

Variance Change Point Detection under A Smoothly-changing Mean Trend with Application to Liver Procurement

Zhenguo Gao

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Pang Du, Chairman

Yili Hong

Inyong Kim

Xiaowei Wu

February 14, 2018

Blacksburg, Virginia

Keywords: Variance change point, Smoothly-changing mean trend, Hypothesis testing in
nonparametric smoothing, Change point consistency, Asymptotic null distribution

Copyright 2018, Zhenguo Gao

Variance Change Point Detection under A Smoothly-changing Mean Trend with Application to Liver Procurement

Zhenguo Gao

(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 that the variance of data may make a significant jump while the mean of data changes in a smooth fashion. It is motivated by a liver procurement experiment with organ surface temperature monitoring. Blindly applying the existing change point analysis methods to the example can yield erratic change point estimates since the smoothly-changing mean violates the sudden-change assumption. In my dissertation, we propose a penalized weighted least squares approach with an iterative estimation procedure that naturally integrates variance change point detection and smooth mean function estimation. 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 the non-invasive organ viability assessment by surface temperature monitoring.

The method above can only yield the variance change point of temperature at a single point on the surface of the organ at a time. In practice, an organ is often transplanted as a whole or in part. Therefore, it is generally of more interest to study the variance change

point for a chunk of organ. With this motivation, we extend our method to study variance change point for a chunk of the organ surface. Now the variances become functions on a 2D space of locations (longitude and latitude) and the mean is a function on a 3D space of location and time. We model the variance functions by thin-plate splines and the mean function by the tensor product of thin-plate splines and cubic splines. However, the additional dimensions in these functions incur serious computational problems since the sample size, as a product of the number of locations and the number of sampling time points, becomes too large to run the standard multi-dimensional spline models. To overcome the computational hurdle, we introduce a multi-stages subsampling strategy into our modified iterative algorithm. The strategy involves several down-sampling or subsampling steps educated by preliminary statistical measures. We carry out extensive simulations to show that the new method can efficiently cut down the computational cost and make a practically unsolvable problem solvable with reasonable time and satisfactory parameter estimates. Application of the new method to the liver surface temperature monitoring data shows its effectiveness in providing accurate status change information for a portion of or the whole organ.

Variance Change Point Detection under A Smoothly-changing Mean Trend with Application to Liver Procurement

Zhenguo Gao

(General Audience Abstract)

The viability evaluation is the key issue in the organ transplant operation. The donated organ must be viable at the time of being transplanted to the recipient. Nowadays, viability evaluation can be assessed by analyzing the temperature data monitored on the organ surface. In my dissertation, I have developed two new statistical methods to evaluate the viability status of a prepared organ by studying the organ surface temperature. The first method I have developed can be used to detect the change of viability status at a spot on the organ surface. The second method I have developed can be used to detect the change of viability condition for the selected organ chunks. In practice, combining these two methods together can provide accurate viability status change information for a portion of or the whole organ effectively.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Prof. Pang Du for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank Prof. Zuofeng Shang from Indiana University - Purdue University at Indianapolis for his help with the theoretical development in the dissertation. Thank Drs. Ran Jin and John Robertson for providing the liver temperature monitoring data and for explaining the experiments. I would also like to thank the rest of my thesis committee: Prof. Yili Hong, Prof. Inyong Kim, and Prof. Xiaowei Wu, for their encouragement, insightful comments, and hard questions.

Furthermore, I want to thank all the faculty and staff members of the Department of Statistics for their help during my five years at Virginia Tech. I want to thank all my fellow students who have offered assistance, encouragement and friendship during the course of my study.

Last but not the least, I would like to thank my family, especially my parents, for their consistent support.

My research thesis is supported by the NSF grant DMS-1620945.

Contents

1	Introduction	1
1.1	Motivating Application	1
1.2	Change Point Analysis Literature Review	3
1.2.1	Parametric Change Point Analysis	3
1.2.2	Nonparametric Change Point Analysis	5
1.3	Smoothing Spline Models	7
1.3.1	Penalized (Weighted) Least Squares Estimation	8
1.3.2	Thin-Plate Spline	10
1.3.3	Tensor product of Cubic splines and Thin-Plate Splines	12
1.4	New Challenges	14
2	Variance Change Point Detection under a Smoothly-changing Mean Trend: The Univariate Case	16

2.1	Introduction	16
2.2	Method	20
2.2.1	Notation and Model	20
2.2.2	Mean Estimation Given τ , σ^2 , and δ^2	22
2.2.3	Variance Change Point Detection Given f	23
2.2.4	Theoretical Properties	24
2.3	Simulations	27
2.3.1	Estimation Performance	27
2.3.2	Power Analysis Study	31
2.4	Application: Temperature Monitoring in Liver Procurement	33
2.5	Conclusion	37
3	Variance Change Point Detection under a Smoothly-changing Mean Trend:	
	The Multivariate Case	56
3.1	Introduction	56
3.2	Method	62

3.2.1	Notation and Model	62
3.2.2	Smoothing Splines Estimation	62
3.2.3	Three-stage Subsampling	66
3.2.4	Iterative Estimation on $T \times Z$	67
3.2.5	Final Estimates	69
3.3	Simulations	71
3.3.1	Computational time comparison	71
3.3.2	Choice of location subsample size	72
3.3.3	Estimation performance	79
3.4	Application: Temperature Monitoring in Liver Procurement	80
3.5	Conclusion	89
	References	96

List of Figures

1.1	Sample of raw temperature on organ surface	2
2.1	Boxplots of change point estimates	29
2.2	Plots for assessing mean estimation performance	30
2.3	Boxplots of the log ratios of variance estimates versus true variances	31
2.4	Plots of power against ratio of the variances	33
2.5	The heat map of estimated variance change points	35
2.6	Example of mean and variance change point estimates	36
3.1	Sample of temperature profile and the heat map of estimated variance change points	57
3.2	Mean function estimate plots with $f_0 = f_{01}$ and $m = 30$	73
3.3	Variance function estimate plots with $f_0 = f_{01}$ and $m = 30$	74

3.4	Mean function estimate plots with $f_0 = f_{02}$ and $m = 30$	75
3.5	Variance function estimate plots with $f_0 = f_{02}$ and $m = 30$	76
3.6	Mean function estimate plots with $f_0 = f_{01}$ and $m = 10$	77
3.7	Variance function estimate plots with $f_0 = f_{01}$ and $m = 10$	78
3.8	Boxplots of change point estimates	79
3.9	Boxplots of MSE	80
3.10	The heat map of estimated variance change points with 5 selected areas	83
3.11	Variance change point estimates for area #1	84
3.12	Variance change point estimates for area #2	85
3.13	Variance change point estimates for area #3	86
3.14	Variance change point estimates for area #4	87
3.15	Variance change point estimates for area #5	88

List of Tables

2.1	Iterative Algorithm I	21
2.2	Test size simulation result	32
3.1	Iterative Algorithm II	63
3.2	Table of computational time comparison	72

Chapter 1 Introduction

1.1 Motivating Application

My dissertation is motivated from an experiment about the procurement of transplant livers. Quality/viability evaluation is a key issue in the procurement of transplant organs. Currently, such evaluations are mostly performed through visual inspection by surgeons or biopsy image assessment by pathologists. Both approaches are subjective judgments. 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. And the viability status of the biopsy sample may not represent that of the whole organ.

In the experiment considered in my dissertation, surface temperature of a severed porcine liver was constantly monitored upon the infusion of the perfusion liquid to the organ. The measurements consisted of surface temperatures measured every 10 minutes on a dense grid covering the whole organ for a span of 24 hours. The left panels in Figure 3.1 is the temperature profile for a randomly selected spot on the organ surface. The temperature of the perfusion liquid is often slightly different from the body temperature. So the temperature of the organ changes in a slow fashion and displayed an overall smooth mean trend. The high oscillations in the first half reflected the resistance of the organ to the abrupt temperature change in the environment. Around the 10th hour, the organ started to lose its viability and this change is reflected in a sudden drop in the variance of the temperature, as shown in the plot of residuals versus time in the right panels of Figure 3.1. Our goal is to design a testing procedure for identifying the variance change point of the residuals after removing the smoothly changing mean trend.

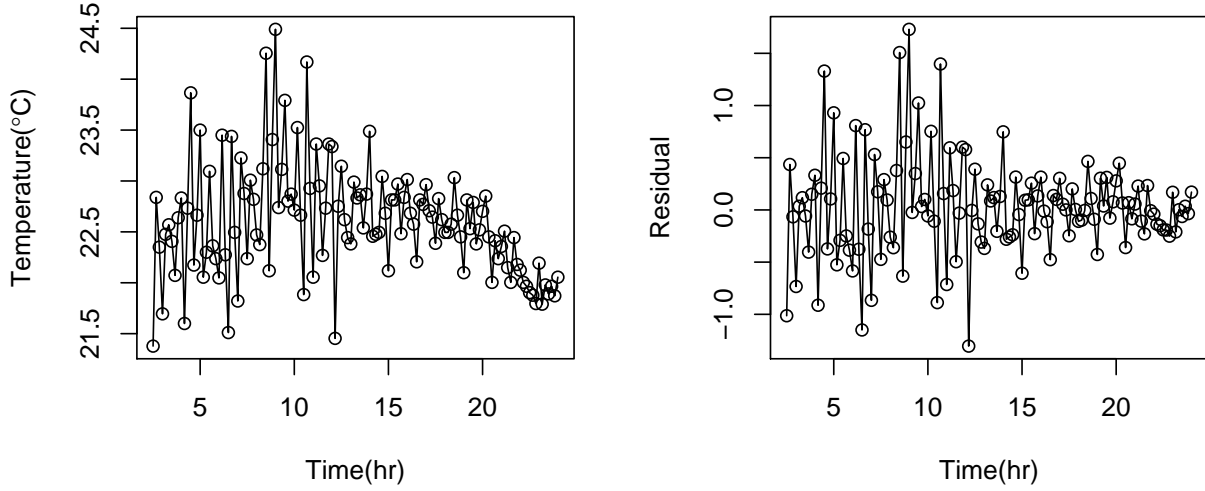


Figure 1.1: Sample of raw temperature profiles (left panels) and detrended temperature profiles (right panels) at three spots of the liver. The x-axis labels in both panels represent 24 hours.

Note that this phenomenon of having a variance change point underlying a smooth mean trend actually occurs in many other settings besides the liver procurement experiment considered here. For example, seismic activity monitoring often sees a smooth mean trend with small variation and a sudden change in variation could be the early sign of an earthquake; the EEG signal for an epilepsy patient generally shows a smooth mean trend and a sudden variation change in the signal might mean the onset of a seizure; the stock price for a big company often shows a smooth mean trend and a sudden increase in variation could mean a turmoil on the stock market or stock holders' rising panic about the company's health. So the change point detection procedure proposed here is a new method that arises naturally from our motivating example on liver procurement and can be also applied to many other areas.

The methods proposed in this dissertation is an integration of changing point analysis and smoothing spline regression. Therefore we will review some relevant literature on change point analysis and some background of smoothing spline regression.

1.2 Change Point Analysis Literature Review

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. In contrast, nonparametric change point analysis methods generally don't assume any specific distribution family and aim to detect a sudden change in the data distribution as a whole entity.

1.2.1 Parametric Change Point Analysis

Detect Mean Change While Variance Remains Constant. There has been considerable efforts devoted to the mean change point analysis when data are assumed to be normally distributed. One of the earliest developments was Page (1957) who proposed a cumulative sum (CUSUM) method. Chernoff and Zacks (1964) presented a Bayesian approach to locate a change in the mean of a normal distribution, as well as estimating the current mean. Hinkley (1970) derived a asymptotic distribution for the maximum likelihood statistics and performed a likelihood ratio test to locate the change points. Gardner (1969), Hawkins (1977), and Worsley (1979) all focused on the study of change in the mean for data follows a Gaussian distribution. Sen and Srivastava (1975a,b) derived the exact and asymptotic distributions from the Bayesian point of view to inspect the changes in mean. After years of dormancy, as change point detection been more widely used in various fields in the past decades, research on the mean change point in a normal distribution has reemerged. Albert and Hunsberger (2004) developed a penalized likelihood ratio test statistics to calculate multiple change points in the mean structure. Zhang and Siegmund (2007) proposed a modified BIC for locating multiple change points. Chen et al. (2011), used the binary segmentation procedure and maximum residual down methods to detect multiple change points.

Detect Variance Change While Mean Remains Constant. When change of variance is the only concern, two representative approaches are the cumulative sum of squares approach in Inclán (1993) and Inclán and Tiao (1994), and the Schwartz information criterion in Chen and Gupta (1997). Inclán (1993) developed a Bayesian procedure for detecting multiple change points using posterior odds, while Inclán and Tiao (1994) proposed an iterative cumulative sum of squares algorithm to locate multiple changes. Chen and Gupta (1997), who proposed a parametric method of testing and locating multiple variance change points in a sequence of independent Gaussian random variables (assuming known and common mean). They used the binary procedure combined with the Schwarz Information Criterion (SIC) to search all of the possible variance change points existed in the sequence. Hinkley (1971) studied the use of cumulative sum tests for detecting a change point and then estimating its location, which are adaptable to the presence of multiple change points through a sequential algorithm. Farley et al. (1975) studied the nonstationarity in the parameters of the one-factor market model by using the technique of robust regression. Wichern et al. (1976) also discussed the variance change problems in their papers. Hsu (1977, 1979) developed two tests one that was locally most powerful and another based on the CUSUMs of Chi Squared values, with an initial variance structure. Davis (1979) studied tests for a single change in the innovations at a specified time point in a univariate AR process. Talwar (1983) Talwar (1983) [229] studied the problem of a shift in location by using several criteria available in the literature.

Detect Change in both Mean and Variance. When simultaneous shifts in mean and variance are considered, one can refer to Chen and Gupta (2012) for a comprehensive list of publications in parametric change point analysis. Horváth (1993) studied the theoretical properties of likelihood ratio test and modified information criterion respectively. Krishnaiah et al. (1990) suggested a so-called local likelihood method for solving the change point problems when the data is distributed as multivariate normal. Horvath (1993) developed the asymptotic distribution of the likelihood ratio statistic. Barry and Hartigan (1992) proposed a product

partition model (PPM) to identify multiple change points, and similar job was done by Loschi and Cruz (2002, 2005) later. Then a modified information criterion (MIC), which penalizes the distance between all preferred change points, was developed by Pan and Chen (2006) for multiple change points detection problems under the assumption that the number of change points is known. This criterion not only works under the normal model, but also can be applied to any other non normally distributed models.

1.2.2 Nonparametric 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. Parametric analysis necessarily assumes that the underlying distributions belong to some known family, and the likelihood function plays a major role, while Nonparametric approaches often rely on the estimation of density functions. Johnson and Bhattacharyya (1968) developed a locally best invariant test, assuming the initial mean was known and applied it to the normal distribution as well as the double exponential. Pettitt (1979) developed a well-known test which uses the maximum of Wilcoxon's test statistics to test all possible change point. One of the most important generalizations is the so-called maximum Kruskal-Wallis statistic, which is a function of the change points, based on the Kruskal-Wallis test. A minimally selected p method for locating single changes in a Binomial Sequence was proposed by Halpern (1999). Later a method for locating multiple change points in a Binomial Sequence was proposed in Halpern (2000). Rozenholc (2001) who developed two Kolmogorov-Smirnov type statistics for the identification of changes in the spectral characteristics of Gaussian tapered data. Hariz et al. (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. Hušková and Kirch (2008) used a block bootstrapping method

to construct confidence intervals for the unknown change point in the mean. Matteson and James (2014a) developed a nonparametric approach by calculating the Euclidean distances between sample observations. Their retrospective analysis of an multidimensional time series data was based on hierarchical clustering. They implemented both divisive and agglomerative algorithms in an R package called “ecp”.

1.3 Smoothing Spline Models

The framework of penalized likelihood method adopted in this study is based on Wahba (1990) and Gu (2013). Given stochastic data “generated” according to an unknown “pattern” function f_0 , the penalized likelihood method estimates f_0 by minimizing a score of the form

$$L(f|\text{data}) + \frac{\lambda}{2}J(f), \quad (1.1)$$

where $L(f)$, usually the negative log likelihood, measures the goodness-of-fit of f , $J(f)$, the roughness penalty, measures the smoothness of f , and the smoothing parameter $\lambda(> 0)$ controls the trade off. The minimization of ((1.1)) is done in a reproducing kernel Hilbert space (RKHS) \mathcal{H} of functions. RKHS provides a theoretical basis for Smoothing Spline ANOVA (SSANOVA) model and unified framework for modeling various data. For multivariate f , it can be decomposed into main effects and interactions similar to the classical ANOVA decomposition.

Through proper specifications of f and $J(f)$ in a variety of problem settings, ((1.1)) yields nonparametric models for Gaussian and non-Gaussian regression, probability density estimation, hazard rate estimation, etc. Kimeldorf and Wahba (1970a), Kimeldorf and Wahba (1970b) and Kimeldorf and Wahba (1971) first proposed penalized least squares regression in univariate case. The general problem of penalized least squares regression with multiple penalty terms was formulated by Wahba (1986). Non-Gaussian regression in such context can be found in Gu (1990) and Gu and Xiang (2001). The penalized likelihood method in the context of density estimation was studied by Good and Gaskins (1971), Wahba et al. (2001) and Gu and Qiu (1993). The formulation of penalized likelihood hazard estimation was proposed by Gu (1996). The setting relevant to our problem is the penalized (weighted) least squares regression. In Chapter 2 we shall deal with regression models on a univariate time domain and in Chapter 3 we extend the model to regression on spaces of time and

location. Therefore, in this chapter we shall introduce in this order: the general formulation for smoothing spline regression, its details for the univariate regression, the thin-plate splines for spatial regression, and the tensor product of thin-plate and cubic smoothing splines for a temporospatial domain.

