal Statististcal Society*, Ser. C, **56** 587–606.

Ingster, Y. I. and Suslina, I. A. (2003), *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, New York: Springer.

James, M. and Stein, C. M. (1960), "Estimation with quadratic loss," In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1** 361–379, Berkeley: University of California.

Kanbur, S. M., Hendry, M. A. and Clarke, D. (1997), "Period-luminosity and period-luminosity-color relations for Mira variables at maximum light," *Monthly Notices of the Royal Astronomical Society*, **289** 428–442.

Koen, C and Lombard, F (2001), "The analysis of indexed astronomical time series - VII: Simultaneous use of times of maxima and minima to test for period changes in long-period variables," *Monthly Notices of the Royal Astronomical Society*, **325** 1124–1132.

Koen, C and Lombard, F (2004), "The analysis of indexed astronomical time series - IX: a period change test," *Monthly Notices of the Royal Astronomical Society*, **353** 98–104.

Li, K.-C. (1989), "Honest confidence regions for nonparametric regression," *The Annals of Statistics*, **17** 1001–1008.

Mattei, J. A., Mayall, M. W. and Waagen E. O. (1990), *Maxima and Minima of Long Period Variables, 1949-1975*, Cambridge: American Association of Variable Star Observers.

Muirhead, C. R. (1986), "Distinguishing outlier types in time series," *Journal of the Royal Statistical Society*, Ser. B, **48** 39–47.

Nunes, M. A., Knight, M. I. and Nason, G. P. (2006), "Adaptive lifting for nonparametric regression," *Statistics and Computing*, **16** 143–159.

Oh, H.-S., Nychka D., Brown, T. and Charbonneau, P. (2004), "Period analysis of variable stars by robust smoothing," *Journal of the Royal Statistical Society*, Ser. C, **53** 15–30.

Olivier, E. A. and Wood, P. R. (2005), "Non-linear pulsation models of red giants." *Monthly Notices of the Royal Astronomical Society*, **263** 1396–1412.

Percy, J. R. and Colivas, T. (1999), "Long-term changes in Mira stars: I, period fluctuations in Mira Stars," *Publications of the Astronomical Society of the Pacific*, **111**, 94–97.

Percy, J. R., Colivas, T., Sloan, W. B. and Mattei, J. (1990), "Long-term changes in Mira variables," *Astronomical Society of the Pacific Conference Series*, **11** 446–449.

Popper, D. M. (1967), "Determination of masses of eclipsing binary stars," *Annual Review of Astronomy and Astrophysics*, **5** 85–104.

Ruppert, D., Sheather, S.J. and Wand, M.P. (1995), "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, **90** 1257–1270.

Shen, Q. and Faraway, J. (2004), "An F test for linear models with functional responses," *Statistica Sinica*, **14** 1239–1257.

Stein, C. M. (1981), "Estimation of the mean of a multivariate normal distribution," *The Annals Statistics*, **9** 1135–1151.

Storey, J. D. (2002), "A direct approach to false discovery rates," *Journal of the Royal Statistical Society*, Ser. B, **64** 479–498.

Wasserman, L. A. (2005), *All of Nonparametric Statistics*, New York: Springer.

Whitelock, P. A. (1999), "Real Time evolution in Mira variables," *New Astronomy Review*, **43** 437–440.

Wood, P.R.(2006), "Long-period variables," *Memorie della Societa Astronomica Italiana* **77** 76–80.

Wornell, G. W.(1996), *Signal Processing with Fractals: A Wavelet Based Approach*, Prentice Hall, Eaglewood Cliffs.

Yao, F., Müller, H.- G. and Wang, J.-L. (2005), "Functional data analysis for sparse longitudinal data," *Journal of the American Statistical Association*, **100** 577–590.