Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Variance Change Point Detection Under a Smoothly-Changing Mean Trend with Application to Liver Procurement

Zhenguo Gao, Zuofeng Shang, Pang Du & John L. Robertson

To cite this article: Zhenguo Gao, Zuofeng Shang, Pang Du & John L. Robertson (2019) Variance Change Point Detection Under a Smoothly-Changing Mean Trend with Application to Liver Procurement, Journal of the American Statistical Association, 114:526, 773-781, DOI: 10.1080/01621459.2018.1442341

To link to this article: https://doi.org/10.1080/01621459.2018.1442341

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 114, NO. 526, 773–781: Theory and Methods
https://doi.org/10.1080/01621459.2018.1442341

Variance Change Point Detection Under a Smoothly-Changing Mean Trend with Application to Liver Procurement

Zhenguo Gao[a], Zuofeng Shang[b], Pang Du[a], and John L. Robertson[c]

[a]Department of Statistics, Virginia Tech, Blacksburg, VA; [b]Department of Mathematical Sciences, IUPUI, Indianapolis, IN; [c]School of Biomedical Engineering, Virginia Tech, Blacksburg, VA

ABSTRACT
Literature on change point analysis mostly requires a sudden change in the data distribution, either in a few parameters or the distribution as a whole. We are interested in the scenario, where the variance of data may make a significant jump while the mean changes in a smooth fashion. The motivation is a liver procurement experiment monitoring organ surface temperature. Blindly applying the existing methods to the example can yield erroneous change point estimates since the smoothly changing mean violates the sudden-change assumption. We propose a penalized weighted least-squares approach with an iterative estimation procedure that integrates variance change point detection and smooth mean function estimation. The procedure starts with a consistent initial mean estimate ignoring the variance heterogeneity. Given the variance components the mean function is estimated by smoothing splines as the minimizer of the penalized weighted least squares. Given the mean function, we propose a likelihood ratio test statistic for identifying the variance change point. The null distribution of the test statistic is derived together with the rates of convergence of all the parameter estimates. Simulations show excellent performance of the proposed method. Application analysis offers numerical support to non invasive organ viability assessment by surface temperature monitoring. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2017
Revised January 2018

KEYWORDS
Asymptotic null distribution; Change point consistency; Hypothesis testing in nonparametric smoothing; Smoothly changing mean trend; Variance change point

1. Introduction

Change point detection is a topic that has attracted significant attention for decades. Efforts have mostly focused on detection of sudden changes in a few parameters, such as the mean and/or variance, of the underlying distribution, or the distribution itself as a whole entity. In this article, we are concerned with variance change point detection under a smoothly changing mean trend. In particular, a constantly changing mean trend violates the assumptions of most existing change point detection methods. As demonstrated in the article, the naïve application of these existing methods to such kind of data would yield erroneous change point estimates.

Our method is motivated by experiments involving the procurement of transplant livers. Quality/viability evaluation is a key issue in the procurement of transplantable organs. Currently, such evaluations are usually performed through visual inspection of potentially transplantable organs by surgeons and organ recovery personnel, and biopsy image assessment by pathologists. Both approaches involve subjective judgements and can lead to unnecessary organ discards. Biopsy is more accurate than surgeons' visual inspection, but it is also invasive and damages the part of the organ where the biopsy sample is collected. Additionally, the morphology of the biopsy sample may not represent that of the whole organ, and may result in potentially transplantable organs being discarded. In the experiment considered in the article, the surface temperature of an explanted porcine liver was constantly monitored during the infusion of a temperature-controlled perfusion liquid. Experimental data consisted of surface temperatures, obtained using high-resolution infrared thermography, and measured every 10 minutes on a dense optical grid covering the whole organ for a span of 24 hours. The left panel in Figure 1 demonstrates the temperature profile for a randomly selected spot on the surface. The temperature of the perfusion liquid (4°C) was lower than normal body temperature and the temperature of the organ changed slowly during perfusion and displayed an overall smooth mean trend. The high temperature oscillations noted in the first half of perfusion reflected the response of the organ to temperature change from the ambient environment and infusion of perfusion fluid. By the 10th hour of perfusion, the organ started to lose its viability and this change was reflected in a sudden drop in the variance of the temperature, as shown in the plot of residuals versus time in the right panel of Figure 1. Our goal was to design a testing procedure for identifying the variance change point of the residuals after removing the smoothly changing mean trend.

We note that the phenomenon of having a variance change point underlying a smooth mean trend actually occurs in many application settings. For example, seismic activity monitoring often detects a smooth mean trend with small variation and a sudden change in variation could be the early sign of an earthquake. The electroencephalographic (EEG) signal from

CONTACT Pang Du pangdu@vt.edu Department of Statistics, Virginia Tech, 417-A Hutcheson Hall (MC 0439), 250 Drillfield Drive, Blacksburg, VA 24061.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

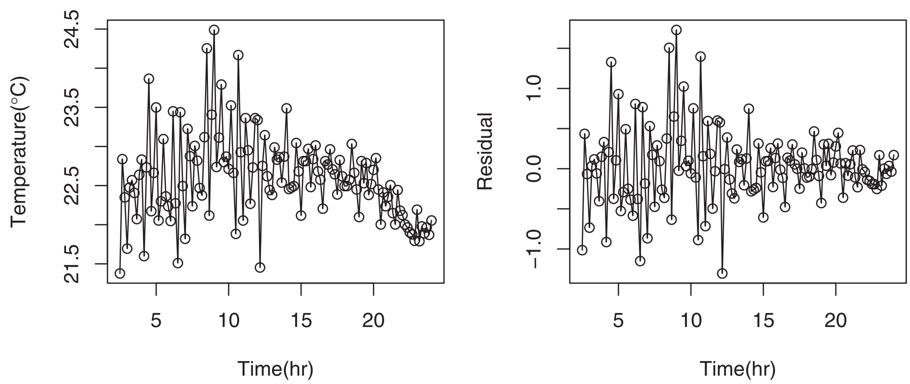

Figure 1. Raw temperature profile (left panel) and detrended temperature profile (right panel) at a randomly selected spot of the liver. The *x*-axis labels in both panels represent 24 hours.

an epilepsy patient may show a smooth mean trend and a sudden variation change in the signal may indicate the onset of a seizure. The stock price for a large corporation often shows a smooth mean trend and a sudden increase in variation could signal turmoil on the stock market due to share holders' anxiety about the company's health. While the change point detection procedure proposed here is a new method that has been applied to liver procurement, it can also be applied to many other areas.