1.3.1 Penalized (Weighted) Least Squares Estimation

Observing $Y_i = f(x_i) + \epsilon_i, i = 1, \dots, n$, with $\boldsymbol{\epsilon} \sim N(0, \Sigma)$, the minus log likelihood functional $L(f)$ in $L(f) + (\lambda/2)J(f)$ of ((1.1)) reduces to the least squares functional proportional to $\sum_{i=1}^n (Y_i - f(x_i))^2$. The general form of penalized weighted least squares functional in a reproducing kernel Hilbert space \mathcal{H} can be written as

$$\frac{1}{n}(\mathbf{y} - \mathbf{f})^T \Sigma^{-1}(\mathbf{y} - \mathbf{f}) + \lambda J(f). \quad (1.2)$$

If we assume $\Sigma = \sigma^2 I$, then ((1.2)) can be reduced to

$$\frac{1}{n}(\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda J(f), \quad (1.3)$$

where $f \in \mathcal{H} = \{f \mid f : \mathcal{X} \rightarrow \mathbb{R}, J(f) < \infty\}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ and $\mathbf{f} = (f(x_1), f(x_2), \dots, f(x_n))^T$ are respectively the vectors of observed responses and fitted values. Note that the common variance σ^2 has been absorbed by the smoothing parameter λ . Here the predictor x_i can be of one dimension such as time (that is, $x_i = t_i$) in Chapter 2, of two dimensions such as location (that is, $x_i = \mathbf{z}_i = (z_{i1}, z_{i2})$), or of three dimensions such as (time, location) (that is, $x_i = (t_i, \mathbf{z}_i)$) as in Chapter 3.

The first part of ((1.2)) is the sum of square errors weighted by the inverse of the covariance matrix Σ . $J(f)$ is a squared semi-norm on \mathcal{H} , acting as a roughness penalty on f , and $\lambda > 0$ is the smoothing parameter balancing the trade-off between the smoothness

of the mean function estimate and the goodness-of-fit represented by the weighted sum of squared errors. \mathcal{H} is a reproducing kernel Hilbert space of functions on the domain \mathcal{X} . A reproducing kernel Hilbert space (RKHS) is a Hilbert space \mathcal{H} where the evaluation functional $[x] : \mathcal{H} \rightarrow \mathbb{R}, f \mapsto f(x)$ is continuous for every $x \in \mathcal{X}$. The Riesz Representation Theorem then indicates that for all $x \in \mathcal{X}$ there exists a unique function $R_x \in \mathcal{H}$ with the reproducing property $\langle R_x, f \rangle = [x](f) = f(x)$, 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 $R : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that $R(s, x) = \langle R_s, R_x \rangle$. One can show that each RKHS is uniquely associated with a reproducing kernel and vice versa.

Note that the penalty functional J in ((1.2)) 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 .

We now introduce an example (See, e.g., Gu (2013, Chapter 2) for more details on RKHSs) of cubic smoothing splines to illustrate these concepts. We shall use the cubic smoothing splines in Chapter 2 of the dissertation.

Example 1.3.1 (Cubic Smoothing Splines). *Without loss of generality assume $\mathcal{X} = [0, 1]$. 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, Section 2.3.3). \square*

The RKHS \mathcal{H} is of infinite dimensions, so a direct optimization of (1.2) on \mathcal{H} seems infeasible. However, since the weighted least squares part in (1.2) depends on f only through its evaluations at the observation points $x_i, i = 1, \dots, n$, the Representer Theorem (Wahba,

1990) guarantees that the exact minimizer of (1.2) actually resides in a finite dimensional subspace of \mathcal{H} , namely, $\mathcal{N}_J \oplus \text{span}\{R_J(x_1, \cdot), \dots, R_J(x_n, \cdot)\}$. Let $\phi_l, l = 1, \dots, m$ be the basis functions of \mathcal{N}_J and $\xi_j = R_J(x_j, \cdot), j = 1, \dots, n$. 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 $R_J(x_i, x_j)$. So for a fixed λ , the objective function (1.2) 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, Section 3.2).

1.3.2 Thin-Plate Spline

In Chapter 3, the variances before and after the change point are both functions of locations on a 2D Euclidean space. They will be estimated by thin-plate splines which we introduce here. Suppose that s_i are independent observations generated from the following model

$$s_i = v_0(\mathbf{z}_i) + \epsilon_i, i = 1, \dots, n \quad (1.4)$$

where $\mathbf{z}_i = (z_{i1}, \dots, z_{id}) \in \mathcal{X} = (-\infty, \infty)^d$ is the location of the observation, ϵ_i are independent random errors from a zero-mean normal distribution. v_0 is an unknown smooth function on the reproducing kernel Hilbert space $\mathcal{H} = \{f : f \text{ is a function on } \mathcal{X} \text{ with } J_r^d(f) < \infty\}$. Here

$$J_r^d(f) = \sum_{\alpha_1 + \dots + \alpha_d = r} \frac{r!}{\alpha_1! \dots \alpha_d!} \times \int \dots \int \left(\frac{\partial^r f}{\partial z_1^{\alpha_1} \dots \partial z_d^{\alpha_d}} \right)^2 dz_d \dots dz_1. \quad (1.5)$$

The estimate of the function v_0 through the minimization of the penalized least squares (1.3) with J replaced by J_r^d is called a d -dimensional thin-plate splines function with order r . The null space of $J_r^d(f)$ in \mathcal{H} consists of polynomials of up to $(r - 1)$ total order, which is of dimension $M = \binom{d+r-1}{d}$. The functional $J_r^d(f)$ is invariant under a rotation of the

coordinates. It is a square semi-norm on \mathcal{H} , and acts as a roughness penalty. It is necessary that $2m - d > 0$ for the evaluation functional $[\mathbf{z}]f = f(\mathbf{z})$ to be continuous; see Duchon (1977), Meinguet (1979), and Wahba and Wendelberger (1980) for more details.

When the parametric least squares estimate in the null space of $J_r^d(f)$ uniquely exists, the minimizer f_λ of the PLS uniquely exists; see Gu (2013, Theorem 2.9). From Duchon (1977, Theorem 4 bis), f_λ has an expression

$$f_\lambda(\mathbf{z}) = \sum_{\nu=1}^r d_\nu \phi_\nu(\mathbf{z}) + \sum_{i=1}^n c_i E(|\mathbf{z}_i - \mathbf{z}|), \quad (1.6)$$

where $\{\phi_\nu\}_{\nu=1}^M$ span the null space of $J_r^d(f)$, c_i 's are subject to the constraints $S^T \mathbf{c} = \mathbf{0}$ with S the $n \times M$ matrix with the (i, ν) th entry $\phi_\nu(\mathbf{z}_i)$, $|\mathbf{z} - \mathbf{y}|$ is the Euclidean distance, and

$$E(\mathbf{z}) = \begin{cases} \theta_{r,d} \mathbf{z}^{2r-d} \log \mathbf{z}, & d \text{ even, for } \theta_{r,d} = \frac{(-1)^{d/2+r+1}}{2^{2r-1} \pi^{d/2} (r-1)! (r-d/2)!}, \\ \theta_{r,d} \mathbf{z}^{2r-d}, & d \text{ odd, for } \theta_{r,d} = \frac{\Gamma(d/2-r)}{2^{2r} \pi^{d/2} (r-1)!}. \end{cases} \quad (1.7)$$

The constant $\theta_{r,d}$ in (1.7) is not really needed for (1.6), as it is readily absorbed into c_i 's. However, the reproducing kernels, to be introduced shortly, are expressed in terms of $E(\mathbf{z})$ with $\theta_{r,d}$ attached.

Example 1.3.2. With $d = 2$ and $r = 2$, one has $J_2^2(f) = \iint (\ddot{f}_{(11)}^2 + 2\ddot{f}_{(12)}^2 + \ddot{f}_{(22)}^2) dz_1 dz_2$.

Obviously, $d/2 + r + 1 = 4$ is even, so $E(|\mathbf{z} - \mathbf{y}|) \propto |\mathbf{z} - \mathbf{y}|^2 \log |\mathbf{z} - \mathbf{y}|$. It follows that

$$\eta_\lambda(\mathbf{z}) = d_1 + d_2 z_1 + d_3 z_2 + \sum_{i=1}^n c_i |\mathbf{z}_i - \mathbf{z}|^2 \log |\mathbf{z}_i - \mathbf{z}|,$$

with \mathbf{c} and \mathbf{d} the solution of the constrained minimizer of $f_\lambda(\mathbf{z})$ in (1.6).

Next we will introduce the reproducing kernels for thin-plate splines. Let ψ_ν be a set of

polynomials that span \mathcal{N}_J , the null space of $J_r^d(f)$. Define

$$(f, g)_0 = \sum_{i=1}^N p_i f(u_i) g(u_i), \quad (1.8)$$

where $u_i \in (-\infty, \infty)^d$, $p_i > 0$, $\sum_{i=1}^N p_i = 1$ are specified such that the Gram matrix with the (ν, μ) th entry $(\psi_\nu, \psi_\mu)_0$ is nonsingular. Following some standard orthogonalization procedure, one can find an orthonormal basis ϕ_ν , $\nu = 1, \dots, M$, for \mathcal{N}_J with $\phi_1(\mathbf{z}) = 1$ and $(\phi_\nu, \phi_\mu)_0 = \delta_{\nu, \mu}$, where $\delta_{\nu, \mu}$ is the Kroneker delta. The RK in \mathcal{N}_J is seen to be

$$R_0(\mathbf{z}, \mathbf{y}) = \sum_{\nu=1}^M \phi_\nu(\mathbf{z}) \phi_\nu(\mathbf{y}), \quad (1.9)$$

and for fixed \mathbf{x} , one has,

$$\begin{aligned} R_1(\mathbf{z}, \mathbf{u}) &= E(|\mathbf{z} - \mathbf{u}|) - \sum_{\nu=1}^M \phi_\nu(\mathbf{z}) \sum_{i=1}^N p_i \phi_\nu(\mathbf{u}_i) E(|\mathbf{u}_i - \mathbf{u}|) + \pi(\mathbf{u}) \\ &= E(|\mathbf{z} - \mathbf{u}|) - \sum_{i=1}^N c_i(\mathbf{z}) E(|\mathbf{u}_i - \mathbf{u}|) + \pi(\mathbf{u}) \end{aligned} \quad (1.10)$$

where p_i is the number of basis ϕ_ν in R , $\pi(\mathbf{u}) \in \mathcal{N}_J$ and $c_i(\mathbf{z}) = -\sum_{\nu=1}^M p_i \phi_\nu(\mathbf{u}_i) \phi_\nu(\mathbf{z})$.

1.3.3 Tensor product of Cubic splines and Thin-Plate Splines

In my dissertation, we will use tensor produce of cubic splines and thin-plate splines to estimate the mean temperature as a function of both time and location. Without loss of generality, we will introduce only a special case in this section: the tensor product of cubic splines and the thin-plate splines with $r = 2$ and $d = 2$. Consider covariates on the product space $\mathcal{T} \times \mathcal{Z}$. One continuous covariate, say $x \in \mathcal{T} = [0, 1]$, and the other covariate, say $\mathbf{z} = \{z_1, z_2\} \in \mathbb{R}^2$, is a vector of longitude and latitude location. The RKHS $\mathcal{H}_{\langle \mathbf{Z} \rangle}$ on the 2D location domain can be decomposed as $\mathcal{H}_{00\langle \mathbf{Z} \rangle} = \text{span}\{1\}$, $\mathcal{H}_{01\langle \mathbf{Z} \rangle} = \text{span}\{z_1, z_2\}$, and $\mathcal{H}_{1\langle \mathbf{Z} \rangle} = \mathcal{H}_{\langle \mathbf{Z} \rangle} - (\mathcal{H}_{00\langle \mathbf{Z} \rangle} + \mathcal{H}_{01\langle \mathbf{Z} \rangle})$

with RKs $R_{00\langle Z \rangle}$, $R_{01\langle Z \rangle}$, and $R_{1\langle Z \rangle}$.

On the other hand, the construction in Example 1.3.1 gives a decomposition of the RKHS $\mathcal{H}_{\langle T \rangle}$ on the \mathcal{T} domain

$$\begin{aligned}\mathcal{H}_{\langle T \rangle} &= \{f : \int_0^1 (f'')^2 dx < \infty\} = \mathcal{H}_{00\langle T \rangle} \oplus \mathcal{H}_{01\langle T \rangle} \oplus \mathcal{H}_{1\langle T \rangle} \\ &= \text{span}\{1\} \oplus \text{span}\{k_1(x)\} \oplus \{f : \int_0^1 f dx = \int_0^1 \dot{f} dx = 0, \int_0^1 (f'')^2 dx < \infty\},\end{aligned}$$

with RKs $R_{00\langle T \rangle}(v_1, v_2) = 1$, $R_{01\langle T \rangle}(v_1, v_2) = k_1(v_1)k_1(v_2)$, and $R_{1\langle T \rangle}(v_1, v_2) = k_2(v_1)k_2(v_2) - k_4(v_1 - v_2)$. Taking tensor product of $\mathcal{H}_{\langle T \rangle}$ and $\mathcal{H}_{\langle Z \rangle}$, one obtains six tensor sum terms $\mathcal{H}_{\nu, \mu} = \mathcal{H}_{\nu\langle T \rangle} \otimes \mathcal{H}_{\mu\langle Z \rangle}$ on $\mathcal{T} \times \mathcal{Z}$, $\nu = 00, 01, 1$ and $\mu = 00, 01, 1$, with RKs $R_{\nu, \mu}(x_1, x_2) = R_{\nu}(v_1, v_2)R_{\mu}(u_1, u_2)$, where $x_i = (v_i, u_i)$. The two subspaces with $\nu = 00, 01$ and $\mu = 00, 01$ can be lumped together as \mathcal{N}_J . The other four subspaces can be put together as \mathcal{H}_J with the RK

$$R_J = \theta_{00,1}R_{00\langle T \rangle, 1\langle Z \rangle} + \theta_{01,1}R_{01\langle T \rangle, 1\langle Z \rangle} + \theta_{1,00}R_{1\langle T \rangle, 00\langle Z \rangle} + \theta_{1,01}R_{1\langle T \rangle, 01\langle Z \rangle} + \theta_{1,1}R_{1\langle T \rangle, 1\langle Z \rangle},$$

where $\theta_{\nu, \mu}$ are a set of extra smoothing parameters adjusting the relative weights of the roughness of different components. For more detail about multiple term RKHS with multiple smoothing parameters, see Gu (2013, Section 2.4.5).

For interpretation, the six subspaces readily define an ANOVA decomposition

$$f(T, Z) = f_{\emptyset} + f_T(T) + f_Z(Z) + f_{T,Z}(T, Z)$$

for functions on $\mathcal{T} \times \mathcal{Z}$, with $f_{\emptyset} \in \mathcal{H}_{00\langle T \rangle} \otimes \mathcal{H}_{00\langle Z \rangle}$ being the constant term, $f_T \in \{\mathcal{H}_{01\langle T \rangle} \oplus \mathcal{H}_{1\langle T \rangle}\} \otimes \mathcal{H}_{00\langle Z \rangle}$ the T main effect, $f_Z \in \mathcal{H}_{00\langle T \rangle} \otimes \{\mathcal{H}_{01\langle Z \rangle} \oplus \mathcal{H}_{1\langle Z \rangle}\}$ the Z main effect, and $f_{T,Z} \in \{\mathcal{H}_{01\langle T \rangle} \oplus \mathcal{H}_{1\langle T \rangle}\} \otimes \{\mathcal{H}_{01\langle Z \rangle} \oplus \mathcal{H}_{1\langle Z \rangle}\}$ the interaction.

1.4 New Challenges

As we have reviewed above, change point detection is a classical topic that has attracted a lot of 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 dissertation, we are concerned with variance change point detection under a smoothly-changing mean trend. In some cases, variance shifts is a more important criterion, since it is a clear sign of system failure as in our liver procurement experiment. Particularly, the constantly changing mean trend violates the assumptions of most existing change point detection methods. As demonstrated in my dissertation, a naïve application of these existing methods to such kind of data would yield erratic change point estimates.

In Chapter 2, we propose a new method, which focus on a single location on the surface of the organ. Particularly, we developed a penalized weighted least squares approach with an iterative estimation procedure that naturally integrates variance change point detection and smooth mean function estimation. 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 the non-invasive organ viability assessment by surface temperature monitoring.