The existing literature on change point analysis can be roughly divided into two categories. In the domain of parametric change point analysis, researchers assume that the underlying distribution belongs to some known family and sudden shift changes in the mean, variance, or both are considered. For example, when change of variance is the only concern, two representative approaches are the cumulative sum of squares approach in Inclán and Tiao (1994) and the Schwartz information criterion in Chen and Gupta (1997). When simultaneous shifts in mean and variance are considered, Horváth (1993) and Pan and Chen (2006) studied the theoretical properties of likelihood ratio test and modified information criterion respectively. One can refer to the monographs Carlstein, Müller, and Siegmund (1994), Qiu (2005), and Chen and Gupta (2012) for a comprehensive list of publications in parametric change point analysis. In the domain of nonparametric change point analysis, the assumption is that there is a sudden change in the probability distribution of the data. Various measures for such change have been developed in the literature to describe the differences between probability distributions. For example, Hariz, Wylie, and Zhang (2007) developed a semi-norm to measure the difference between empirical probability distributions and estimated the change point as the position where a weighted version of such difference is maximized. Matteson and James (2014) used hierarchical clustering to estimate the number of change points and their positions simultaneously for multivariate data. However, none of these existing methods in these two domains can address the problem posed in our experiment, where the variance change happened underneath a smoothly-changing mean trend. In particular, a smooth mean trend implies that the mean, and thus the distribution of the data, are constantly changing over time besides the sudden change in variance. Neither the parametric nor the nonparametric change point analysis methods can adequately describe a gradually changing mean trend. As demonstrated in our numerical experiments, erratic behavior and unreliable results occur when blindly applying these methods to such kind of data ignoring the underlying smooth mean trend.

Nonparametric smoothing and change point detection are often viewed as two conflicting issues in statistics since the former emphasizes on continuity and the latter represents discontinuity. The variance change point detection method proposed here naturally integrates these two domains in both numerical and theoretical senses. There has been other work combining nonparametric regression with change point detection. For example, both Loader (1996) and Grégoire and Hamrouni (2002) considered the problem of detecting jump points in smooth curves. However, they both focused on jumps in the mean curve, whereas our application clearly showed a jump in the variance. So, the method proposed in this article is uniquely suited to tackling the change point problem found in our liver procurement experiment.

Our variance change point detection method is formulated under the framework of penalized weighted least squares estimation. Particularly, the estimates of the mean function, the change point, and the variances are a local minimizer of a penalized weighted least squares score whose global minimizer may not exist. This objective functional consists of three parts: the weighted sum of squared errors represents the goodness of fit, the roughness penalty on the mean function estimate enforces smoothness on the mean, and the smoothing parameter balances the tradeoff. The optimization of the objective functional is carried out in an iterative fashion starting with a consistent initial mean estimate. When the mean function is given, the variance change point and the corresponding variances are estimated through a testing procedure generalizing the one in Chen and Gupta (1997). When the variance change point and the variances of two subsequences of data are given, the mean function is estimated by smoothing splines through the standard optimization of the penalized weighted least squares with known weights. The initial mean estimate is the minimizer of the penalized least squares under the working independence assumption.

For theoretical properties, we derive the asymptotic null distribution of our test statistic for the variance change point and we show that our change point estimate is consistent when the function space for the mean function is a periodic Soblev space. We note that these results have their own theoretical values too. Testing procedures under nonparametric null and alternative hypotheses are known to be very difficult problems since both the null and alternative spaces are of infinite dimensions. They become even harder in the penalized estimation scenario since the smoothing parameter in the penalty adds additional complexity to the derivation of asymptotic theory. For example, the rigorous theory for statistical inference with smoothing spline regression under the constant variance assumption was established by Shang and Cheng (2013) only a few years ago. And their work focused on the inference of the mean function. But our work studies hypothesis testing on the variance component. Our consistency result on the mean and variance component estimates is also new. Recognizing that the global minimizer of the penalized weighted least squares may not exist, we have proved the consistency of the estimates obtained from an iterative algorithm starting with a consistent initial mean estimate. This opens a new venue for studying the asymptotic theory of a nonparametric regression model when the random errors are not IID. So the theoretical developments here are novel and nontrivial.

In our simulations, we first demonstrate the pitfall of blindly applying the existing change point procedures without removing the smoothly changing mean trend when such a trend is present. Then, we show the excellent performance of our method in estimating the variance change point, the mean functions and the variances. The application of our method to the temperature profiles collected in the liver procurement experiment yield critical information about the viability status of the organ. In summary, our method has the following distinguishing features: (1) it is uniquely qualified to address the scientific hypothesis raised in our application experiment; (2) it is an innovative addition to the existing rich literature on change point analysis, (3) it naturally integrates smoothing and change point analysis in a way distinct from others, and (4) its theoretical development opens new fronts for the inference theory of nonparametric smoothing.

The rest of the article is organized as follows. In Section 2, we introduce the notation and model, the iterative algorithm, the mean estimation given the variances and change point, the test procedure for variance change point given the mean function, and the theoretical properties of the proposed method. In Section 3, we present all the simulations. We analyze the liver procurement data in Section 4. Discussion in Section 5 concludes the article. Proofs of the theorems are collected in the Appendix.

2. Method

2.1. Notation and Model

Suppose that y_i are independent observations generated from the following model:

$$y_i = f_0(i/n) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where f_0 is an unknown smooth function, $\epsilon_i \sim N(0, \sigma_i^2)$ with $\sigma_i = \sigma_0$ when $i \le \tau_0$ and $\sigma_i = \delta_0$ when $i > \tau_0$, and σ_0^2 and δ_0^2 are unknown variances. When $\sigma_0 = \delta_0$, there is no variance change point; when $\sigma_0 \neq \delta_0$, there is an unknown variance change point at τ_0. Assume that f_0 belongs to a reproducing kernel Hilbert space $\mathcal{H} = \{f | f : [0, 1] \to \mathbb{R}, J(f) < \infty\}$, where J is a seminorm on $\mathcal{H}$. For example, we consider $J(f) = \int_0^1 \{f^{(m)}(t)\}^2 dt$ in this article for some positive integer m. We propose to estimate $(f_0, \tau_0, \sigma_0^2, \delta_0^2)$ through the minimization of the penalized weighted least squares

$$\frac{1}{n}(\mathbf{y} - \mathbf{f})^T \Sigma_{n,\tau,\sigma,\delta}^{-1} (\mathbf{y} - \mathbf{f}) + \lambda J(f), \tag{2}$$

where f is a function in $\mathcal{H}$, $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{f} = (f(1/n), f(2/n), \ldots, f(1))^T$ are respectively the vectors of observed responses and fitted values, $\Sigma_{n,\tau,\sigma,\delta}$ is a diagonal matrix with the first τ diagonals equal to σ^2 and the rest equal to δ^2, $J(f)$ acts as a roughness penalty, and $\lambda > 0$ is the smoothing parameter balancing the tradeoff between the smoothness of the mean function estimate and the goodness of fit represented by the weighted sum of squared errors.

We note that the global minimizer of (2) does not exist since it approaches zero as σ^2 goes to infinity. Hence, we propose the estimates $(\widehat{f}, \widehat{\tau}, \widehat{\sigma}^2, \widehat{\delta}^2)$ as the local minimizer of (2) obtained through the following iterative algorithm. We shall show in Section 2.4 that the estimates are consistent with proper rates of convergence.

Algorithm.

1. Initialize $\widehat{f}^{(0)}$ with the mean function estimate assuming constant variance. That is, $\widehat{f}^{(0)}$ minimizes
$$\frac{1}{n}(\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda J(f). \tag{3}$$
Note that when $\sigma^2 = \delta^2$, the covariance matrix in (2) reduces to $\sigma^2 I$ and σ^2 can be absorbed into the smoothing parameter λ.
2. Each iteration consists of two steps. At the ιth iteration,
   (a) given the mean estimate $\widehat{f}^{(\iota-1)}$, we first use the testing procedure in Section 2.3 to find an estimate $\widehat{\tau}^{(\iota)}$ for τ_0. Then, we estimate the variance parameters respectively by the maximum likelihood variance estimates, $[\widehat{\sigma}^2]^{(\iota)}$ and $[\widehat{\delta}^2]^{(\iota)}$, of the subsequences of residuals, $\{y_i - \widehat{f}^{(\iota-1)}(i/n) : i = 1, \ldots, \widehat{\tau}^{(\iota)}\}$ and $\{y_i - \widehat{f}^{(\iota-1)}(i/n) : i = \widehat{\tau}^{(\iota)} + 1, \ldots, n\}$.
   (b) Now given the estimates $\widehat{\tau}^{(\iota)}$, $[\widehat{\sigma}^2]^{(\iota)}$ and $[\widehat{\delta}^2]^{(\iota)}$, we update the mean estimate by the minimizer of (2), where τ, σ^2, and δ^2 are replaced respectively by their current estimates.
3. Iterate until the algorithm converges.

The above algorithm is general in allowing multiple iterations. As will be demonstrated in Theorem 2.1, the updated estimators will satisfy desirable convergence properties, as long as the estimator from previous step is “good” enough. A simplified version is based on one iteration which will also satisfy Theorem 2.1 if the initial estimator $\widehat{f}^0$ converges sufficiently fast. Our numerical experiments, not reported here due to space concern, actually show that a single iteration can often yield satisfactory estimates. To achieve better and stable estimation, we use the full iteration approach here. The convergence criterion we use is the maximum absolute difference between the residuals of the current iteration versus the previous iteration. In our numerical experiments, our algorithm can usually converge in a few iterations. The selection of the smoothing parameter λ will be introduced in Section 2.2.

2.2. Mean Estimation Given τ, σ^2, and δ^2

When τ, σ^2, and δ^2 are given, the mean function f_0 is estimated as the minimizer of the penalized weighted least squares (2) in a reproducing kernel Hilbert space $\mathcal{H}$ of functions on the domain $\mathcal{T}$. A reproducing kernel Hilbert space (RKHS) is a Hilbert space $\mathcal{H}$ where the evaluation functional $[t] : \mathcal{H} \to \mathbb{R}$, $f \mapsto f(t)$ is continuous for every $t \in \mathcal{T}$. The Riesz representation theorem then indicates that for all $t \in \mathcal{T}$ there exists a unique function $R_t \in \mathcal{H}$ with the reproducing property $\langle R_t, f\rangle = [t](f) = f(t)$, where $\langle\cdot,\cdot\rangle$ is the inner product on $\mathcal{H}$. Now the reproducing kernel R of $\mathcal{H}$ is defined as a function $R : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ such that $R(s, t) = \langle R_s, R_t\rangle$. One can show that each RKHS is uniquely associated with a reproducing kernel and vice versa.

Note that the penalty functional J in (2) is a squared semi-norm on $\mathcal{H}$. The null space of J, namely $\mathcal{N}_J = \{f : J(f) = 0\}$, induces a direct sum decomposition $\mathcal{H} = \mathcal{N}_J \oplus \mathcal{H}_J$, where $\mathcal{H}_J$ is the complement of $\mathcal{N}_J$ in $\mathcal{H}$. This then yields a decomposition of the reproducing kernel $R = R_0 + R_J$, where R_0 and R_J are respectively the reproducing kernels on the subspaces $\mathcal{N}_J$ and $\mathcal{H}_J$. See, for example, Gu (2013, chap. 2) for more details on RKHSs.

We now introduce an example of cubic smoothing splines to illustrate these concepts. We shall use the cubic smoothing splines in all the numerical studies of the article.

Example 2.1 (Cubic Smoothing Splines). Without loss of generality assume $\mathcal{T} = [0, 1]$. A choice of $J(f)$ is $\int_0^1 (f'')^2 dt$, which yields the popular cubic splines. If the inner product in $\mathcal{N}_J$ is $(\int_0^1 f\,dt)(\int_0^1 g\,dt) + (\int_0^1 f'\,dt)(\int_0^1 g'\,dt)$, then $\mathcal{H}_J = \mathcal{H} \ominus \mathcal{N}_J = \{f : \int_0^1 f dt = \int_0^1 f' dt = 0, J(f) < \infty\}$ and the reproducing kernel $R_J(s, t) = k_2(s)k_2(t) - k_4(|s - t|)$, where $k_\nu(t) = B_\nu(t)/\nu!$ are scaled Bernoulli polynomials for $t \in [0, 1]$. The null space $\mathcal{N}_J$ has a basis $\{1, k_1(t)\}$ of 2 functions, where $k_1(t) = t - 0.5$ for $t \in [0, 1]$. See Gu (2013, sec. 2.3.3). □

The RKHS $\mathcal{H}$ is of infinite dimensions, so a direct optimization of (2) on $\mathcal{H}$ seems infeasible. However, since the weighted least squares part in (2) depends on f only through its evaluations at the observation points t_i, $i = 1, \ldots, n$, the Representer Theorem (Wahba 1990) guarantees that the exact minimizer of (2) actually resides in a finite dimensional subspace of $\mathcal{H}$, namely, $\mathcal{N}_J \oplus \text{span}\{R_J(t_1, \cdot), \ldots, R_J(t_n, \cdot)\}$. Let $\phi_l, l = 1, \ldots, m$ be the basis functions of N_J and $\xi_j = R_J(t_j, \cdot)$, $j = 1, \ldots, n$. Write $f = \boldsymbol{\phi}^T\mathbf{d} + \boldsymbol{\xi}^T\mathbf{c}$, where $\mathbf{c}$ and $\mathbf{d}$ are the corresponding coefficient vectors. Also note that $J(f)$ can be written as a quadratic form $J(f) = \mathbf{c}^T Q\mathbf{c}$, where Q is the $n \times n$ matrix with the (i, j)th entry equal to $R_J(t_i, t_j)$. So for a fixed λ, the objective function (2) is reduced to a quadratic function of the coefficient vectors $\mathbf{c}$ and $\mathbf{d}$. Its minimizer can be obtained analytically. To select the smoothing parameter λ, an outer loop for minimizing the generalized cross-validation (GCV) score is sufficient for the job; see Gu (2013, sec. 3.2).