In Chapter 3, we will extend our method to study the variance change point for a chunk of the organ surface with 2D locations (longitude and latitude) involved. We will use thin-plate splines to estimate variances as functions on a two-dimensional surface and the mean as a function on a three-dimensional space of time and location. Recall that the whole organ surface

in our data consists of over 36,000 grid points. It is easy to get thousands of observations even for only a small chunk of it. With such a large sample size, a standard thin-plate splines model would incur serious computational issues even equipped with the recent efficient reduction methods. Therefore in this part of the dissertation we propose a subsampling strategy that can be incorporated into the iterative algorithm introduced in Chapter 2. The strategy has three components: the first is subsampling locations, the second is downsampling times around a preliminary estimate of the change point, and the last is a subsampling based on the preliminary estimates. Our simulations show that the proposed computational strategy can significantly reduce the computational cost, essentially making it manageable, as well as deliver a reasonable estimation performance. When applying the method to the procurement data, we can provide a change point result that relays immediately useful information to practitioners.

Chapter 2 Variance Change Point Detection under a Smoothly-changing Mean Trend: The Univariate Case

2.1 Introduction

As reviewed in Chapter 1, 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 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 et al. (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 (2014b) 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 in our experiment where the variance change happened underneath a smoothly-changing mean trend. Particularly, 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 capture the gradually changing mean trend. As demonstrated in our numerical experiments, erratic behavior occurs 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 paper 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 Sobolev space. We note that these results have their own theoretical values too. Testing procedures under nonparametric null and alternative hypotheses are 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 this chapter is organized as follows. In Section 3.2, we introduce in the order: the notation and model, the iterative algorithm, the mean estimation given the variances and change point, the test procedure for variance change point give the mean function, and the theoretical properties of the proposed method. In Section 3.3 we present all the simulations. We analyze the liver procurement data in Section 3.4. Discussion in Section 3.5 concludes this chapter. Proofs of the theorems are collected in the Appendix.

2.2 Method

2.2.1 Notation and Model

Suppose that y_i are independent observations generated from the following model

$$y_i = f_0(i/n) + \epsilon_i, i = 1, \dots, n, \quad (2.1)$$

where f_0 is an unknown smooth function, $\epsilon_i \sim N(0, \sigma_i^2)$ with $\sigma_i = \sigma_0$ when $i \leq \tau_0$ 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 $\mathcal{H} = \{f | f : [0, 1] \rightarrow \mathbb{R}, J(f) < \infty\}$, where J is a semi-norm on \mathcal{H} . For example, we consider $J(f) = \int_0^1 \{f^{(m)}(t)\}^2 dt$ in this paper 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

$$\frac{1}{n}(\mathbf{y} - \mathbf{f})^T \Sigma_{n,\tau,\sigma,\delta}^{-1}(\mathbf{y} - \mathbf{f}) + \lambda J(f), \quad (2.2)$$

where f is a function in \mathcal{H} , $\mathbf{y} = (y_1, \dots, y_n)^T$ and $\mathbf{f} = (f(1/n), f(2/n), \dots, f(1))^T$ 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 (3.2) does not exist since it approaches zero as σ^2 goes to infinity. Hence, we propose the estimates $(\hat{f}, \hat{\tau}, \hat{\sigma}^2, \hat{\delta}^2)$ as the local minimizer of (3.2) obtained through the following iterative algorithm. We shall show in Section 2.2.4 that the

Algorithm I

1. Initialize $\hat{f}^{(0)}$ with the mean function estimate assuming constant variance. That is, $\hat{f}^{(0)}$ minimizes

$$\frac{1}{n}(\mathbf{y} - \mathbf{f})^T(\mathbf{y} - \mathbf{f}) + \lambda J(f). \quad (2.3)$$

Note that when $\sigma^2 = \delta^2$, the covariance matrix in (3.2) 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 ι th iteration,
 - (a) given the mean estimate $\hat{f}^{(\iota-1)}$, we first use the testing procedure in Section 2.2.3 to find an estimate $\hat{\tau}^{(\iota)}$ for τ_0 . Then we estimate the variance parameters respectively by the maximum likelihood variance estimates, $[\hat{\sigma}^2]^{(\iota)}$ and $[\hat{\delta}^2]^{(\iota)}$, of the subsequences of residuals, $\{y_i - \hat{f}^{(\iota-1)}(i/n) : i = 1, \dots, \hat{\tau}^{(\iota)}\}$ and $\{y_i - \hat{f}^{(\iota-1)}(i/n) : i = \hat{\tau}^{(\iota)} + 1, \dots, n\}$.
 - (b) Now given the estimates $\hat{\tau}^{(\iota)}$, $[\hat{\sigma}^2]^{(\iota)}$ and $[\hat{\delta}^2]^{(\iota)}$, we update the mean estimate by the minimizer of (3.2) where τ , σ^2 and δ^2 are replaced respectively by their current estimates.
 3. Iterate until the algorithm converges.
-

Table 2.1: Iterative Algorithm I

estimates are consistent with proper rates of convergence.

The above algorithm is general in allowing multiple iterations. As will be demonstrated in Theorem 2.2.1, 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 2.2.1 if the initial estimator \hat{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 criteria 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.

2.2.2 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 (3.2) 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 $[t] : \mathcal{H} \rightarrow \mathbb{R}, f \mapsto f(t)$ 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 $R : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ 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 (3.2) 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, e.g., Gu (2013, Chapter 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 paper.

Example 2.2.1 (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, Section 2.3.3). \square*

The RKHS \mathcal{H} is of infinite dimensions, so a direct optimization of (3.2) on \mathcal{H} seems infeasible. However, since the weighted least squares part in (3.2) depends on f only through its evaluations at the observation points $t_i, i = 1, \dots, n$, the Representer Theorem (Wahba, 1990) guarantees that the exact minimizer of (3.2) actually resides in a finite dimensional subspace of \mathcal{H} , namely, $\mathcal{N}_J \oplus \text{span}\{R_J(t_1, \cdot), \dots, R_J(t_n, \cdot)\}$. Let $\phi_l, l = 1, \dots, m$ be the basis functions of \mathcal{N}_J and $\xi_j = R_J(t_j, \cdot), j = 1, \dots, n$. 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 (3.2) 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, Section 3.2).

2.2.3 Variance Change Point Detection Given f

Given \hat{f} , we now introduce a testing procedure to find an estimate $\hat{\tau}$ for the variance change point τ_0 . Then we compute the maximum likelihood estimates for σ^2 and δ^2 respectively by $\hat{\sigma}^2 = \hat{\tau}^{-1} \sum_{i=1}^{\hat{\tau}} \{y_i - \hat{f}(i/n)\}^2$ and $\hat{\delta}^2 = (n - \hat{\tau})^{-1} \sum_{i=\hat{\tau}+1}^n \{y_i - \hat{f}(i/n)\}^2$. 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

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2 \text{ versus } H_1 : \sigma_1^2 = \dots = \sigma_\tau^2 \neq \sigma_{\tau+1}^2 = \dots = \sigma_n^2, \quad (2.4)$$

for a potential change point position τ . Let

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

Note that $\ell(n) = -2L_0(\hat{\sigma}^2) - n - n \log 2\pi$ and $\ell(\tau) = -2L_1(\hat{\sigma}^2, \hat{\delta}^2) - n - n \log 2\pi$, where L_0 and L_1 are respectively the log likelihood functions under the null and alternative hypotheses of (3.4). 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 in 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 $-2 \log L(\hat{\theta}) + p \log n$, where $L(\hat{\theta})$ is the likelihood function for the model, $\hat{\theta}$ is the maximum likelihood estimate of the parameter θ , and p is the dimension of θ . In our case, given f and τ we have two models corresponding to the null and alternative hypotheses with their SICs respectively defined by $\text{SIC}(n) = -2L_0(\hat{\sigma}^2) + \log n$ and $\text{SIC}(\tau) = -2L_1(\hat{\sigma}^2, \hat{\delta}^2) + 2 \log n$. By the principle of minimum information criterion, we do not reject H_0 if $\text{SIC}(n) \leq \min_{\tau} \text{SIC}(\tau)$, or equivalently $\ell(n) \leq \min_{1 < \tau < n} \ell(\tau)$, 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 $\hat{\tau} = \arg \min_{1 < \tau < n} \ell(\tau)$. 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 2.2.4.

2.2.4 Theoretical Properties

In this section we present the asymptotic theories for the proposed method. For simplicity, we only consider the special case when \mathcal{H} is the m th 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 $k = 1, 2, \dots$, $\varphi_{2k-1}(t) = \sqrt{2} \cos(2\pi kt)$, $\varphi_{2k}(t) = \sqrt{2} \sin(2\pi kt)$, and $\gamma_{2k-1} = \gamma_{2k} = (2\pi k)^{2m}$. Note that $J(f) = \int_0^1 \{f^{(m)}(t)\}^2 dt = \sum_{\nu=1}^{\infty} f_{\nu}^2 \gamma_{\nu}$ for $f \in S^m$ and $R_J(s, t) = (2\pi m)^{-2m} \sum_{\nu=1}^{\infty} 2 \cos\{2\pi \nu s\} \cos\{2\pi \nu t\}$.

$t)\}/(2\pi m\nu)^{2m}$.

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

Theorem 2.2.1 (Consistency of Parameter Estimates). *Under Conditions 1-3 in the Appendix, the estimates $(\hat{f}, \hat{\tau}, \hat{\sigma}^2, \hat{\delta}^2)$ from the algorithm in Section 3.2.1 are consistent with the following rates of convergence:*

$$\begin{aligned} \|\hat{f} - f_0\|_n^2 &= O_P(\lambda + (nh)^{-1} + h^{-1}\tilde{r}_n^2), \quad |\hat{\tau} - \tau_0| = O_P((\log n)^4(\log \log n)^2), \\ |\hat{\sigma}^2 - \sigma_0^2| &= O_P(\tilde{r}_n), \quad |\hat{\delta}^2 - \delta_0^2| = O_P(\tilde{r}_n), \end{aligned}$$

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

Note that when $m \geq 1$ and $\lambda \asymp n^{-2m/(2m+1)}$, it can be verified that $\tilde{r}_n = O(n^{-1/2})$. Then this implies that $\hat{\sigma}^2$ and $\hat{\delta}^2$ are \sqrt{n} -consistent, and that $\|\hat{f} - f_0\|_n = O_P(n^{-m/(2m+1)})$ or \hat{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 (3.4).

Theorem 2.2.2 (Asymptotic Null Distribution of Test Statistic). *Suppose that as $n \rightarrow \infty$, $h \rightarrow 0$ and $r_n^2 \log n \rightarrow 0$. Under H_0 in (3.4) and Conditions 1, 2, and 3' in the Appendix, for any $t \in \mathbb{R}$,*

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

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 $1 - \alpha$:

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

2.3 Simulations

2.3.1 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 in Matteson and James (2014b) 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 in 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 $\text{MSE} = n^{-1} \sum_{i=1}^n \{\hat{f}(i/n) - f_0(i/n)\}^2$ and the log ratios $\log(\hat{\sigma}^2/\sigma_0^2)$ and $\log(\hat{\delta}^2/\delta_0^2)$.

Figure 2.1 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 paper.

Figure 2.2 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

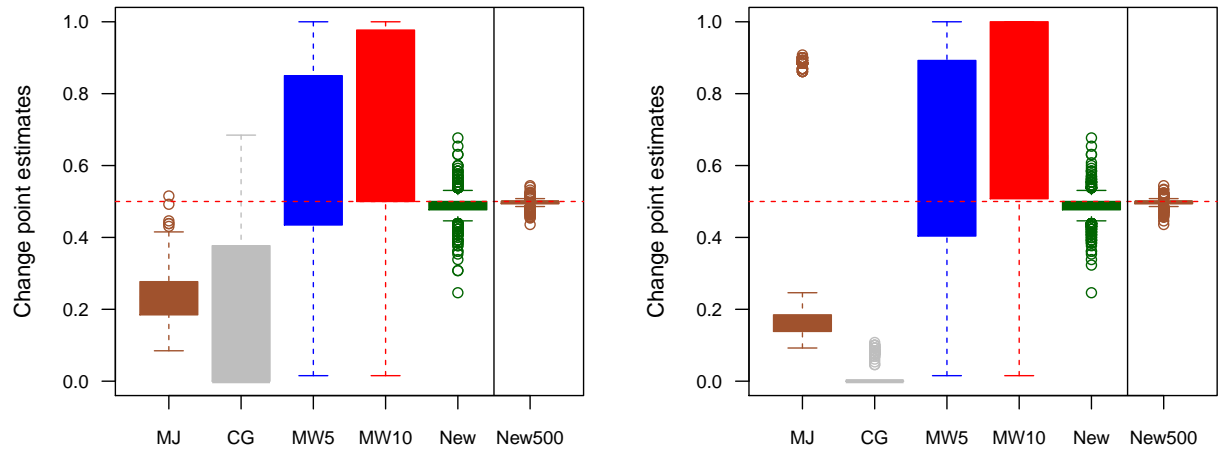


Figure 2.1: Boxplots of change point estimates. Left panel: Simulations with the true mean function= f_{01} ; Right panel: Simulations with the true mean function= f_{02} . The five plots on the left in each panel were the change point estimates with $n = 130$ respectively for the methods in Matteson and James (2014b) (MJ), Chen and Gupta (1997) (CG), moving-window approaches with 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 $\tau_0/n = 0.5$.

area around the change point, which was reasonable considering the fluctuations in that area. Also, the estimation accuracies improved as the sample size increased.

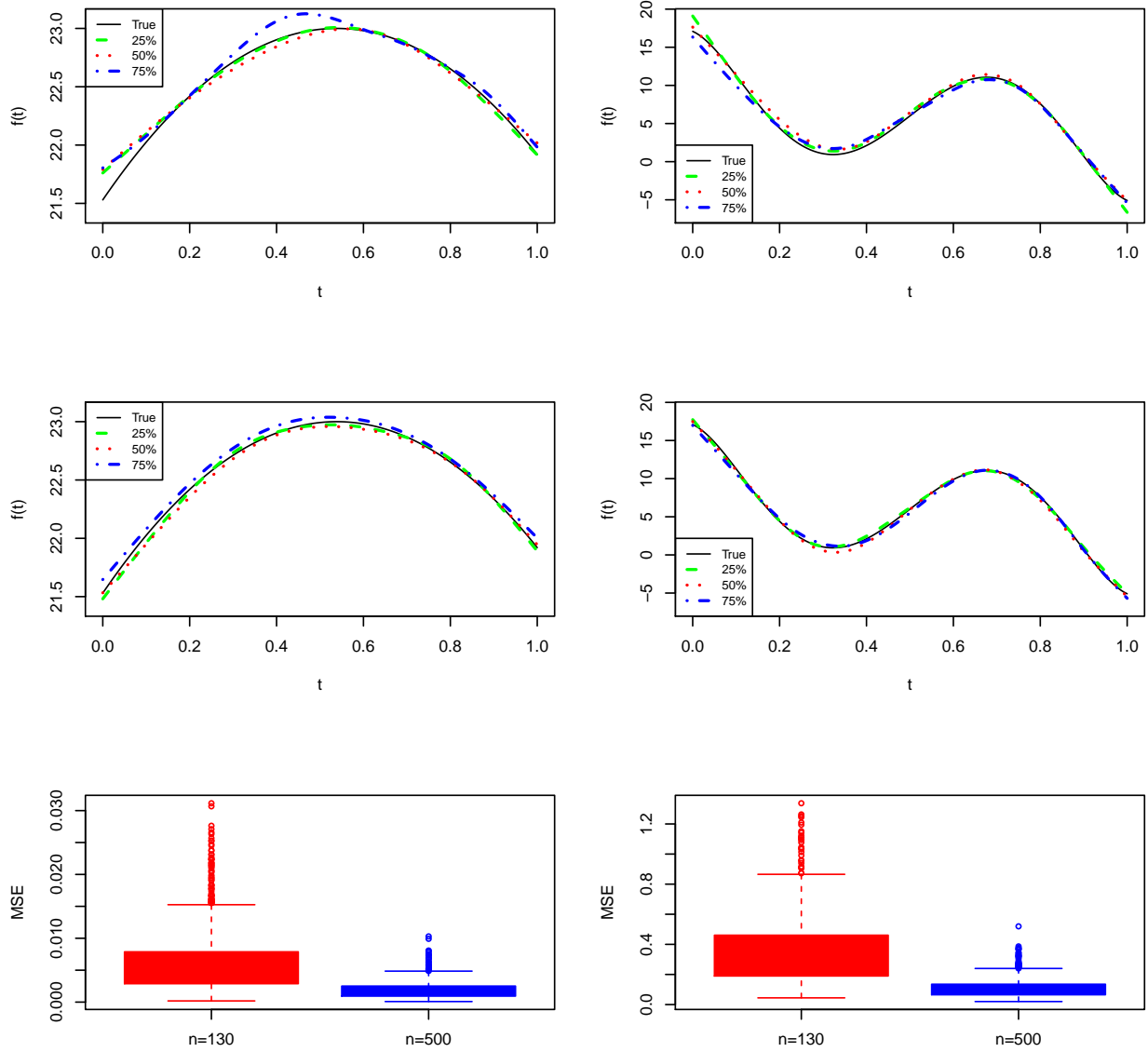


Figure 2.2: Plots for assessing mean estimation performance. Left panels: Simulations with the true mean function = f_{01} ; Right panels: Simulations with the true mean function = f_{02} . Top: True mean function (solid black) versus the mean estimates with $n = 130$ whose MSE were the 25th (dashed 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 2.3 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 estimated with the accuracies also improved as the sample size increased.