2.3. Variance Change Point Detection Given f

Given $\widehat{f}$, we now introduce a testing procedure to find an estimate $\widehat{\tau}$ for the variance change point τ_0. Then we compute the maximum likelihood estimates for σ^2 and δ^2 respectively by $\widehat{\sigma}^2 = \widehat{\tau}^{-1}\sum_{i=1}^{\widehat{\tau}}\{y_i - \widehat{f}(i/n)\}^2$ and $\widehat{\delta}^2 = (n - \widehat{\tau})^{-1}\sum_{i=\widehat{\tau}+1}^{n}\{y_i - \widehat{f}(i/n)\}^2$. We propose a testing procedure that generalizes the one introduced by Chen and Gupta (1997) for the parametric case of normal data with a fixed mean.

We want to test the hypothesis

$$H_0 : \sigma_1^2 = \cdots = \sigma_n^2 \text{ versus}$$
$$H_1 : \sigma_1^2 = \cdots = \sigma_\tau^2 \neq \sigma_{\tau+1}^2 = \cdots = \sigma_n^2, \tag{4}$$

for a potential change point position τ. Let

$$\ell(\tau) = \tau \log\left[\frac{1}{\tau}\sum_{i=1}^{\tau}\{y_i - \widehat{f}(i/n)\}^2\right] + (n - \tau)\log\left[\frac{1}{n-\tau}\sum_{i=\tau+1}^{n}\{y_i - \widehat{f}(i/n)\}^2\right].$$

Note that $\ell(n) = -2L_0(\widehat{\sigma}^2) - n - n\log 2\pi$ and $\ell(\tau) = -2L_1(\widehat{\sigma}^2, \widehat{\delta}^2) - n - n\log 2\pi$, where L_0 and L_1 are respectively the log-likelihood functions under the null and alternative hypotheses of (4). So, we define the test statistic to be $\Delta_n^2 = \max_{1<\tau<n}\{\ell(n) - \ell(\tau)\}$.

To gain further insight for the test statistic Δ_n^2, we recap the motivation illustrated by Chen and Gupta (1997) by referring to the Schwartz information criterion (SIC) from Schwarz (1978). As a criterion for model selection, the SIC is defined as $-2\log L(\widehat{\boldsymbol{\theta}}) + p\log n$, where $L(\widehat{\boldsymbol{\theta}})$ is the likelihood function for the model, $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of the parameter θ, and p is the dimension of $\boldsymbol{\theta}$. In our case, given f and τ we have two models corresponding to the null and alternative hypotheses with their SICs respectively defined by $\text{SIC}(n) = -2L_0(\widehat{\sigma}^2) + \log n$ and $\text{SIC}(\tau) = -2L_1(\widehat{\sigma}^2, \widehat{\delta}^2) + 2\log n$. By the principle of minimum information criterion, we do not reject H_0 if $\text{SIC}(n) \le \min_\tau \text{SIC}(\tau)$, or equivalently $\ell(n) \le \min_{1<\tau<n}\ell(\tau)$, and reject H_0 if $\text{SIC}(n) > \text{SIC}(\tau)$ for some τ, or equivalently $\ell(n) > \ell(\tau)$ for some τ. In the case of rejection(s), we estimate the position of change point by $\widehat{\tau} = \arg\min_{1<\tau<n}\ell(\tau)$. So our test statistic can also be written as $\Delta_n^2 = \log n - \min_{1<\tau<n}\{\text{SIC}(\tau) - \text{SIC}(n)\}$. We shall present the asymptotic distribution of Δ_n^2 under the null hypothesis in Section 2.4.

2.4. Theoretical Properties

In this section, we present the asymptotic theories for the proposed method. Due to concern of the article's length, the conditions and technical proofs are collected in the Online Supplmentary Material. For simplicity, we only consider the special case when $\mathcal{H}$ is the mth order Sobolev space of periodic functions on [0, 1] with period 1, namely,

$$\mathcal{H} = S^m \equiv \left\{f : f(t) = \sum_{\nu=1}^{\infty} f_\nu\varphi_\nu(t) \text{ with } t \in [0, 1] \text{ and } \sum_{\nu=1}^{\infty} f_\nu^2\gamma_\nu < \infty\right\},$$

where for $k = 1, 2, \ldots$, $\varphi_{2k-1}(t) = \sqrt{2}\cos(2\pi kt)$, $\varphi_{2k}(t) = \sqrt{2}\sin(2\pi kt)$, and $\gamma_{2k-1} = \gamma_{2k} = (2\pi k)^{2m}$. Note that

$J(f)=\int_0^1\{f^{(m)}(t)\}^2dt=\sum_{\nu=1}^{\infty} f_\nu^2\gamma_\nu$ for $f\in S^m$ and $R_J(s,t)=(2\pi m)^{-2m}\sum_{\nu=1}^{\infty}2\cos\{2\pi\nu(s-t)\}/(2\pi m\nu)^{2m}$.

Let $h=\lambda^{1/(2m)}$, $r_n=\sqrt{\log n/(nh)}+h^{m-1/2}$, and $\widetilde{r}_n=r_n^2+(nh)^{-3/4}+(\log n)^5(\log\log n)^2/n+n^{-1/2}$. We shall first show the consistency of the estimates $(\widehat{f},\widehat{\tau},\widehat{\sigma}^2,\widehat{\delta}^2)$.

Theorem 2.1 (Consistency of Parameter Estimates). Under Conditions 1–3 in the online supplementary material, the estimates $(\widehat{f},\widehat{\tau},\widehat{\sigma}^2,\widehat{\delta}^2)$ from the algorithm in Section 2.1 are consistent with the following rates of convergence:

$$\|\widehat{f}-f_0\|_n^2=O_P\left(\lambda+(nh)^{-1}+h^{-1}\widetilde{r}_n^2\right),$$
$$|\widehat{\tau}-\tau_0|=O_P((\log n)^4(\log\log n)^2),$$
$$|\widehat{\sigma}^2-\sigma_0^2|=O_P(\widetilde{r}_n),\ |\widehat{\delta}^2-\delta_0^2|=O_P(\widetilde{r}_n),$$

where $\|f\|_n=\sqrt{\sum_{i=1}^n f(i/n)^2/n}$ is the empirical norm of a function f.

Note that when $m\ge 1$ and $\lambda\asymp n^{-2m/(2m+1)}$, it can be verified that $\widetilde{r}_n=O(n^{-1/2})$. Then, this implies that $\widehat{\sigma}^2$ and $\widehat{\delta}^2$ are $\sqrt{n}$-consistent, and that $\|\widehat{f}-f_0\|_n=O_P(n^{-m/(2m+1)})$ or $\widehat{f}$ achieves the optimal convergence rate of a spline function estimate.

We then derive the asymptotic sampling distribution of the test statistic Δ_n^2 under the null hypothesis H_0 in (4).

Theorem 2.2 (Asymptotic Null Distribution of Test Statistic). Suppose that as $n\to\infty$, $h\to 0$ and $r_n^2\log n\to 0$. Under H_0 in (4) and Conditions 1, 2, and 3' in the online supplementary material, for any $t\in\mathbb{R}$,

$$P(a_n(\log n)^{1/2}\Delta_n-b_n\log n\le t)\to\exp(-2\exp(-t)),$$

where $a_n=(2\log\log n)^{1/2}/\log n$, and $b_n=\{2\log\log n+\frac{1}{2}\log\log\log n-\log\Gamma(1/2)\}/\log n$.

The limit distribution turns out to be an extreme value distribution. Based on this result, we propose the following testing rule at the significance level $1-\alpha$:

$$\text{Reject } H_0 \Leftrightarrow a_n(\log n)^{1/2}\Delta_n-b_n\log n>-\log\{-\log(1-\alpha)/2\}.$$

3. Simulations

3.1. Estimation Performance

We compared the change point estimation performance of the proposed variance detection method with three change point detection methods. The first two are existing methods, one from the parametric domain and the other from the nonparametric domain. The parametric method is the SIC approach in Chen and Gupta (1997) hereafter denoted by the CG method. We used the implementation in the `changepoint` package of R. The nonparametric method is the hierarchical clustering approach by Matteson and James (2014) hereafter denoted by the MJ method. We used the authors' implementation in their R package `ecp`. Besides these approaches, we also designed a moving-window approach following the idea by Niu and Zhang (2012). They considered a local method for detecting mean change points. In each moving window of a fixed size, they computed the difference between the means of data on the left and right hand sides of the center of the window. When the difference exceeds a carefully chosen threshold, the center of the window is claimed as a mean change point. We adapted this approach to the variance change point detection here. For each moving window, we computed the log ratio of the sample variances for data on both sides of the center. Since our simulated data contained only one change point, we simply took the center of the window with the largest log ratio as the variance change point identified by the moving window approach. We considered two window sizes, denoted by MW5 and MW10, such that there are respectively 5 and 10 data points on each side of the window. Furthermore, we examined the performance of the proposed method in estimating the mean curve and the variances.

We considered two mean functions $f_{01}(t)=20+12t(1-t)$ and $f_{02}(t)=\sin(t)+t^5-8t^3+10t+6$. The first function f_{01} had a trend similar to the mean temperature profile in the liver procurement study and the second function f_{02} represented a more complex smooth trend. Two sample sizes $n=130$ and 500 were used. The true variance change point was set at $\tau_0=65$ when $n=130$, and $\tau_0=250$ when $n=500$. The true variances were $\sigma_0^2=0.219$ and $\delta_0^2=0.057$ when f_{01} was the true mean function, and $\sigma_0^2=9$ and $\delta_0^2=2$ when f_{02} was the true mean function. We simulated 1000 data replicates for each combination of mean function and sample size.

For each data replicate, we applied the five variance change point detection methods (CG, MJ, MW5, MW10, and New) to obtain the change point estimate. These estimates were divided by n to rescale them to the range of (0, 1) for easier comparison. For the proposed method, we also obtained the mean function estimate and the two variance estimates. To evaluate their performances, we computed the mean squared error MSE $=n^{-1}\sum_{i=1}^n\{\widehat{f}(i/n)-f_0(i/n)\}^2$ and the log ratios $\log(\widehat{\sigma}^2/\sigma_0^2)$ and $\log(\widehat{\delta}^2/\delta_0^2)$.

Figure 2 displayed the boxplots of change point estimates from the five methods. We can clearly see that both the CG and the MJ methods suffered when blindly applied to the data without removing the mean trend. The moving window approaches with window sizes 5 and 10 both suffered too. On the other hand, the proposed method did a decent job in estimating the location of the change point. And the estimation accuracy clearly improved as the sample size n increased from 130 to 500. Note that since the moving window approaches essentially treat the mean within each window as a constant, they can be considered as variance change point detection methods where the underlying mean trend is roughly approximated by moving constants. Therefore, their poor performance fully demonstrates the importance of an accurate mean estimation in the variance change point detection scenario considered in this article.

Figure 3 assesses the performance of mean estimation. The top panels plotted the mean estimates that attained the 25th, 50th, and 75th percentiles of the MSEs for sample sizes $n=130$ and 500. The mean function estimates all matched well with the true functions. The 75th percentile estimate for the true function f_{01} with $n=130$ was slightly off in the area around the change point, which was reasonable considering the fluctuations in that area. Also, the estimation accuracies improved as the sample size increased.

Figure 4 uses the log ratios of variance estimates versus true variances to assess the estimation performance for both variances. We can see that both variances were accurately

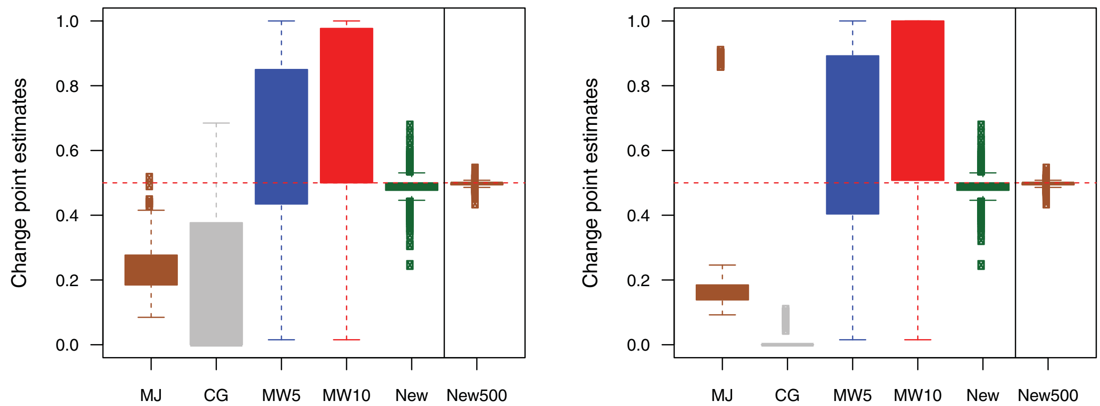

Figure 2. Boxplots of change point estimates. Left panel: simulations with the true mean function $= f_{01}$; Right panel: simulations with the true mean function $= f_{02}$. The five plots on the left in each panel were the change point estimates with $n = 130$ respectively for the methods by Matteson and James (2014) (MJ), Chen and Gupta (1997) (CG), moving-window approaches with half-window sizes 5 (MW5) and 10 (MW10), and the newly proposed method (New). The rightmost plot in each panel was the proposed method with $n = 500$ (New500). The red dashed line is the true change point $\tau_0/n = 0.5$.

estimated with the accuracies also improved as the sample size increased.