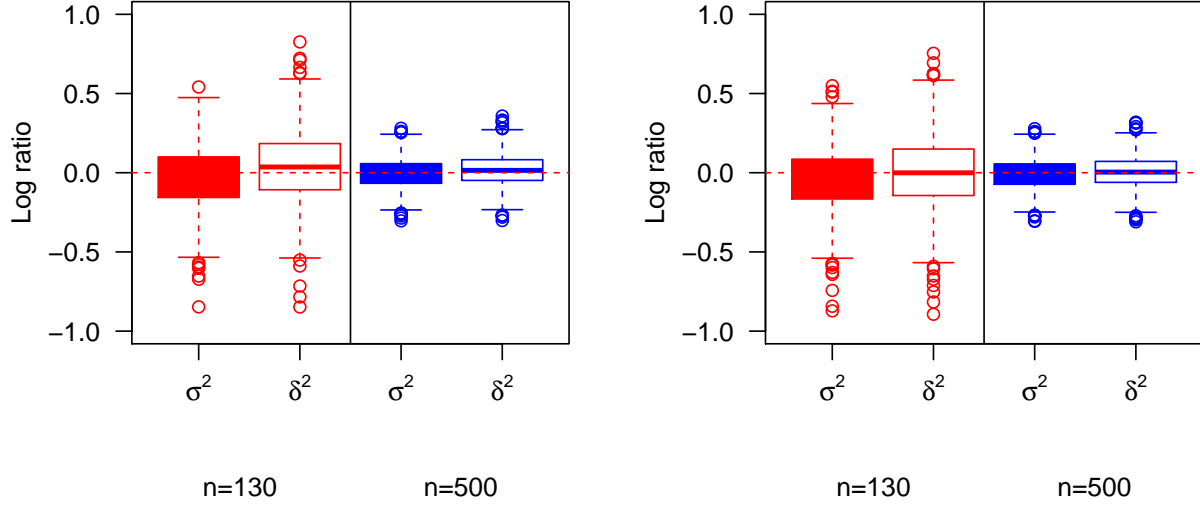


Figure 2.3: Boxplots of the log ratios of variance estimates versus true variances. Left panels: Simulations with the true mean function= f_{01} ; Right panels: Simulations with the true mean function= f_{02} . Red: $n = 130$; Blue: $n = 500$. Filled boxes: σ^2 ; Unfilled boxes: δ^2 .

2.3.2 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 $\theta = 1$, $\delta_0^2 = \sigma_0^2$ 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 $n = 130, 500, 2000, 10000$. For each combination of α and n , we simulated 10000 data replicates. Table 2.2 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 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.

Test Level	Sample Size n			
α	130	500	2000	10000
0.05	0.0097	0.0114	0.0141	0.015
0.1	0.0346	0.04	0.0461	0.0504

Table 2.2: Test size simulation result: proportions of rejections when H_0 is true. The true mean function is $f_{01}(t)$.

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 $\{1.25, 1.50, \dots, 5.00\}$. 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 10000 data replicates. The resulting power function plots are in Figure 2.4. 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.

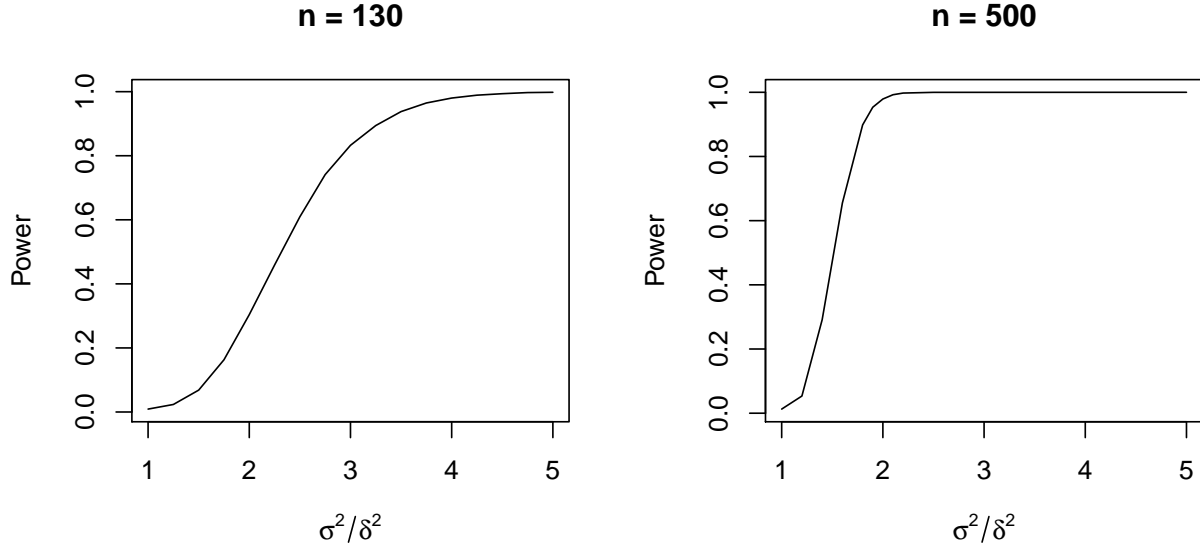


Figure 2.4: Plots of power against ratio of the variances.

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 paper, a lobe of porcine liver, as shown in Figure 2.5, 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 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 legitimate following the principle of the BHY procedure.

The heat map of all the estimated change points were plotted in Figure 2.5. 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 severing and perfusion.

Figure 2.6 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 paces at the three spots in the first 12 hours or so

and shared a common trend of a quicker 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.

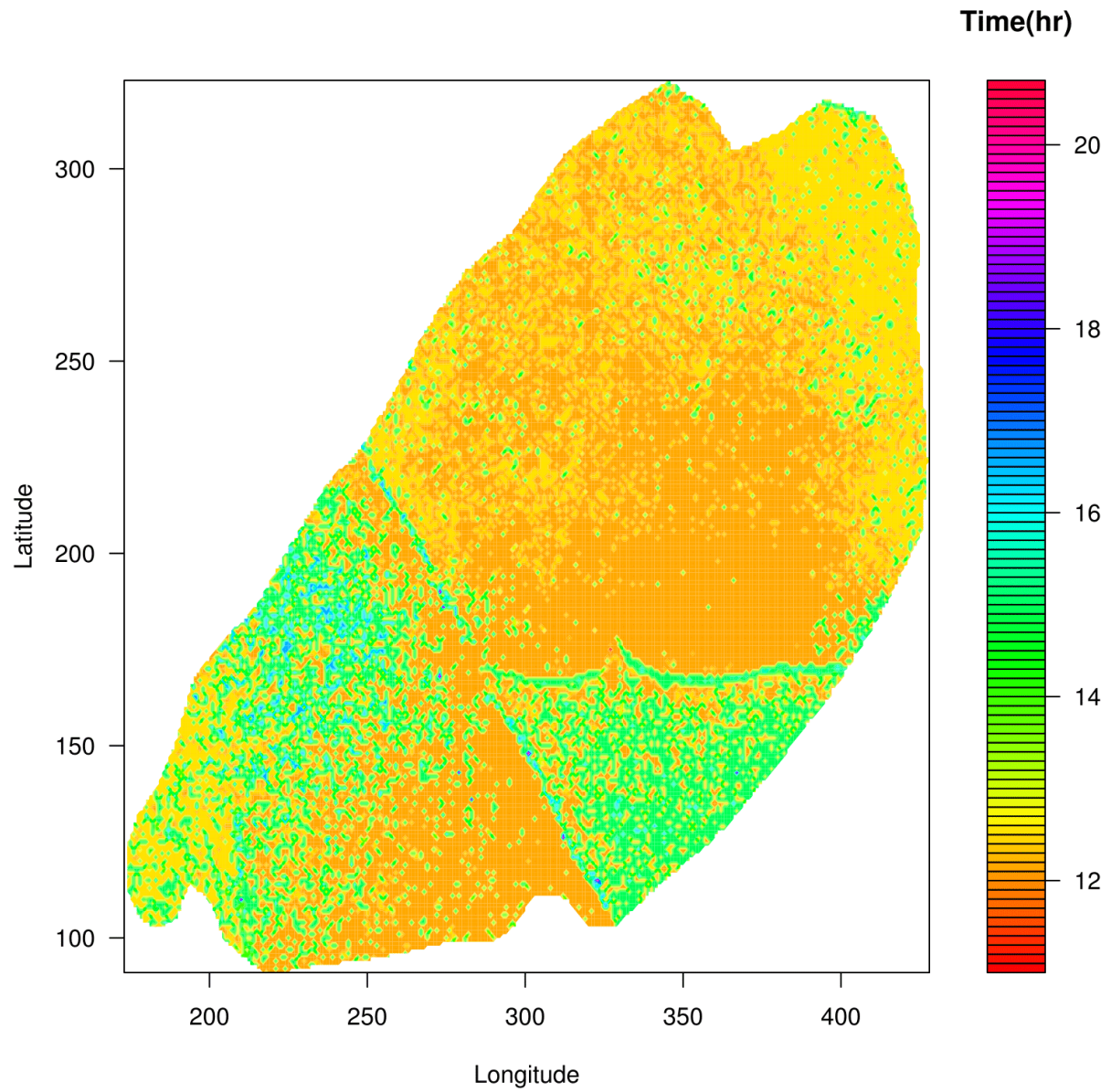


Figure 2.5: The heat map of estimated variance change points of temperatures on the lobe of liver in the procurement experiment.

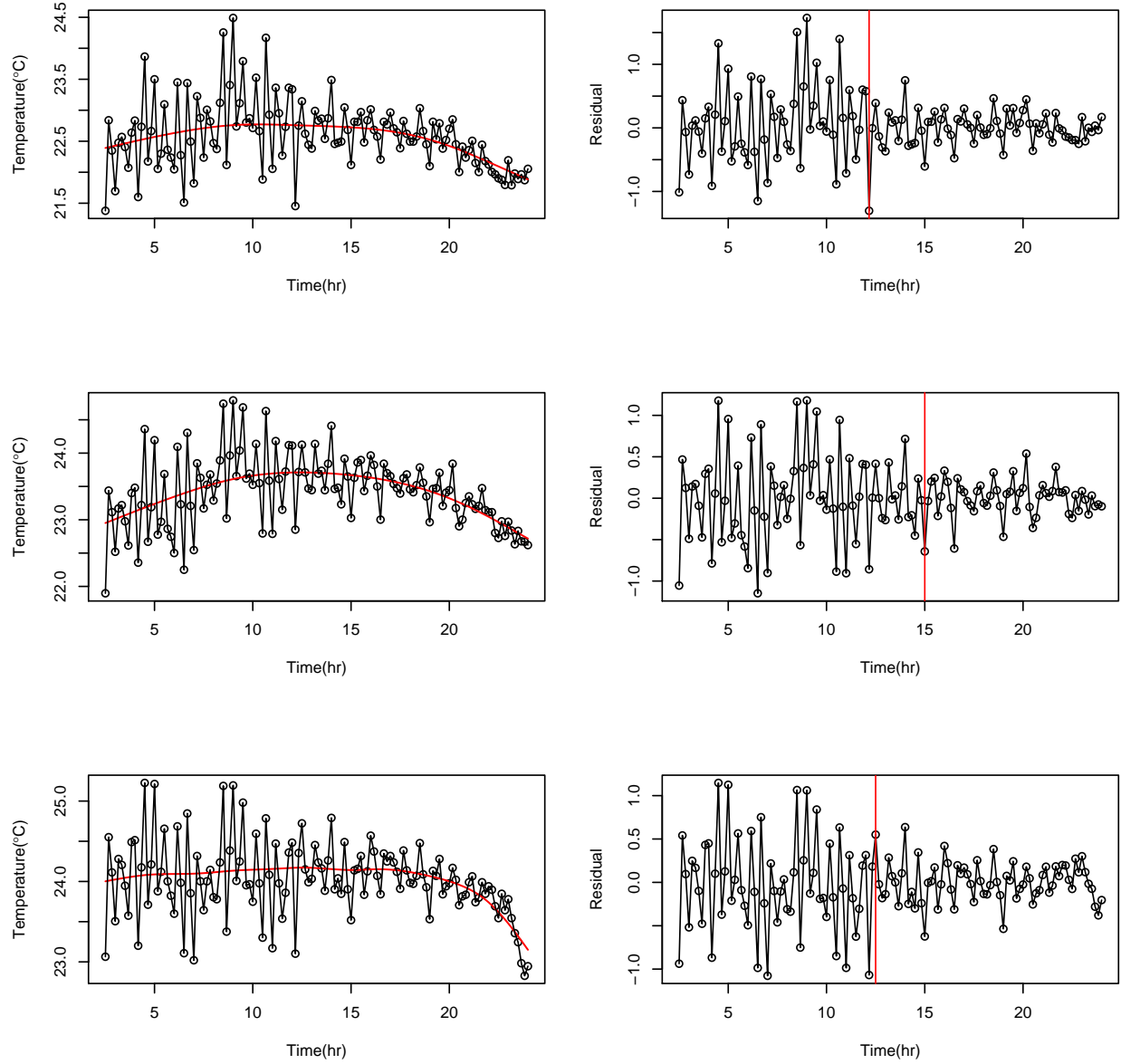


Figure 2.6: Example of mean and variance change point estimates imposed respectively on the raw and de-trended temperature profiles at three randomly selected spots.

2.5 Conclusion

In this chapter, 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 handled 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 also has its own innovation in the perspective of nonparametric regression with non-IID errors. The comparison simulations show that having an accurately estimated mean function is indeed critically important to the correct identification of the variance change point. The application of our method to the liver procurement experiment 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 derived with a combination of a proper characterization of in-control data.

It is also of interest to understand why the global minimizer of (3.2) 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 lets 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 (3.2) by introducing some penalty on the variance parameters into (3.2) 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 paper.

SUPPLEMENTARY MATERIAL

The supplementary material collects all the conditions and technical proofs for the theoretical results in Section 2.2.4.

A.1 Conditions and Technical Lemmas

Conditions:

1. Suppose that when there is a variance change point the true change point $\tau_0 \in [cn/\log n, n - cn/\log n]$ for some $c > 0$. And assume that $\tau_0/n \rightarrow q_0 \in (0, 1)$ as $n \rightarrow \infty$.
2. The true mean function $f_0 \in S^m$, the m th order Sobolev space of periodic functions on $[0, 1]$ with period 1.
3. The random errors $\epsilon_i, i = 1, \dots, n$ are independent normal random variables with mean 0 and variance σ_i^2 , where $\sigma_i = \sigma_0$ when $i \leq \tau_0$ and $\sigma_i = \delta_0$ when $i > \tau_0$.
- 3'. The random errors $\epsilon_i, i = 1, \dots, n$ are independent and identically distributed normal random variables with mean 0 and variance σ_0^2 .

Condition 1 is common in change point analysis literature. It basically ensures that the change point is away from the boundaries. Condition 2 restricts our theory to the case when $\mathcal{H} = S^m$. The periodic RKHS is necessary when the design is fixed. A more general function space \mathcal{H} is possible, but the matrix calculation involved in the technical proof would be much harder. For the general RKHS with a random design, the theory needs substantial modification. For instance, one needs to first arrange the random covariates in a descending order and then establish the SIC test based on the ordered samples. Since the ordered samples

are not independent, the theoretical results do not directly go through. New techniques are needed to address such difficulties. Condition 3 spells our assumption about the error distribution and variances. The normality assumption is not necessary here. Any distribution with sub-Gaussian tails would be sufficient but the proof would be more tedious, though not necessarily harder. Condition 3' is the corresponding assumption about the error distribution under the null hypothesis that there is no variance change point.

We first show two technical lemmas that will be used in the proofs of our main theorems. Let $\delta_i = \widehat{f}^{(0)}(i/n) - E\{\widehat{f}^{(0)}\}(i/n)$ and $\delta_i^0 = E\{\widehat{f}^{(0)}\}(i/n) - f_0(i/n)$.

Lemma 2.5.1. *There exists constant c_m (depending only on m) s.t.*

$$\|E\{\widehat{f}^{(0)}\} - f_0\|_{\sup} \leq c_m \sqrt{J(f_0)}(h^{m-1/2} + (nh)^{-1/2}), \quad (2.5)$$

Lemma 2.5.2. *Suppose hypothesis H_1 holds true. Then it holds that*

$$\max_{1 \leq k_1 < k_2 \leq n} (k_2 - k_1)^{-1/2} \left| \sum_{i=k_1+1}^{k_2} [\epsilon_i^2 - E(\epsilon_i^2)] \right| = O_P(\log n), \quad (2.6)$$

$$\max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left| \sum_{i=1}^k \epsilon_i(\widehat{f}^{(0)}(i/n) - f_0(i/n)) \right| = O_P(n^{-1/4}h^{-3/4}), \quad (2.7)$$

$$\max_{1 \leq k \leq n} \frac{1}{\sqrt{n-k}} \left| \sum_{i=k+1}^n \epsilon_i(\widehat{f}^{(0)}(i/n) - f_0(i/n)) \right| = O_P(n^{-1/4}h^{-3/4}), \quad (2.8)$$

$$\max_{1 \leq k \leq n} |\widehat{f}^{(0)}(i/n) - f_0(i/n)| = O_P(\sqrt{\log n/(nh)} + h^{m-1/2}). \quad (2.9)$$

The above results (2.7), (2.6), (2.8) also hold true under hypothesis H_0 .