3.2. Power Analysis Study

Next, we present a power analysis study on the change point testing procedure. We considered the data simulated from the mean function $f_{01}(t)$. The variance $\delta_0^2 = 0.06$ and $\sigma_0^2 = \theta\delta_0^2$, where $\theta \geq 1$ was the ratio of σ_0^2 over δ_0^2.

Note that when $\theta = 1$, $\delta_0^2 = \delta_0^2$ and there was no change point. So, we used this setting to investigate the size of the test. We considered two levels $\alpha = 0.05, 0.1$ and four sample sizes $n = 130, 500, 2000, 10{,}000$. For each combination of α and n, we simulated 10,000 data replicates. Table 1 summarizes the results about the size of the test. We can see the empirical sizes of the test were smaller than the levels of the test. It indicates that the test is a bit more conservative than expected in claiming a change point when there is no change point. One

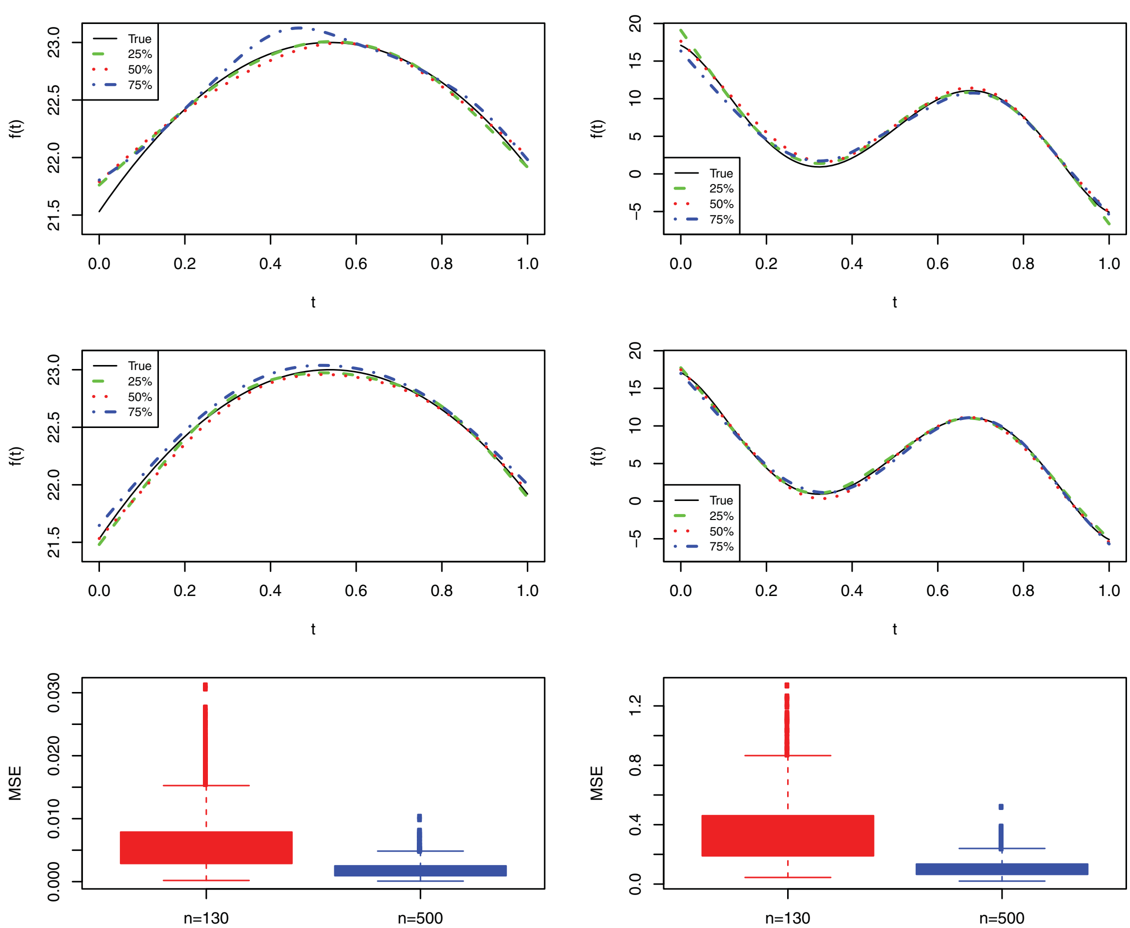

Figure 3. Plots for assessing mean estimation performance. Left panels: simulations with the true mean function $= f_{01}$; Right panels: simulations with the true mean function $= f_{02}$. Top: true mean function (solid black) versus the mean estimates with $n = 130$ whose MSE were the 25th (dashed green), 50th (dotted red), and 75th (dot-dashed blue) percentiles of the 1000 MSEs obtained in each setting. Middle: same as top but with $n = 500$. Bottom: boxplots of the 1000 MSEs in each setting.

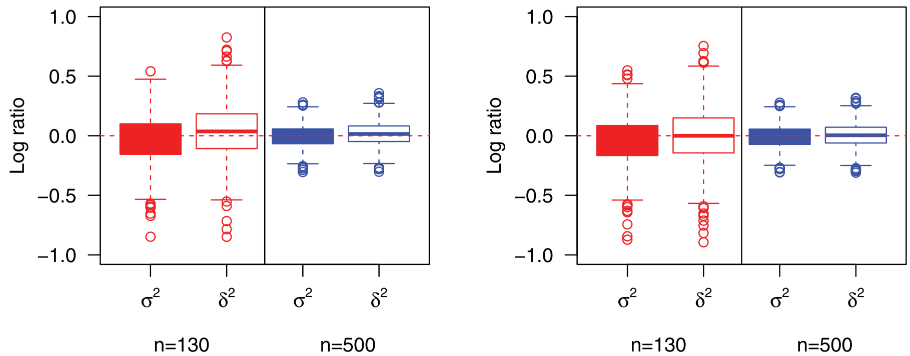

Figure 4. Boxplots of the log ratios of variance estimates versus true variances. Left panels: simulations with the true mean function = f_{01}; right panels: simulations with the true mean function = f_{02}. Red: $n = 130$; Blue: $n = 500$. Filled boxes: σ^2; Unfilled boxes: δ^2.

Table 1. Test size simulation result: proportions of rejections when H_0 is true. The true mean function is $f_{01}(t)$.

Test Level	Sample Size n			
α	130	500	2000	10,000
0.05	0.0097	0.0114	0.0141	0.015
0.1	0.0346	0.04	0.0461	0.0504

possible reason for this phenomenon is that the asymptotic null distribution, an extreme value distribution, is a heavy-tailed distribution and may require a larger sample size to achieve the desired size of the hypothesis test.

For $\theta > 1$, the simulations were used to assess the power of the test procedure. Particularly, we considered the variance ratio θ taking values from a set of grid points on $(1, 5]$. We considered two sample sizes $n = 130$ and 500. When $n = 130$, the grid points for θ were $\{1.25, 1.50, \ldots, 5.00\}$. When $n = 500$, the grid points were $\{1.2, 1.4, 1.6, 1.8, 1.9, 2.0, 2.1, 2.2, 2.5, 5.0\}$. Then for each combination of θ and n, we simulated 10,000 data replicates. The resulting power function plots are in Figure 5. For $n = 130$, the power became greater than 0.8 when the variance ratio was 3. For $n = 500$, the power got higher than 0.9 even before the variance ratio hit 2.

4. Application: Temperature Monitoring in Liver Procurement

Viability assessment is a critical step in organ transplant procedures. The current assessment procedure purely relies on visual inspection of potentially transplantable organs by physicians and organ procurement personnel and biopsy of the organ. While the former suffers from subjective judgment (influenced by the experience of the observer, and gross evaluations of organ color and shape), the latter is an intrusive approach that destroys the part of organ where the biopsy sample is collected. Aiming to find a new, noninvasive way of assessing the viability of organs, a biomedical engineering team at Virginia Tech designed a temperature monitoring system such that the surface temperature of a perfused organ can be densely and continuously monitored, using high-resolution infrared thermography. In the experiment considered in this article, a lobe of porcine liver, as shown in Figure 6, was perfused with a physiologic perfusion fluid (modified Krebs' solution). Whole organ surface temperature was intensively monitored for a continuous period of 24 hours. The liver lobe was optically (not physically) divided into a dense grid of 36,795 spots with each spot producing a 24-hour temperature profile. Temperature measurements were collected every 10 minutes, yielding a total of 145 points in each profile. The first 2.5 hours of data were discarded since it took about one to two hours for the perfusion fluid to completely infuse and stabilize the liver. There were $n = 130$ points left in each profile, after correction and elimination of the data collected during initial infusion and stabilization.

We applied the proposed variance change point detection method to the 36,795 temperature profiles in the data. Since a large number of hypothesis tests were involved here, we considered the Benjamini–Hochberg–Yekutieli (BHY) procedure (Benjamini and Yekutieli 2001) to address the multiple

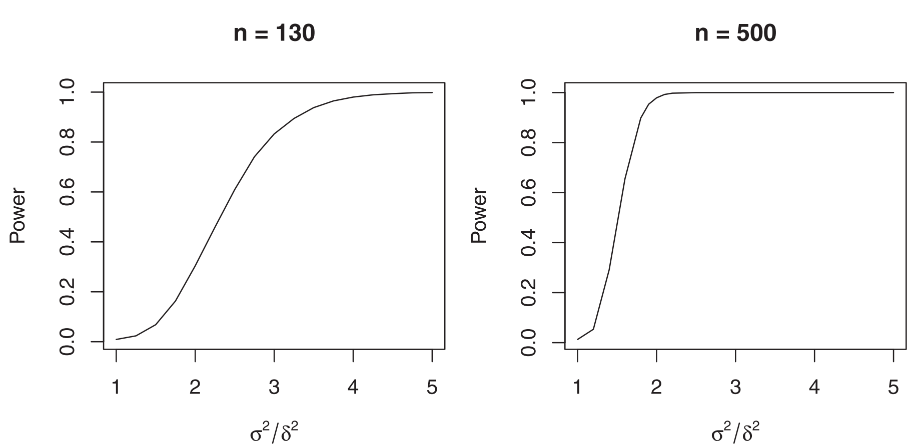

Figure 5. Plots of power against ratio of the variances.

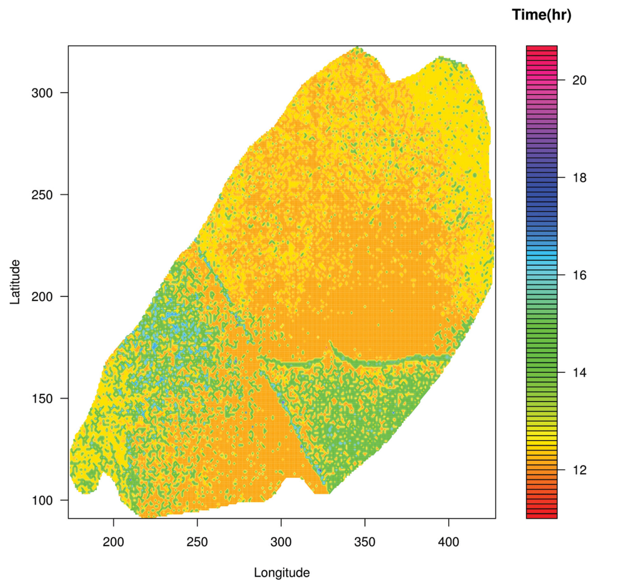

Figure 6. The heat map of estimated variance change points of temperatures on the lobe of liver in the procurement experiment.

comparison issue with the control of false discovery rate. This procedure is an extension of the well-known Benjamini–Hochberg procedure (Benjamini and Hochberg 1995) to the case of dependent tests. Due to the positive correlation between our temperature profiles, we used the positive dependency version of the procedure with the false discovery rate controlled at level 0.05. The largest p-value among all the 36,795 tests of variance change points was 0.019. Hence, all the change points were considered to be legitimate following the principle of the BHY procedure.

The heat map of all the estimated change points is plotted in Figure 6. Note that an earlier change point in variance meant an earlier drop in the viability of the cells around the spot. We can see that the top half and the middle bottom parts of the liver mostly failed around 12 hours while the bottom left and right portions of the liver lasted beyond 14 hours. There were also several clearly visible straight-line type of boundaries between the early and late failure areas. These may be parts where the porcine liver lobe was deformed during dissection and perfusion.

Figure 7 plotted the mean estimates and variance change point estimates at three randomly selected spots, imposed respectively on the raw and de-trended temperature profiles. All the mean estimates matched well with the trends shown in the data. As we can see, the mean temperature increased at different rates at the three spots in the first 12 hours or so and shared a common trend of a more rapid drop in the second half of the 24-hour period. The variance change points at the three points were all between 12 and 15 hours.