Proof of Lemma 2.5.1. Since $\widehat{f}^{(0)}$ is the minimizer of (3.3), we have

$$-\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{f}^{(0)}(i/n)) R_J(i/n, \cdot) + \lambda \widehat{f}^{(0)} = 0, \quad (2.10)$$

where R_J is the reproducing kernel associated with S^m (and J), and $R_J(x, \cdot)$ denotes the univariate function derived from R_J with its first argument fixed at x . Taking expectations, we get that

$$\frac{1}{n} \sum_{i=1}^n (\bar{f}(i/n) - f_0(i/n)) R_J(i/n, \cdot) + \lambda \bar{f} = 0, \quad (2.11)$$

where $\bar{f} = E\{\hat{f}^{(0)}\}$. Therefore, \bar{f} is the minimizer to the following functional

$$\ell_0(f) = \frac{1}{n} \sum_{i=1}^n (f(i/n) - f_0(i/n))^2 + \lambda J(f).$$

Since $\ell_0(\bar{f}) \leq \ell_0(f_0)$, we get

$$\frac{1}{n} \sum_{i=1}^n (\bar{f}(i/n) - f_0(i/n))^2 + \lambda J(\bar{f}) \leq \lambda J(f_0).$$

This means that $J(\bar{f}) \leq J(f_0)$. Let $g(t) = (\bar{f}(t) - f_0(t))^2$. Meanwhile, by Eggermont and LaRiccia (2009, Lemma (2.24), pp. 58) we get that

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n g(i/n) - \int_0^1 g(t) dt \right| &\leq \frac{1}{n} \int_0^1 |g'(t)| dt \\ &= \frac{2}{n} \int_0^1 |\bar{f}'(t) - f_0'(t)| \times |\bar{f}(t) - f_0(t)| dt \\ &\leq \frac{2}{n} \|\bar{f}' - f_0'\|_{L^2[0,1]} \|\bar{f} - f_0\|_{L^2[0,1]} \\ &\leq \frac{2}{n} \|\bar{f}^{(m)} - f_0^{(m)}\|_{L^2[0,1]}^2 \leq \frac{8}{n} J(f_0). \end{aligned} \quad (2.12)$$

In the meantime, (2.11) leads to

$$\frac{1}{n} \sum_{i=1}^n g(i/n) + \lambda J(\bar{f} - f_0) = -\lambda J(f_0, \bar{f} - f_0),$$

so that

$$\begin{aligned}
\|\bar{f} - f_0\|^2 &\equiv \|\bar{f} - f_0\|_{L^2}^2 + \lambda J(\bar{f} - f_0) \\
&= -\lambda J(f_0, \bar{f} - f_0) + \int_0^1 g(t)dt - \frac{1}{n} \sum_{i=1}^n g(i/n) \\
&\leq 2J(f_0)\lambda + 8J(f_0)/n = 2J(f_0)(\lambda + 4/n).
\end{aligned}$$

It follows from Eggermont and LaRiccia (2009) that

$$\|\bar{f} - f_0\|_{\sup} \leq c'_m h^{-1/2} \|\bar{f} - f_0\| \leq c_m \sqrt{J(f_0)} (h^{m-1/2} + (nh)^{-1/2}),$$

where c_m, c'_m are positive constants depending only on m . Thus (2.5) holds. \square

Proof of Lemma 2.5.2. Let $\sigma_i^2 = E(\epsilon_i^2)$. Without loss of generality, assume $\sigma^2 = \sigma_1^2 = \dots = \sigma_{k_0}^2 < \sigma_{k_0+1}^2 = \dots = \sigma_n^2 = \delta^2$. Since $\epsilon_i^2 - \sigma_i^2$ are independent centered sub-exponential random variables, by Vershynin (2012), there exist constants $c, d > 0$ such that, for any $1 \leq k_2 < k_2 \leq n$,

$$\begin{aligned}
&P\left(\left|\sum_{i=k_1+1}^{k_2} [\epsilon_i^2 - \sigma_i^2]\right| \geq C\sqrt{k_2 - k_1} \log n\right) \\
&\leq 2 \exp\left(-c \min\left\{C^2(k_2 - k_1)(\log n)^2/(d^2(k_2 - k_1)), C\sqrt{k_2 - k_1} \log n/d\right\}\right) \\
&\leq 2 \exp(-c \min\{C^2/d^2, C/d\} \log n) \leq 2 \exp(-3 \log n) = 2/n^3,
\end{aligned}$$

where $C = \max\{\sqrt{3d^2/c}, 3d/c\} > 0$. Hence, as $n \rightarrow \infty$,

$$P\left(\max_{1 \leq k_1 < k_2 \leq n} (k_2 - k_1)^{-1/2} \left|\sum_{i=k_1+1}^{k_2} [\epsilon_i^2 - \sigma_i^2]\right| \geq C \log n\right) \leq 2/n \rightarrow 0.$$

This shows (2.6).

Next we show (2.7). We only prove the results under H_1 . The results under H_0 can be

proved similarly. Define $\Omega = (\Omega_1^T, \dots, \Omega_n^T)^T$ with $\Omega_i = (R_J(1/n, i/n), \dots, R_J(n/n, i/n))/n$. Then by the representer theorem (Wahba, 1990) it can be shown that

$$(\widehat{f}^{(0)}(1/n), \dots, \widehat{f}^{(0)}(n/n))^T = \Omega(\Omega + \lambda I_n)^{-1} \mathbf{y}.$$

From Wahba (1990) we know that

$$R_J(x, y) = \sum_{\nu=1}^{\infty} \frac{\varphi_{\nu}(x)\varphi_{\nu}(y)}{\gamma_{\nu}} = 2 \sum_{k=1}^{\infty} \frac{\cos(2\pi k(x-y))}{(2\pi k)^{2m}}, \quad x, y \in \mathbb{I}.$$

For $0 \leq l \leq n-1$, let $c_l = 2/n \sum_{k=1}^{\infty} \cos(2\pi kl/n)/(2\pi k)^{2m}$. Since $c_l = c_{n-l}$ for $l = 1, 2, \dots, n-1$, Ω is symmetric circulant of order n .

Let $\zeta = \exp(2\pi\sqrt{-1}/n)$. The normalized eigenvectors of Ω can be specified as

$$\mathbf{x}_k = \frac{1}{\sqrt{n}}(1, \zeta^k, \zeta^{2k}, \dots, \zeta^{(n-1)k})^T, \quad k = 0, 1, \dots, n-1.$$

Let $M = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1})$. Denote M^* as the conjugate transpose of M . Clearly, $MM^* = I_n$ and Ω admits the decomposition $\Omega = M\Lambda M^*$, where $\Lambda = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$ with $\lambda_l = c_0 + c_1\zeta^l + \dots + c_{n-1}\zeta^{(n-1)l}$.

Direct calculations show that

$$\lambda_l = \begin{cases} 2 \sum_{k=1}^{\infty} \frac{1}{(2\pi kn)^{2m}}, & l = 0, \\ \sum_{k=1}^{\infty} \frac{1}{[2\pi(kn-l)]^{2m}} + \sum_{k=0}^{\infty} \frac{1}{[2\pi(kn+l)]^{2m}}, & 1 \leq l \leq n-1. \end{cases}$$

It is easy to examine that $\lambda_0 = 2\bar{c}_m(2\pi n)^{-2m}$ where $\bar{c}_m := \sum_{k=1}^{\infty} k^{-2m}$, and for $1 \leq l \leq$

$n - 1$,

$$\begin{aligned} \lambda_l &= \frac{1}{[2\pi(n-l)]^{2m}} + \frac{1}{(2\pi l)^{2m}} \\ &+ \sum_{k=2}^{\infty} \frac{1}{[2\pi(kn-l)]^{2m}} + \sum_{k=1}^{\infty} \frac{1}{[2\pi(kn+l)]^{2m}}. \end{aligned} \quad (2.13)$$

Let $\underline{c}_m = \sum_{k=2}^{\infty} k^{-2m}$. Then

$$\begin{aligned} \underline{c}_m (2\pi n)^{-2m} &\leq \sum_{k=2}^{\infty} \frac{1}{[2\pi(kn-l)]^{2m}} \leq \bar{c}_m (2\pi n)^{-2m}, \\ \underline{c}_m (2\pi n)^{-2m} &\leq \sum_{k=1}^{\infty} \frac{1}{[2\pi(kn+l)]^{2m}} \leq \bar{c}_m (2\pi n)^{-2m}. \end{aligned}$$

Let $\mathbf{e} \sim N(0, I_n)$ be a vector of independent standard normal random variables such that we can write $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T = D_0 \mathbf{e}$, where D_0 is the square-root of the true covariance matrix Σ_0 of $\boldsymbol{\epsilon}$, that is, D_0 is a diagonal matrix with the first τ_0 diagonals equal to σ_0 and the remaining diagonals equal to δ_0 . Let $\boldsymbol{\delta}_{(k)} = (\delta_1, \dots, \delta_k)^T$, $\Omega_{(k)} = (\Omega_1^T, \dots, \Omega_k^T)^T$, $\boldsymbol{\epsilon}_{(k)} = (\epsilon_1, \dots, \epsilon_k)^T$, $\boldsymbol{\epsilon}_{*(n-k)} = (\epsilon_{k+1}, \dots, \epsilon_n)^T$, $\mathbf{e}_{(k)} = (e_1, \dots, e_k)^T$, $\mathbf{e}_{*(n-k)} = (e_{k+1}, \dots, e_n)^T$. Then

$$\begin{aligned} \boldsymbol{\delta}_{(k)} &= \Omega_{(k)} (\Omega + \lambda I_n)^{-1} \boldsymbol{\epsilon} \\ &= (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T \Lambda M^* M (\Lambda + \lambda I_n)^{-1} M^* \boldsymbol{\epsilon} \\ &= (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T \Lambda (\Lambda + \lambda I_n)^{-1} M^* \boldsymbol{\epsilon} \\ &= (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T \Lambda (\Lambda + \lambda I_n)^{-1} (\bar{\mathbf{x}}_0, \dots, \bar{\mathbf{x}}_{k-1}) \boldsymbol{\epsilon}_{(k)} \\ &\quad + (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T \Lambda (\Lambda + \lambda I_n)^{-1} (\bar{\mathbf{x}}_k, \dots, \bar{\mathbf{x}}_{n-1}) \boldsymbol{\epsilon}_{*(n-k)}, \end{aligned}$$

where $\bar{\mathbf{x}}_k$ is the conjugate of \mathbf{x}_k .

Define, for $k \leq \tau_0$, $D_k = \text{diag}(\underbrace{\sigma_0, \dots, \sigma_0}_{k \text{ items}})$; for $k > \tau_0$, $D_k = \text{diag}(\underbrace{\sigma_0, \dots, \sigma_0}_{\tau_0 \text{ items}}, \underbrace{\delta, \dots, \delta}_{k - \tau_0 \text{ items}})$.

Define, for $k \leq n - \tau_0$, $D_{*k} = \text{diag}(\underbrace{\delta_0, \dots, \delta_0}_{k \text{ items}})$; for $k > n - \tau_0$, $D_{*k} = \text{diag}(\underbrace{\sigma_0, \dots, \sigma_0}_{k-n+\tau_0}, \underbrace{\delta_0, \dots, \delta_0}_{n-\tau_0})$.

It is easy to see that $\boldsymbol{\epsilon}_{(k)} = D_k \mathbf{e}_{(k)}$ and $\boldsymbol{\epsilon}_{*(k)} = D_{*k} \mathbf{e}_{*(k)}$. Let $\tilde{A}_k = (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T \Lambda (\Lambda + \lambda I_n)^{-1} (\bar{\mathbf{x}}_0, \dots, \bar{\mathbf{x}}_{k-1})$ and $\tilde{B}_k = (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T \Lambda (\Lambda + \lambda I_n)^{-1} (\bar{\mathbf{x}}_k, \dots, \bar{\mathbf{x}}_{n-1})$. Define $A_k = D_k \tilde{A}_k D_k$ and $B_k = D_k \tilde{B}_k D_{*n-k}$. Then

$$\boldsymbol{\epsilon}_{(k)}^T \boldsymbol{\delta}_{(k)} = \mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)} + \mathbf{e}_{(k)}^T B_k \mathbf{e}_{*(n-k)}.$$

By the Hanson-Wright inequality, for any $k = 1, \dots, n$,

$$P \left(|\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)} - E\{\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)}\}| \geq C_n \sqrt{k/(nh)} \right) \leq 2 \exp \left(- \min \left\{ \frac{c^2 C_n^2 k / (nh)}{\|A_k\|_F^2}, \frac{c C_n k / (nh)}{\|A_k\|_{op}} \right\} \right), \quad (2.14)$$

where $\|\cdot\|_F$ and $\|\cdot\|_{op}$ respectively denote the Frobenius norm and operator norm of a matrix, and $C_n > 0$ is a constant depending only n and $c > 0$ is a constant. Let $\Gamma = \Lambda (\Lambda + \lambda I_n)^{-1}$. Since Γ is a diagonal matrix, we can write $\Gamma = \text{diag}(\gamma_0, \dots, \gamma_{n-1})$. Let $M_{(k)} = (\mathbf{x}_0, \dots, \mathbf{x}_{k-1})^T$. Since $M^* M = I_n$, $M_{(k)}^* M_{(k)} \leq I_n$. Let $a_0 = \max\{\sigma^2, \delta^2\}$. We know that

$$\begin{aligned} \|A_k\|_F^2 &\leq a_0^2 \text{Tr}(\tilde{A}_k^* \tilde{A}_k) = a_0^2 \text{Tr}(M_{(k)} \Gamma M_{(k)}^* M_{(k)} \Gamma M_{(k)}^*) \\ &\leq a_0^2 \text{Tr}(M_{(k)} \Gamma^2 M_{(k)}^*) = a_0^2 \sum_{l=0}^{k-1} \mathbf{x}_l^T \text{diag}(\gamma_0^2, \dots, \gamma_{n-1}^2) \bar{\mathbf{x}}_l \\ &= \frac{a_0^2 k}{n} \sum_{r=0}^{n-1} \gamma_r^2 = O\left(\frac{k}{nh}\right), \text{ uniformly for } k. \end{aligned}$$

This also shows that $\|A_k\|_{op} \leq \|A_k\|_F = O(\sqrt{k/(nh)})$ uniformly for k . So for $C_n > 1$, (2.14) becomes

$$P \left(|\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)} - E\{\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)}\}| \geq C_n \sqrt{k/(nh)} \right) \leq 2 \exp(-c C_n).$$

This shows that

$$P \left(\max_{1 \leq k \leq n} \frac{|\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)} - E\{\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)}\}|}{\sqrt{k/(nh)}} \geq C_n \right) \leq 2n \exp(-c C_n).$$

Taking $C_n = (2/c) \log n$, we have shown that

$$\max_{1 \leq k \leq n} \frac{|\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)} - E\{\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)}\}|}{\sqrt{k/(nh)}} = O_P(\log n).$$

In the meantime,

$$P\left(\max_{1 \leq k \leq n} \frac{|\sum_{i=1}^k \delta_i^0 \epsilon_i|}{\sqrt{\sum_{i=1}^k (\delta_i^0 \sigma_i)^2}} \geq C\right) \leq nP(|Z| > C) = O(n \exp(-C^2/2)),$$

which implies

$$\max_{1 \leq k \leq n} \frac{|\sum_{i=1}^k \delta_i^0 \epsilon_i|}{\sqrt{\sum_{i=1}^k (\delta_i^0 \sigma_i)^2}} = O_P(\sqrt{\log n}).$$

So by Lemma 2.5.1 and conditions $h(\log n)^2 = o(1)$ and $nh^{4m+1}(\log n)^2 = o(1)$, we have

$$\begin{aligned} & \frac{1}{\sqrt{k}} \sum_{i=1}^k \epsilon_i (\delta_i + \delta_i^0) \\ &= \frac{1}{\sqrt{k}} \boldsymbol{\epsilon}_{(k)}^T \boldsymbol{\delta}_{(k)} + \frac{1}{\sqrt{k}} \sum_{i=1}^k \epsilon_i \delta_i^0 \\ &= \frac{1}{\sqrt{k}} (\mathbf{e}_{(k)}^T A_k \mathbf{e}_{(k)} + \mathbf{e}_{(k)}^T B_k \mathbf{e}_{*(n-k)}) + \frac{1}{\sqrt{k}} \sum_{i=1}^k \epsilon_i \delta_i^0 \\ &= O_P\left(\frac{\sqrt{k}}{nh} + \frac{\log n}{\sqrt{nh}} + \sqrt{\left(\frac{1}{nh} + \sqrt{\frac{1}{knh}}\right)\left(\frac{n-k}{nh} + \sqrt{\frac{n-k}{nh}}\right)} + \sqrt{\log n}(h^{m-1/2} + (nh)^{-1/2})\right) \\ &= O_P\left(\frac{1}{\sqrt{nh}} + \frac{\log n}{\sqrt{nh}} + \sqrt{\left(\frac{1}{nh} + \sqrt{\frac{1}{nh}}\right)\left(\frac{1}{h} + \sqrt{\frac{1}{h}}\right)} + \sqrt{\log n}(h^{m-1/2} + (nh)^{-1/2})\right) \\ &= O_P\left(\frac{1}{\sqrt{nh}} + \frac{\log n}{\sqrt{nh}} + \left(\frac{1}{nh^3}\right)^{1/4} + \sqrt{\log n}(h^{m-1/2} + (nh)^{-1/2})\right) \\ &= O_P(n^{-1/4}h^{-3/4}), \end{aligned}$$

where the O_P term is uniformly valid for $1 \leq k \leq n$.

Next we will handle $\|\widehat{f}^{(0)} - \bar{f}\|_{\sup}$. It can be seen by the representer theorem that $\widehat{f}^{(0)} - \bar{f} =$

$(R_J(1/n, \cdot), \dots, R_J(n/n, \cdot))(\Omega + \lambda I_n)^{-1} \epsilon/n$. It is easy to see that, with Ω_i being the i th row of Ω , $\delta_i = \Omega_i(\Omega + \lambda I_n)^{-1} \epsilon \sim N(0, \Omega_i(\Omega + \lambda I_n)^{-2} \Omega_i^T \sigma_i^2)$. Note that

$$\Omega_i(\Omega + \lambda I_n)^{-2} \Omega_i^T = \mathbf{x}_{i-1}^T \Lambda (\Lambda + \lambda I_n)^{-2} \Lambda \bar{\mathbf{x}}_{i-1} = \frac{1}{n} \sum_{r=0}^{n-1} \frac{\lambda_r^2}{(\lambda + \lambda_r)^2} \asymp \frac{1}{nh}.$$

Therefore, as $n \rightarrow \infty$,

$$P \left(\max_{1 \leq k \leq n} |\delta_i| \geq C \sqrt{\log n / \sqrt{nh}} \right) \leq nP(|\delta_i| \geq C \sqrt{\log n / \sqrt{nh}}) \leq n \exp(-C \log n) \rightarrow 0.$$

This shows that $\max_{1 \leq k \leq n} |\delta_i| = O_P(\sqrt{\log n / (nh)})$. The result follows from Lemma 2.5.1. \square

A.2 Proof of Theorem 2.2.1

The consistency result in Theorem 2.2.1 is proved in three steps: (1) the consistency of the initial mean function estimate, (2) the consistency of variance change point estimate and variance estimates given a consistent mean function estimate, and (3) the consistency of the mean estimate given consistent variance change point estimate and variance estimates. Particularly, we shall prove the following results.

1. $\max_{1 \leq i \leq n} |\hat{f}^{(0)}(i/n) - f_0(i/n)| = O_P(r_n)$.
2. Given that $\max_{1 \leq i \leq n} |\hat{f}^{(\iota-1)}(i/n) - f_0(i/n)| = O_P(r_n)$, we have $|\hat{\tau}^{(\iota)} - \tau_0| = O_P((\log n)^4 (\log \log n)^2)$, $[\hat{\sigma}^2]^{(\iota)} = \sigma_0^2 + O_P(\tilde{r}_n)$ and $[\hat{\delta}^2]^{(\iota)} = \delta_0^2 + O_P(\tilde{r}_n)$.
3. Given that $|\hat{\tau}^{(\iota)} - \tau_0| = O_P((\log n)^4 (\log \log n)^2)$, $[\hat{\sigma}^2]^{(\iota)} = \sigma_0^2 + O_P(\tilde{r}_n)$ and $[\hat{\delta}^2]^{(\iota)} = \delta_0^2 + O_P(\tilde{r}_n)$, we have $\max_{1 \leq i \leq n} |\hat{f}^{(\iota)}(i/n) - f_0(i/n)| = O_P(r_n)$.

These results, combined together, immediately guarantees the consistency result in Theorem 2.2.1. There is often a practical concern when combining iterative convergence results.

The results are all in-probability statements. After many iterations there could be a practical deterioration in the “high probability” that the results hold. However, our algorithm can usually converge in a few iterations. Therefore we feel comfortable that this issue is likely to be negligible in our case. For simplicity of notation, we shall drop the superscripts $(\iota - 1)$ and (ι) in this section of proof.

STEP 1. Consistency of the initial mean function estimate $\widehat{f}^{(0)}$.

This follows directly from Lemma 2.5.2.

STEP 2. Consistency of the estimates of τ_0 , σ_0^2 and δ_0^2 given a consistent mean estimate.

Without loss of generality assume $\sigma_0^2 < \delta_0^2$. The idea is to show that, $\ell(k) > \ell(\tau_0)$ uniformly for $k \in [cn/\log n, n - cn/\log n]$ with $|k - \tau_0| \geq (\log n)^4(\log \log n)^2$. We only consider $cn/\log n \leq k < \tau_0 - (\log n)^4(\log \log n)^2$ since the case for $n - cn/\log n \geq k > \tau_0 + (\log n)^4(\log \log n)^2$ is similar. Define $\eta_i = \widehat{f}(i/n) - f_0(i/n)$ and let

$$\widehat{\sigma}_0^2 = \frac{1}{\tau_0} \sum_{i=1}^{\tau_0} (\eta_i + \epsilon_i)^2, \quad \widehat{\sigma}_n^2 = \frac{1}{n - \tau_0} \sum_{i=\tau_0+1}^n (\eta_i + \epsilon_i)^2, \quad \widehat{\sigma}_k^2 = \frac{1}{\tau_0 - k} \sum_{i=k+1}^{\tau_0} (\eta_i + \epsilon_i)^2.$$

It follows by Lemmas 2.5.1 and 2.5.2 that $\frac{1}{\tau_0 - k} \sum_{i=k+1}^{\tau_0} \eta_i \epsilon_i = O_P(r_n \sqrt{\log n})$ uniformly for $k \leq \tau_0 - (\log n)^4(\log \log n)^2$. Hence we have

$$\begin{aligned} \widehat{\sigma}_k^2 &= \frac{1}{\tau_0 - k} \sum_{i=k+1}^{\tau_0} \eta_i^2 + \frac{2}{\tau_0 - k} \sum_{i=k+1}^{\tau_0} \eta_i \epsilon_i + \frac{1}{\tau_0 - k} \sum_{i=k+1}^{\tau_0} \epsilon_i^2 \\ &= \sigma^2 + O_P\left(r_n^2 + r_n \sqrt{\log n} + (\log n \log \log n)^{-1}\right) \\ &= \sigma^2 + O_P(r_{1n}^2), \end{aligned}$$

where $r_{1n}^2 = r_n^2 + r_n \sqrt{\log n} + (\log n \log \log n)^{-1}$ which is $o(1)$ by assumptions. Meanwhile, using similar argument we have $\widehat{\sigma}_0^2 = \sigma^2 + O_P(r_{1n}^2)$, $\widehat{\sigma}_n^2 = \delta^2 + O_P(r_{1n}^2)$.

Therefore, with probability approaching one, uniformly for $k \leq \tau_0 - (\log n)^4 (\log \log n)^2$,

$$|\hat{\sigma}_k^2 / \hat{\sigma}_n^2 - \sigma^2 / \delta^2| = O_P(r_{1n}^2), \quad |\hat{\sigma}_0^2 / \hat{\sigma}_n^2 - \sigma^2 / \delta^2| = O_P(r_{1n}^2),$$

$$\left| 1 - \frac{\hat{\sigma}_k^2}{\hat{\sigma}_0^2} \right| = \left| \frac{\hat{\sigma}_0^2 - \hat{\sigma}_k^2}{\hat{\sigma}_0^2} \right| = O_P(r_{1n}^2).$$

It is easy to see that

$$\frac{1}{k} \sum_{i=1}^k (y_i - \hat{f}(i/n))^2 = \frac{1}{k} \sum_{i=1}^k (\delta_i + \epsilon_i)^2 = \hat{\sigma}_0^2 + \frac{\tau_0 - k}{k} (\hat{\sigma}_0^2 - \hat{\sigma}_k^2),$$

$$\frac{1}{n-k} \sum_{i=k+1}^n (y_i - \hat{f}(i/n))^2 = \frac{1}{n-k} \sum_{i=k+1}^n (\delta_i + \epsilon_i)^2 = \hat{\sigma}_n^2 + \frac{\tau_0 - k}{n-k} (\hat{\sigma}_k^2 - \hat{\sigma}_n^2).$$

Therefore,

$$\begin{aligned} & \ell(k) \\ &= \frac{k}{n} \log \left(\hat{\sigma}_0^2 + \frac{\tau_0 - k}{k} (\hat{\sigma}_0^2 - \hat{\sigma}_k^2) \right) + \frac{n-k}{n} \log \left(\hat{\sigma}_n^2 + \frac{\tau_0 - k}{n-k} (\hat{\sigma}_k^2 - \hat{\sigma}_n^2) \right) \\ &= \ell(\tau_0) - \frac{\tau_0 - k}{n} \log \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_n^2} \right) + \frac{n-k}{n} \log \left(1 + \frac{\tau_0 - k}{n-k} \left(\frac{\hat{\sigma}_k^2}{\hat{\sigma}_n^2} - 1 \right) \right) + \frac{k}{n} \log \left(1 + \frac{\tau_0 - k}{k} \left(1 - \frac{\hat{\sigma}_k^2}{\hat{\sigma}_0^2} \right) \right). \end{aligned}$$

Note that

$$\begin{aligned} & \frac{n-k}{n} \log \left(1 + \frac{\tau_0 - k}{n-k} \left(\frac{\hat{\sigma}_k^2}{\hat{\sigma}_n^2} - 1 \right) \right) \\ &= \frac{n-k}{n} \log \left(1 + \frac{\tau_0 - k}{n-k} \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_n^2} - 1 + \frac{\hat{\sigma}_k^2 - \hat{\sigma}_0^2}{\hat{\sigma}_n^2} \right) \right) \\ &= \frac{n-k}{n} \log \left(1 + \frac{\tau_0 - k}{n-k} \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_n^2} - 1 \right) \right) + \frac{\tau_0 - k}{n} O_P(r_{1n}^2). \end{aligned}$$

Therefore, with probability approaching one, uniformly for $k < \tau_0 - (\log n)^4 (\log \log n)^2$, we

have

$$\begin{aligned}
& \ell(k) - \ell(\tau_0) \\
&= -\frac{\tau_0 - k}{n} \log \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2} \right) + \frac{n - k}{n} \log \left(1 + \frac{\tau_0 - k}{n - k} \left(\frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_n^2} - 1 \right) \right) + \frac{k}{n} \log \left(1 + \frac{\tau_0 - k}{k} \left(1 - \frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_0^2} \right) \right) \\
&= -\frac{\tau_0 - k}{n} \log \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2} \right) + \frac{n - k}{n} \log \left(1 + \frac{\tau_0 - k}{n - k} \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2} - 1 \right) \right) + \frac{\tau_0 - k}{n} O_P(r_{1n}^2) \\
&= \frac{\tau_0 - k}{n} \sum_{l=1}^{\infty} \frac{1}{l} \left(1 - \frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2} \right)^l \left[1 - \left(\frac{\tau_0 - k}{n - k} \right)^{l-1} \right] + \frac{\tau_0 - k}{n} O_P(r_{1n}^2) \\
&\geq \frac{\tau_0 - k}{2n} \left(1 - \frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2} \right)^2 \frac{n - \tau_0}{n - k} + \frac{\tau_0 - k}{n} O_P(r_{1n}^2) \\
&\geq \frac{\tau_0 - k}{2n} [(1 - \sigma^2/\delta^2)^2(1 - q_0) + O_P(r_{1n}^2)] > 0,
\end{aligned}$$

where the last inequality follows by $r_{1n}^2 = o(1)$. This means that $\widehat{\tau} \geq \tau_0 - (\log n)^4(\log \log n)^2$ with probability approaching one. Similarly, it can be shown that with probability approaching one, $\ell(k) - \ell(\tau_0) > 0$ uniformly for $k > \tau_0 + (\log n)^4(\log \log n)^2$, which implies $\widehat{\tau} \leq \tau_0 + (\log n)^4(\log \log n)^2$. Therefore,

$$|\widehat{\tau} - \tau_0| = O_P((\log n)^4(\log \log n)^2). \quad (2.15)$$

To show the consistency of $\widehat{\sigma}^2$ and $\widehat{\delta}^2$, note that by Lemma 2.5.2 we have that

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{\widehat{\tau}} \sum_{i=1}^{\widehat{\tau}} \eta_i^2 + \frac{2}{\widehat{\tau}} \sum_{i=1}^{\widehat{\tau}} \epsilon_i \eta_i + \frac{1}{\widehat{\tau}} \sum_{i=1}^{\widehat{\tau}} \epsilon_i^2 \\
&= O_P \left(\frac{\log n}{nh} + h^{2m-1} + (nh)^{-3/4} \right) + \frac{1}{\widehat{\tau}} \sum_{i=1}^{\widehat{\tau}} \epsilon_i^2 \\
&= O_P \left(r_n^2 + (nh)^{-3/4} \right) + \frac{1}{\widehat{\tau}} \left(\sum_{i=1}^{\widehat{\tau}} \epsilon_i^2 - \sum_{i=1}^{\tau_0} \epsilon_i^2 \right) + \frac{1}{\widehat{\tau}} \sum_{i=1}^{\tau_0} \epsilon_i^2.
\end{aligned}$$

By Lemma 2.5.2 and (2.15), we have

$$\begin{aligned}
\frac{1}{\widehat{\tau}} \left| \sum_{i=1}^{\widehat{\tau}} \epsilon_i^2 - \sum_{i=1}^{\tau_0} \epsilon_i^2 \right| &\leq \frac{1}{\widehat{\tau}} \sum_{i=\tau_0+1-|\widehat{\tau}-\tau_0|}^{\tau_0+1+|\widehat{\tau}-\tau_0|} \epsilon_i^2 \\
&= \frac{2|\widehat{\tau}-\tau_0|+1}{\widehat{\tau}} \times \frac{1}{2|\widehat{\tau}-\tau_0|+1} \sum_{i=\tau_0+1-|\widehat{\tau}-\tau_0|}^{\tau_0+1+|\widehat{\tau}-\tau_0|} \epsilon_i^2 \\
&= \frac{2|\widehat{\tau}-\tau_0|+1}{\widehat{\tau}} \times O_P \left(1 + \frac{\sqrt{|\widehat{\tau}-\tau_0|} \log n}{|\widehat{\tau}-\tau_0|+1} \right) \\
&= O_P((\log n)^5 (\log \log n)^2 / n), \\
\text{and } \frac{1}{\widehat{\tau}} \sum_{i=1}^{\tau_0} \epsilon_i^2 &= \frac{\tau_0}{\widehat{\tau}} \frac{1}{\tau_0} \sum_{i=1}^{\tau_0} \epsilon_i^2 = \sigma_0^2 + O_P(n^{-1/2}).
\end{aligned}$$

Therefore, we have proved that

$$\widehat{\sigma}^2 = \sigma_0^2 + O_P(r_n^2 + (nh)^{-3/4} + (\log n)^5 (\log \log n)^2 / n + n^{-1/2}) = \sigma_0^2 + O_P(\widetilde{r}_n).$$

The proof for $\widehat{\delta}^2 = \delta_0^2 + O_P(\widetilde{r}_n)$ is similar.

STEP 3. Consistency of the mean function estimate given the consistent estimates of τ_0 , σ_0^2 and δ_0^2 .

Recall from the proof of Lemma 2.5.2, we write $\boldsymbol{\epsilon} = D_0 \mathbf{e}$, where D_0 is the square-root of the true covariance matrix of $\boldsymbol{\epsilon}$ and \mathbf{e} is a vector of n independent standard normal random variables. By the representer theorem, the estimates $\widehat{f}^{(0)}$ and \widehat{f} have explicit expressions

$$\widehat{f}^{(0)} = \Omega(\Omega + \lambda I)^{-1} f_0 + \Omega(\Omega + \lambda I)^{-1} D_0 \mathbf{e},$$

$$\widehat{f} = \Omega(\Omega + \lambda_{new} \widehat{\Sigma})^{-1} f_0 + \Omega(\Omega + \lambda_{new} \widehat{\Sigma})^{-1} D_0 \mathbf{e}.$$

Without loss of generality, assume $\sigma^2 < \delta^2$. Let $\widehat{c} = \widehat{\delta}^2 / \widehat{\sigma}^2$ and $c_0 = \delta^2 / \sigma^2$. By consistency of $\widehat{\sigma}^2$, $\widehat{\delta}^2$, and $\widehat{\tau}$, with probability approaching one, $\mathcal{E}_n \equiv \{\widehat{\sigma}^2 < \widehat{\delta}^2, |\widehat{c} - c_0| \leq C \widetilde{r}_n\}$ holds, where

$\varepsilon > 0$ is arbitrarily small. Let 1_k be a vector of k 1's and 0_k be a vector of k 0's. It is easy to see that on \mathcal{E}_n ,

$$\begin{aligned}\lambda_{new}\widehat{\Sigma} &= \lambda \text{diag}(\widehat{c}1_{\widehat{\tau}}, 1_{n-\widehat{\tau}}) \\ &= \lambda \text{diag}(c_0 1_{\widehat{\tau}}, 1_{n-\widehat{\tau}}) + \lambda(\widehat{c} - c_0) \text{diag}(1_{\widehat{\tau}}, 0_{n-\widehat{\tau}}) \\ &\equiv \lambda \Gamma + \lambda(\widehat{c} - c_0)E,\end{aligned}$$

where $\Gamma = \text{diag}(c_0 1_{\widehat{\tau}}, 1_{n-\widehat{\tau}})$ and $E = \text{diag}(1_{\widehat{\tau}}, 0_{n-\widehat{\tau}})$. Then by the Sherman-Wooldbury formula,

$$\begin{aligned}(\Omega + \lambda_{new}\widehat{\Sigma})^{-1} - (\Omega + \lambda\Gamma)^{-1} \\ = -(\widehat{c} - c_0)\lambda(\Omega + \lambda\Gamma)^{-1}E(I + (\widehat{c} - c_0)\lambda E(\Omega + \lambda\Gamma)^{-1}E)^{-1}E(\Omega + \lambda\Gamma)^{-1} \equiv -(\widehat{c} - c_0)\Delta,\end{aligned}$$

where $\Delta = \lambda(\Omega + \lambda\Gamma)^{-1}E(I + (\widehat{c} - c_0)\lambda E(\Omega + \lambda\Gamma)^{-1}E)^{-1}E(\Omega + \lambda\Gamma)^{-1}$. Hence, on \mathcal{E}_n , $0 \leq \Delta \leq \frac{\lambda}{1-\varepsilon}(\Omega + \lambda\Gamma)^{-2} \leq \frac{1}{1-\varepsilon}(\Omega + \lambda I)^{-1}$.