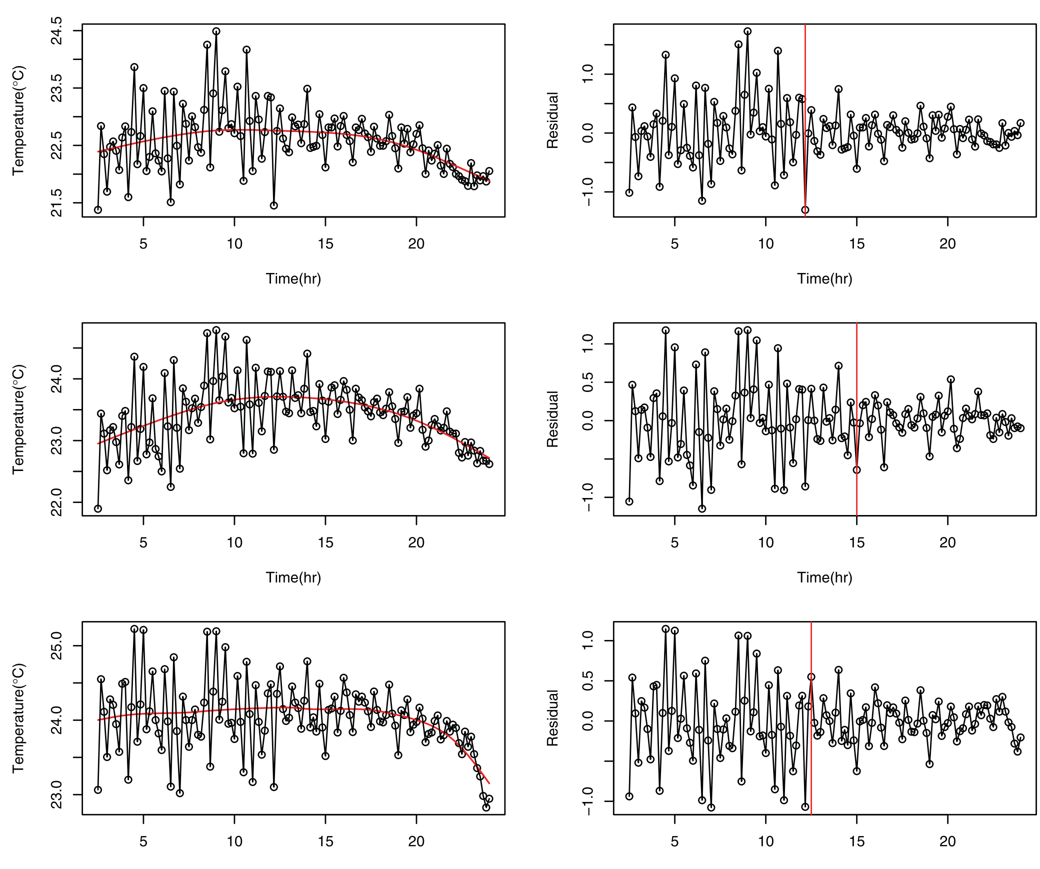

Figure 7. Mean and variance change point estimates imposed respectively on the raw and de-trended temperature profiles at three randomly selected spots.

5. Conclusion

In this article, we have presented a new variance change point detection method when the underlying mean trend changes smoothly. Motivated from a liver procurement experiment, the proposed method naturally integrates the seemingly conflicting goals of estimating a smooth mean and detecting a jump point in variance under the framework of penalized weighted least squares. As demonstrated in the simulations, this is not something that can be resolved by the existing change point detection methods. Furthermore, the testing procedure under our nonparametric smoothing setting is shown to have theoretical properties similar to that under a parameter model. The consistency result is also innovative in the perspective of nonparametric regression with non-IID errors. The comparison simulations show that having an accurately estimated mean function is indeed critical to the correct identification of the variance change point. The application of our method to the liver procurement experimental data provided critical information about the viability status of the liver lobe at different locations. A direction that merit further investigation is the development of an online version of our procedure. This can be developed in combination with a proper characterization of in-control data.

It is also of interest to understand why the global minimizer of (2) does not exist but the proposed iterative procedure with a consistent initial mean estimate can converge to consistent estimates of the parameters. Intuitively, if one allows all the model parameters to vary freely, then an exceptionally large variance can tolerate any kind of mean estimate and accurate estimation becomes hopeless. On the other hand, if one can start from a consistent mean estimate, then the candidate estimates within the neighborhood can all have well-controlled variances and consistent estimation thus becomes feasible. Alternatively, one may pursue a global minimizer of (2) by introducing some penalty on the variance parameters to punish larger variances. But the optimal choice of the penalty as well as the ensuing complication of the theoretical development are beyond the scope of this article.

Acknowledgements

The authors are grateful to the AE and the reviewers for their insightful comments that have significantly improved the article.

Funding

Du's research was supported by U.S. National Science Foundation under grant DMS-1620945. Shang's research was supported by U.S. National Science Foundation under grant DMS-1712919.

References

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*, Series B, 57, 289–300. [780]

Benjamini, Y., and Yekutieli, D. (2001), "The Control of the False Discovery rate in Multiple Testing under Dependency," *The Annals of Statistics*, 29, 1165–1188. [779]

Carlstein, E. G., Müller, H.-G., and Siegmund, D. (1994), *Change-Point Problems*, Hayward, CA: Institute of Mathematical Statistics. [774]

Chen, J., and Gupta, A. K. (1997), "Testing and Locating Variance Change-points with Application to Stock Prices," *Journal of the American Statistical Association*, 92, 739–747. [774,776,777]

——— (2012), *Parametric Statistical Change Point Analysis* (2nd Ed.), Basel; Cambridge, MA: Birkhäuser Verlag. [774]

Grégoire, G., and Hamrouni, Z. (2002), "Change Point Estimation by Local Linear Smoothing," *Journal of Multivariate Analysis*, 83, 56–83. [774]

Gu, C. (2013), *Smoothing Spline ANOVA Models* (2nd Ed.), New York: Springer-Verlag. [776]

Hariz, S. B., Wylie, J. J., and Zhang, Q. (2007), "Optimal Rate of Convergence for Nonparametric Change-Point Estimators for Nonstationary Sequences," *The Annals of Statistics*, 35, 1802–1826. [774]

Horváth, L. (1993), "The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations," *The Annals of Statistics*, 21, 671–680. [774]

Inclán, C., and Tiao, G. C. (1994), "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance," *Journal of the American Statistical Association*, 89, 913–923. [774]

Loader, C. R. (1996), "Change Point Estimation using Nonparametric Regression," *The Annals of Statistics*, 24, 1667–1678. [774]

Matteson, D. S., and James, N. A. (2014), "A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data," *Journal of the American Statistical Association*, 109, 334–345. [774,777]

Niu, Y. S., and Zhang, H. (2012), "The Screening and Ranking Algorithm to Detect DNA Copy Number Variations," *The Annals of Applied Statistics*, 6, 1306–1326. [777]

Pan, J., and Chen, J. (2006), "Application of Modified Information Criterion to Multiple Change Point Problems," *Journal of Multivariate Analysis*, 97, 2221–2241. [774]

Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, New York: Wiley. [774]

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [776]

Shang, Z., and Cheng, G. (2013), "Local and Global Asymptotic Inference in Smoothing Spline Models," *The Annals of Statistics*, 41, 2608–2638. [775]

Wahba, G. (1990), "*Spline Models for Observational Data*," vol. 59 of CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia, PA: SIAM. [776]