Notice that $\lambda I \leq \lambda_{new}\widehat{\Sigma} \leq \lambda\widehat{c}I$. Hence, it holds that

$$\begin{aligned}(\widehat{f} - f_0)'(\widehat{f} - f_0) \\ \leq 2f_0'(\Omega + \lambda_{new}\widehat{\Sigma})^{-1}(\lambda_{new}\widehat{\Sigma})^2(\Omega + \lambda_{new}\widehat{\Sigma})^{-1}f_0 + 2\mathbf{e}^T D_0(\Omega + \lambda_{new}\widehat{\Sigma})^{-1}\Omega^2(\Omega + \lambda_{new}\widehat{\Sigma})^{-1}D_0\mathbf{e} \\ \leq 2\widehat{c}^2\lambda^2 f_0'(\Omega + \lambda I)^{-2}f_0 + 4\mathbf{e}^T D_0(\Omega + \lambda\Gamma)^{-1}\Omega^2(\Omega + \lambda\Gamma)^{-1}D_0\mathbf{e} + 4(\widehat{c} - c_0)^2\mathbf{e}^T D_0\Delta\Omega^2\Delta D_0\mathbf{e}.\end{aligned}$$

We will handle the three terms respectively. The first term is bounded by $2\widehat{c}^2(E\{\widehat{f}^{(0)}\} -$

$f_0)'(E\{\widehat{f}^{(0)}\} - f_0) = O_P(n\lambda)$ by Wahba (1990). To handle the second term, note that

$$\begin{aligned}
E\{\mathbf{e}^T D_0(\Omega + \lambda\Gamma)^{-1}\Omega^2(\Omega + \lambda\Gamma)^{-1}D_0\mathbf{e}\} &= \text{Tr}(D_0(\Omega + \lambda\Gamma)^{-1}\Omega^2(\Omega + \lambda\Gamma)^{-1}D_0) \\
&= \text{Tr}(\Omega(\Omega + \lambda\Gamma)^{-1}D_0^2(\Omega + \lambda\Gamma)^{-1}\Omega) \\
&\leq \delta^4 \text{Tr}(\Omega(\Omega + \lambda\Gamma)^{-2}\Omega) \\
&\leq \delta^4 \text{Tr}(\Omega(\Omega + \lambda I)^{-2}\Omega) = O(h^{-1}),
\end{aligned}$$

so the second term is $O_P(h^{-1})$. As for the third term, notice that

$$\begin{aligned}
\mathbf{e}^T D_0 \Delta \Omega^2 \Delta D_0 \mathbf{e} &\leq \mathbf{e}^T \mathbf{e} \text{Tr}(D_0 \Delta \Omega^2 \Delta D_0) \\
&\leq \delta^4 \mathbf{e}^T \mathbf{e} \text{Tr}(\Omega \Delta^2 \Omega) \\
&\leq \delta^4 \mathbf{e}^T \mathbf{e} \times \frac{1}{(1 - \varepsilon)^2} \text{Tr}(\Omega(\Omega + \lambda I)^{-2}\Omega) = O_P(nh^{-1}),
\end{aligned}$$

hence the last term is $O_P(nh^{-1}\widehat{r}_n^2)$. Therefore, $\|\widehat{f} - f_0\|_n^2 = O_P(\lambda + (nh)^{-1} + h^{-1}\widehat{r}_n^2)$.

A.3 Proof of Theorem 2.2.2

Under H_0 , the samples Y_i come from conventional nonparametric model with Gaussian errors of equal variance. Without loss of generality, assume that the variance of ϵ_i is one. Note that the mean function estimate under H_0 is $\widehat{f}^{(0)}$.

Recall that $\delta_i = \widehat{f}^{(0)}(i/n) - E\{\widehat{f}^{(0)}\}(i/n)$ and $\delta_i^0 = E\{\widehat{f}^{(0)}\}(i/n) - f_0(i/n)$. Let $\eta_i = \delta_i + \delta_i^0 = \widehat{f}^{(0)}(i/n) - f_0(i/n)$. For any $1 < k < n$, we have by Taylor's expansion and results

from Lemmas 2.5.1 and 2.5.2 that

$$\begin{aligned}
\ell(k) - \ell(n) &= n \log \left(1 + \frac{\sum_{i=1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=1}^n \eta_i \epsilon_i + \sum_{i=1}^n \eta_i^2}{n} \right) \\
&\quad - k \log \left(1 + \frac{\sum_{i=1}^k (\epsilon_i^2 - 1) + 2 \sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2}{k} \right) \\
&\quad - (n - k) \log \left(1 + \frac{\sum_{i=k+1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=k+1}^n \eta_i \epsilon_i + \sum_{i=k+1}^n \eta_i^2}{n - k} \right) \\
&= -\frac{1}{2n} \left(\sum_{i=1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=1}^n \eta_i \epsilon_i + \sum_{i=1}^n \eta_i^2 \right)^2 \\
&\quad + \frac{1}{2k} \left(\sum_{i=1}^k (\epsilon_i^2 - 1) + 2 \sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2 \right)^2 \\
&\quad + \frac{1}{2(n-k)} \left(\sum_{i=k+1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=k+1}^n \eta_i \epsilon_i + \sum_{i=k+1}^n \eta_i^2 \right)^2 \\
&\quad + O_P \left(n \left[\frac{\sum_{i=1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=1}^n \eta_i \epsilon_i + \sum_{i=1}^n \eta_i^2}{n} \right]^3 \right) \\
&\quad + O_P \left(k \left[\frac{\sum_{i=1}^k (\epsilon_i^2 - 1) + 2 \sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2}{k} \right]^3 \right) \\
&\quad + O_P \left((n - k) \left[\frac{\sum_{i=k+1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=k+1}^n \eta_i \epsilon_i + \sum_{i=k+1}^n \eta_i^2}{n - k} \right]^3 \right)
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{2n} \left[\sum_{i=1}^n (\epsilon_i^2 - 1) \right]^2 + \frac{1}{2k} \left[\sum_{i=1}^k (\epsilon_i^2 - 1) \right]^2 + \frac{1}{2(n-k)} \left[\sum_{i=k+1}^n (\epsilon_i^2 - 1) \right]^2 \\
&\quad - \frac{1}{n} \sum_{i=1}^n (\epsilon_i^2 - 1) \left[2 \sum_{i=1}^n \eta_i \epsilon_i + \sum_{i=1}^n \eta_i^2 \right] + \frac{1}{k} \sum_{i=1}^k (\epsilon_i^2 - 1) \left[2 \sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2 \right] \\
&\quad + \frac{1}{n-k} \sum_{i=k+1}^n (\epsilon_i^2 - 1) \left[2 \sum_{i=k+1}^n \eta_i \epsilon_i + \sum_{i=k+1}^n \eta_i^2 \right] \\
&\quad - \frac{1}{2n} \left[2 \sum_{i=1}^n \eta_i \epsilon_i + \sum_{i=1}^n \eta_i^2 \right]^2 + \frac{1}{2k} \left[2 \sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2 \right]^2 + \frac{1}{2(n-k)} \left[2 \sum_{i=k+1}^n \eta_i \epsilon_i + \sum_{i=k+1}^n \eta_i^2 \right]^2 \\
&\quad + O_P \left(n \left[\frac{\sum_{i=1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=1}^n \eta_i \epsilon_i + \sum_{i=1}^n \eta_i^2}{n} \right]^3 \right) \\
&\quad + O_P \left(k \left[\frac{\sum_{i=1}^k (\epsilon_i^2 - 1) + 2 \sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2}{k} \right]^3 \right) \\
&\quad + O_P \left((n-k) \left[\frac{\sum_{i=k+1}^n (\epsilon_i^2 - 1) + 2 \sum_{i=k+1}^n \eta_i \epsilon_i + \sum_{i=k+1}^n \eta_i^2}{n-k} \right]^3 \right) \\
&= -n \log \left(\frac{\sum_{i=1}^n \epsilon_i^2}{n} \right) + k \log \left(\frac{\sum_{i=1}^k \epsilon_i^2}{k} \right) + (n-k) \log \left(\frac{\sum_{i=k+1}^n \epsilon_i^2}{n-k} \right) \\
&\quad + O_P(\log n (n^{-1/4} h^{-3/4} + n^{-1/2} h^{-1} \log n + n^{1/2} h^{2m-1})) \\
&= \text{SIC}(k) - \text{SIC}(n) + O_P(r_n),
\end{aligned}$$

where the O_P term holds uniformly for k and $r_n = \log n (n^{-1/4} h^{-3/4} + n^{-1/2} h^{-1} \log n + n^{1/2} h^{2m-1})$. It then follows

$$\max_{1 \leq k \leq n} [\ell(k) - \ell(n)] = \max_{1 \leq k \leq n} [\text{SIC}(k) - \text{SIC}(n)] + O_P(r_n).$$

By Chen and Gupta (1997) we have for any $x \in \mathbb{R}$,

$$P(a_n (\log n)^{1/2} \lambda_{\star n} - b_n \log n \leq x) \rightarrow \exp(-2 \exp(-x)).$$

Since r_n satisfies $r_n \log^2 n = o(1)$, we have $a_n (\log n) (\lambda_n - \lambda_{\star n}) = o_P(1)$. Therefore,

$$P(a_n (\log n)^{1/2} \lambda_n - b_n \log n \leq x) \rightarrow \exp(-2 \exp(-x)).$$

Chapter 3 Variance Change Point Detection under a Smoothly-changing Mean Trend: The Multivariate Case

3.1 Introduction

According to the Organ Procurement and Transplantation Network, there are 119121 people needing a lifesaving organ transplant, 76385 of whom are active waiting list. Every ten minutes someone is added to the national transplant waiting list, and on average 22 people die each day while waiting for a transplant. The whole process, from obtaining a donated organ, finding a matching recipient, to the organ being transplanted to the recipient, can often take hours to days. On the other hand, the maximum viability of vital organs, even under intense procurement/preservation conditions, is 4-6 hours for lungs and hearts and 8-12 hours for livers and pancreas. The actual viability of a specific organ can vary from case to case, due to the variation in donor's health status and quality of procurement. Therefore, an accurate assessment of an organ's viability status is critical in the process. A conservative assessment can result in the wasteful disposal of a viable organ and an aggressive assessment can lead to a transplant failure and endanger the recipient's life.

Unfortunately, the current practice of organ viability assessment is mainly through two approaches: visual inspection and biopsy. The former approach undoubtedly suffers from the human subjectivity, especially so when the appearance of a viable organ is indistinguishable from a nonviable one. The latter is more accurate but often destroys the portion of organ where the biopsy sample is collected. A research team at Virginia Tech, including several physicians in the surrounding areas, are experimenting on a non-invasive way to monitoring

the viability status of an organ. We have designed a device (a non-invasive thermal imaging system) that can intensively measure the surface temperatures of an organ in machine perfusion. The temperatures are taken at a dense mesh of grid points covering the whole surface of the organ. As an example, the left and center panels in Figure 3.1 show the raw and detrended temperature profiles at a random spot on the liver surface. Our preliminary analysis on the data collected from experiments with porcine livers reveals a close relationship between the viability status and surface temperatures of the organ. Particularly, while following an overall smooth mean trend, the temperatures show a high variation when the corresponding part of the organ is viable and the variation drops sharply when the viability of that part of organ deteriorates. In other words, identifying this change point of the temperature variation essentially determines the viability breakdown point of the liver. Therefore, developing an analytic method for estimating such a change point is crucial for the progress of research on this non-invasive method of organ viability assessment.

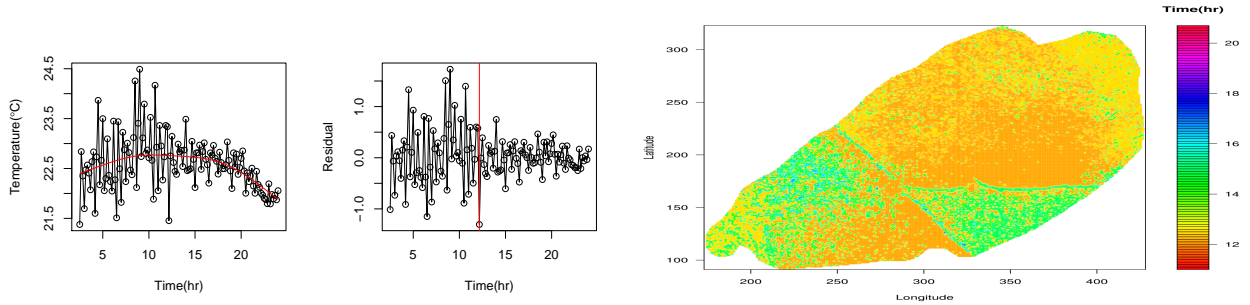


Figure 3.1: Sample of temperature profile and the heat map of estimated variance change points. Left and middle: The raw temperature profile and detrended temperature profile at a random spot of the liver. The x-axis labels in both panels represent 24 hours. Right: The heat map of estimated spot-wise variance change points of temperatures on a lobe of liver in the procurement experiment.

The existing change point detection methods mostly assume a *sudden* change in a few parameters of the data distribution or the distribution itself as a whole entity. However, the surface temperature data in our experiments have a smoothly changing mean function and thus a constantly-changing distribution. Blindly applying these methods would result in erroneous change point estimates. In a recent product of the project, Gao et al. (2018) focused on *spot-*

wise data analysis and developed a new variance change point detection method that allows the existence of a smoothly-changing mean function. They designed an iterative procedure to simultaneously estimate the mean function, variances, and variance change point. Their asymptotic theory included the asymptotic null distribution of the test statistic on variance change point as well as the convergence rates of all the parameter estimates. The final product from the application of the method was a heat map of change points on the surface of the liver as shown in the right panel of Figure 3.1. As we know, a liver is always transplanted as a whole or partially. In either case, the surgeon needs to know the viability of the whole organ or a chunk of it. Although the heat map provides some useful *spot-wise* viability information of the liver, it does not necessarily provide immediate information for the surgeon to make decision. An ideal statistical procedure should provide the viability status of a chunk or the whole organ. This would require extending the method of Gao et al. (2018) to estimating the change point of a variance function on a 2D surface in the presence of a smoothly-changing mean function on a 3D space of time and location.

However, such an extension is not trivial at all. Firstly, as revealed in Qiu et al. (2010), change point detection for functions is not an easy problem by itself. The complication of estimating a mean function makes it even harder. Fortunately, the work of Gao et al. (2018) offers us some methodological and theoretical guidance in this aspect. Secondly and more critically, the computational concern is a more severe issue here. The natural approach is to model the variance function by thin-plate splines on a 2D space and the mean function by the tensor product of thin-plate splines and cubic splines. The surface shown in Figure 3.1 consists of over 30,000 grid points. Even a small trunk of the organ can have thousands or tens of thousand locations. At each location there are over 100 time points. Therefore, the total sample size can easily reach the magnitude of millions. This is a daunting task for multivariate spline smoothing even equipped with modern computing power. An earlier work by Kim and Gu (2004) proposed to reduce the number of knots for smoothing splines from the sample size n to a number q of the order $n^{2/9}$. Consequently, comparing with the original smoothing

splines estimation, the computational cost can be reduced to $O(nq^2)$ rather than $O(n^3)$ and the function estimate can achieve the same convergence rate. This approach works well for data with sample sizes in hundreds or thousands. However, it does not offer a cure to the computational problem when the sample size goes beyond thousands, since the enormous data matrix would be too computationally challenging for all the matrix operations involved in the estimation procedure. Therefore, our goal is to provide an extension of the variance change point analysis method in Gao et al. (2018) to the multivariate scenario that is computationally feasible for data sets with large sample sizes and can provide a variance function change point estimate for a whole or a chunk of an organ in the surface temperature monitoring experiment.

The existing work on nonparametric smoothing with large data sets mostly uses the idea of divide-and-conquer (D&C). The D&C approach consists of three steps: divide the whole data into subsets, fits the model to each subset, and recombine estimates from all the subsets into an overall estimate. For example, Zhang et al. (2015) proposed a divide-and-conquer method for kernel ridge regression. Shang and Cheng (2017) studied the divide-and-conquer method for periodic smoothing splines and argued from a theoretical perspective that sample splitting may be viewed as an alternative form of regularization, playing a role similar to a smoothing parameter. Xu and Wang (2018) proposed two pairs of D&C approaches to fitting a cubic smoothing splines model on a large data set. One pair of approaches divide data in a simple random way, and the other in a sequential way by partitioning the domain into subintervals. We note that both Shang and Cheng (2017) and Xu and Wang (2018) only considered smoothing splines on a univariate domain. And the extension of D&C to modeling of our temperature monitoring experiment may have some other issues. On one hand, a simple random partition of the data does not take full advantage of the dense and regular sampling feature of our data. On the other hand, a sequential partition similar to Xu and Wang (2018) is much harder to implement for a 2D or 3D space. In addition, no matter how the data are partitioned, the ensuing recombining step can be cumbersome since we need to estimate a 3D mean function and two 2D variance functions as well as test on the existence of a variance

function change point.

Therefore, we adopt the subsampling approach in this work. The subsampling strategy has attracted a lot of attention in the past decade due to the modern data explosion and the urgent need of efficient statistical analysis methods to handle massive data sets. For example, Rokhlin and Tygert (2008), Drineas et al. (2011), Ma et al. (2015), Ma and Sun (2015), and Wang et al. (2018) designed various subsampling algorithms for linear regression with massive data. Fithian and Hastie (2014) and Wang et al. (2017) studied subsampling for logistic regression on massive data.

The main computational difficulty in extending Gao et al. (2018) to the multivariate case is the exponential increase in the sample size. The variance functions are now functions of 2D locations to be estimated by thin-plate splines while the mean function is a function of time and 2D location to be estimated by tensor product splines of cubic and thin-plate splines. The enormous sample size essentially makes such estimation infeasible on a standard computer server. To overcome such computational difficulty, we introduce a three-stage subsampling strategy to reduce the sample size for such spline estimation. The first stage is a block-wise subsampling of location, the second stage is a reduction in time points and the third stage a probability-weighted further subsampling of locations. The latter two stages take advantage of a preliminary spot-wise analysis on the stage one subsample using the method of Gao et al. (2018). After the subsampling, we then extend the estimation procedure in Gao et al. (2018) to the multivariate case and apply the extension to the subsample data. Our simulations demonstrate the computational efficiency of the proposed method while maintaining a decent estimation performance. The application of our method to the liver procurement experiment yield direct viability information about any selected chunk 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 due to its ability

to detect change point in 2D variance functions, (3) it provides a nontrivial extension of the existing subsampling approach to a rather complex problem setting.

The rest of the paper is organized as follows. We introduce our method in Section 3.2. In Section 3.3 we present all the simulations. We analyze the liver procurement data in Section 3.4. Discussion in Section 3.5 concludes the paper.

3.2 Method

3.2.1 Notation and Model

Suppose that surface temperature measurements y_{ij} are independent observations generated from the following model

$$y_{ij}^0 = f_0(i/n, \mathbf{z}_j^0) + \epsilon_{ij}, i = 1, \dots, N, j = 1, \dots, M \quad (3.1)$$

where M is the number of locations, N is the number of sampling time points, $\mathbf{z}_j^0 = (z_{1j}^0, z_{2j}^0) \in \mathbb{R}^2$ is the j th location (latitude and longitude), $f_0(t, \mathbf{z})$ is an unknown smooth mean function of time t and location \mathbf{z} , and the random errors $\epsilon_{ij} \sim N(0, \sigma_i^2(\mathbf{z}_j^0))$, where the variance function $\sigma_i^2(\mathbf{z}) = \sigma_0^2(\mathbf{z})$ when $i \leq \tau_0$ and $\sigma_i^2(\mathbf{z}) = \delta_0^2(\mathbf{z})$ when $i > \tau_0$. Both $\sigma_0^2(\mathbf{z})$ and $\delta_0^2(\mathbf{z})$ are unknown smooth positive functions of location \mathbf{z} . When $\sigma_0^2(\cdot) = \delta_0^2(\cdot)$, there is no variance change point; when $\sigma_0^2(\cdot) \neq \delta_0^2(\cdot)$, there is an unknown variance change point at τ_0 . The superscript ⁰ here is to distinguish the original data from the subsample data to be introduced later.

We first present the computational algorithm for our method and leave the details of each step to subsequent subsections. In the algorithm, the final estimates for the variance functions σ_0^2 and δ_0^2 uses thin-plate splines. All the estimates, intermediate and final, of the mean function f_0 uses tensor product splines of cubic and thin-plate splines.

3.2.2 Smoothing Splines Estimation

All the function parameters will be estimated by smoothing splines under the setting of non-parametric regression. Therefore we will give a short review of smoothing splines regression in this section. We start with a generic setting. Let (Y_i, x_i) be independent observations

Algorithm II

1. Three-stage subsampling (see Section 3.2.3):
 - (a) Stage I: Divide the region of interest into m_1 equal-sized small blocks and randomly select one location from each block to form the initial subsample Z_1 of m_1 locations.
 - (b) Stage II: Perform the spot-wise analysis of Gao et al. (2018) at the locations in Z_1 , calculate the average of all the spot-wise variance change points, and compute a time window $T = \{t_1, \dots, t_n\}$ of size n centered at the average change point.
 - (c) Stage III: Calculate a probability weight for each location in Z_1 based on statistics from the spot-wise analysis. Sample m locations from Z_1 with these probability weights to obtain the final location subsample $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$.
 2. Iterative estimation of mean function and variance change point on $T \times Z$.
 - (a) Initialize $\hat{f}^{(0)}$ with the mean function estimate assuming constant variance.
 - (b) Iterate on the following two steps until convergence. At the ι th iteration,
 - i. Given the mean estimate $\hat{f}^{(\iota-1)}$, we first use the testing procedure in Section 3.2.4 to find an estimate $\hat{\tau}^{(\iota)}$ for τ_0 . Then at each location \mathbf{z}_j in Z , we compute the maximum likelihood estimates $[\hat{\sigma}_j^2]^{(\iota)}$ and $[\hat{\delta}_j^2]^{(\iota)}$ using the respective subsequences of residuals, $\{y_{ij} - \hat{f}^{(\iota-1)}(t_i, \mathbf{z}_j) : i = 1, \dots, \hat{\tau}^{(\iota)}, \}$ and $\{y_{ij} - \hat{f}^{(\iota-1)}(t_i, \mathbf{z}_j) : i = \hat{\tau}^{(\iota)} + 1, \dots, n\}$.
 - ii. Now given the variance component estimates $\hat{\tau}^{(\iota)}$, $[\hat{\sigma}_j^2]^{(\iota)}$ and $[\hat{\delta}_j^2]^{(\iota)}$, we update the mean estimate by the minimizer of (3.2) where the covariance matrix Σ is updated with the current variance component estimates.
 3. Repeat Step 1 subsampling and Step 2 estimation K times. Denote the K sets of mean function and change point estimates as $\{(\hat{f}_1, \hat{\tau}_1), \dots, (\hat{f}_K, \hat{\tau}_K)\}$. Then our final mean and change point estimates are respectively $\hat{f} = \sum_{k=1}^K \hat{f}_k / K$ and $\hat{\tau} = \sum_{k=1}^K \hat{\tau}_k / K$.
 4. Use the whole data and the mean estimate \hat{f} to compute the residuals $\{\hat{\epsilon}_{ij} = y_{ij}^0 - \hat{f}(i/N, \mathbf{z}_j^0), i = 1, \dots, N; j = 1, \dots, M\}$. Use thin-plate splines to smooth the squared residuals $\{\hat{\epsilon}_{ij}^2, i = 1, \dots, \hat{\tau}; j = 1, \dots, M\}$ and $\{\hat{\epsilon}_{ij}^2, i = \hat{\tau} + 1, \dots, N; j = 1, \dots, M\}$ respectively to obtain the final variance function estimates $\hat{\sigma}^2(\mathbf{z})$ and $\hat{\delta}^2(\mathbf{z})$.
-

Table 3.1: Iterative Algorithm II

generated from the regression model $Y_i = \eta_0(x_i) + \epsilon_i$, $i = 1, \dots, I$, where Y_i is a continuous response variable, $x_i \in \mathcal{X}$ is a generic predictor that can be univariate or multivariate, η_0 is the unknown smooth function of interest, and the random errors $\epsilon_i \sim N(0, \sigma_i^2)$. Assume η_0 belongs to a reproducing kernel Hilbert space $\mathcal{H} = \{\eta | \eta : \mathcal{X} \rightarrow \mathbb{R}, J(\eta) < \infty\}$, where J is semi-norm on \mathcal{H} . Some examples of J and \mathcal{H} to be used in the paper will be given later. Let $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_I^2)$. Given Σ , the smoothing splines estimate of η is the minimizer of the penalized weighted least squares

$$\frac{1}{I}(\mathbf{Y} - \boldsymbol{\eta})^T \Sigma^{-1}(\mathbf{Y} - \boldsymbol{\eta}) + \lambda J(\eta), \quad (3.2)$$

in \mathcal{H} , where $\mathbf{Y} = (Y_1, \dots, Y_I)^T$, $\boldsymbol{\eta} = (\eta(x_1), \dots, \eta(x_I))^T$, $J(\eta)$ acts as a roughness penalty, and $\lambda > 0$ is the smoothing parameter balancing the trade-off between the goodness-of-fit and smoothness of the mean function estimate. When the random errors have a common variance, the matrix Σ^{-1} can be dropped from (3.2) since the common variance can be absorbed into λ . Then (3.2) reduces to the penalized least squares

$$\frac{1}{I}(\mathbf{Y} - \boldsymbol{\eta})^T (\mathbf{Y} - \boldsymbol{\eta}) + \lambda J(\eta), \quad (3.3)$$

A reproducing kernel Hilbert space (RKHS) is a Hilbert space \mathcal{H} where the evaluation functional $[x] : \mathcal{H} \rightarrow \mathbb{R}, \eta \mapsto \eta(x)$ is continuous for every $x \in \mathcal{X}$. By the Riesz Representation Theorem, for each $x \in \mathcal{X}$ there exists a unique function $R_x \in \mathcal{H}$ with the reproducing property $\langle R_x, \eta \rangle = [x](\eta) = \eta(x)$, 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 $R : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that $R(x, u) = \langle R_x, R_u \rangle$. One can show that each RKHS is uniquely associated with a reproducing kernel and vice versa. Note that the penalty functional J in (3.2) 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, e.g., Gu (2013, Chapter 2) for more details on RKHSs.

The RKHS \mathcal{H} is of infinite dimensions, so a direct optimization of (3.2) on \mathcal{H} seems infeasible. However, since the weighted least squares part in (3.2) depends on η only through its evaluations at the observation points $x_i, i = 1, \dots, I$, the Representer Theorem (Wahba, 1990) guarantees that the exact minimizer of (3.2) actually resides in a finite dimensional subspace of \mathcal{H} , namely, $\mathcal{N}_J \oplus \text{span}\{R_J(x_1, \cdot), \dots, R_J(x_I, \cdot)\}$. Let $\phi_l, l = 1, \dots, a$ be the basis functions of \mathcal{N}_J and $\xi_j = R_J(x_j, \cdot), j = 1, \dots, I$. Write $f = \phi^T \mathbf{d} + \boldsymbol{\xi}^T \mathbf{c}$, where \mathbf{c} and \mathbf{d} are the corresponding coefficient vectors. Also note that $J(\eta)$ can be written as a quadratic form $J(f) = \mathbf{c}^T Q \mathbf{c}$, where Q is the $I \times I$ matrix with the (i, j) th entry equal to $R_J(x_i, x_j)$. So for a fixed λ , the objective function (3.2) 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, Section 3.2).

We now introduce three examples of smoothing splines to illustrate these concepts. The cubic smoothing splines (Example 3.2.1) will be used for estimating mean functions of time only in the preliminary spot-wise analysis. The thin-plate splines (Example 3.2.2) will be used for computing the final estimates of the variance functions of location. The tensor product splines of cubic and thin-plate splines (Example 3.2.3) will be used for estimating the mean functions of time and location.

Example 3.2.1 (Cubic Smoothing Splines). *Suppose x is a variable of time and without loss of generality assume $\mathcal{X} = \mathcal{T} = [0, 1]$. 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, Section 2.3.3). \square*

Example 3.2.2 (Thin-plate Splines). *Please refer to Example 1.3.2 in Section 1.3.2.*

Example 3.2.3 (Tensor Product Splines). *Please refer to Section 1.3.3.*

3.2.3 Three-stage Subsampling

As discussed above, the data have over 30,000 densely-sampled locations and over 100 equally-spaced sampling time points. A completely random subsampling strategy does not take full advantage of the dense nature of the data in both space and time. Therefore, the first stage of our subsampling strategy uses a block-wise random sampling to select the locations. Also, we notice that analysis at a fixed location using Gao et al. (2018) is very fast. Therefore, a preliminary spot-wise analysis can provide key information about the variance change point and location-wise estimation accuracy, which can be exploited to guide the selection of time points in stage II and further subsampling of locations in stage III.

Stage I. As described in the algorithm, we first divide the region of interest into m_1 small blocks, all equal-sized rectangles except for the corner blocks. Then one location is randomly selected from each block to form the initial subsample Z_1 of m_1 locations. Note that a spot-wise preliminary analysis by the method of Gao et al. (2018) needs to be conducted at each of these m_1 locations. Therefore, the choice of m_1 needs to balance the size of the region and the computational cost of the preliminary analysis. In our numerical experiments, we have maintained m_1 to be in the range of 100 to 500.

Denote the data points corresponding to $Z_1 = \{\mathbf{z}_j^1, j = 1, \dots, m_1\}$ by $\{(y_{ij}^1, \mathbf{z}_j^1), i = 1, \dots, N; j = 1, \dots, m_1\}$. For each fixed location \mathbf{z}_j , we apply the method of Gao et al. (2018) to the measurements $\{y_{ij}^1, i = 1, \dots, m_1\}$ at the location and obtain the following statistics to assist the next subsampling stages: change point estimate $\hat{\tau}_j^1$ and mean square error of the mean estimate $\text{MSE}_j^1 = \sum_{i=1}^N \{y_{ij}^1 - \hat{f}_j^1(i/N)\}^2 / N$, where \hat{f}_j^1 is the mean function estimate at location \mathbf{z}_j^1 . Let $\bar{\tau}^1 = \sum_{j=1}^{m_1} \hat{\tau}_j^1 / m_1$ be the average change point estimate.

Stage II. With a preliminary change point estimate $\bar{\tau}^1$, we can now narrow down the time window to reduce the sample size in the time direction. Particularly, we compute a window $T = \{t_1, \dots, t_n\}$ of size n centered at $\bar{\tau}^1$. The choice of n should balance between the computational cost and the need of sufficient time points for proper change point analysis. We generally recommend an n no less than 50. In the paper we use $n = 60$.

Stage III. Although the first two stages have significantly reduced the sample size, it is still in the range of thousands or ten thousands. The computation would still be in the scale of hours or even blow up the memory of a standard computer server. Therefore, a further reduction of locations is desired. We shall apply a probability-weighted subsampling approach to Z_1 where the probability weight π_j for each location \mathbf{z}_j^1 is calculated as

$$\pi_j = \frac{\sum_{\ell=1}^{m_1} (\text{MSE}_\ell^1 + \varrho_1)}{\text{MSE}_j^1 + \varrho_1} \times \frac{\sum_{\ell=1}^{m_1} (|\hat{\tau}_\ell^1 - \bar{\tau}^1| + \varrho_2)}{|\hat{\tau}_j^1 - \bar{\tau}^1| + \varrho_2},$$

where ϱ_1 and ϱ_2 are small positive constants to prevent against zero values. Here we use $\varrho_1 = 0.0001$ and $\varrho_2 = 0.5$. The motivation for weight π_j is that those locations with “poor” estimates, represented by larger MSE for mean estimate and large deviation of change point estimate from the average change point, would be less likely to be selected. In our numerical studies we experiment with m values in the range of 10 to 50. Let $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$ be the final subsample of locations. The final subsample of data is now $\{y_{ij} : (t_i, \mathbf{z}_j) \in T \times Z, i = 1, \dots, n, j = 1, \dots, m\}$ which has a sample size of mn .

3.2.4 Iterative Estimation on $T \times Z$

The parameter estimation for the subsample data on $T \times Z$ iterates between: (1) estimate the mean function given estimates of the variance component parameters, and (2) estimate variance component parameters given estimate of the mean function. Note that in these iterations we only estimate the variance functions $\sigma_0^2(\mathbf{z})$ and $\delta_0^2(\mathbf{z})$ in a point-wise fashion,

leaving the estimation of them as smooth functions to the final steps of the whole algorithm after these iterations end.

For the first step, let the current estimates of variance component parameters be respectively $\tilde{\tau}, \tilde{\sigma}^2(\mathbf{z}_j)$ and $\tilde{\delta}^2(\mathbf{z}_j)$, $j = 1, \dots, m$. In the objective functional (3.2), the response vector \mathbf{Y} becomes $(y_{11}, y_{21}, \dots, y_{n1}, \dots, y_{1m}, \dots, y_{nm})^T$ and the covariance matrix Σ becomes a diagonal matrix whose diagonal repeats $\tilde{\tau}$ times with the entry $\tilde{\sigma}^2(\mathbf{z}_1)$, $(n - \tilde{\tau})$ times with the entry $\tilde{\delta}^2(\mathbf{z}_1)$, $\tilde{\tau}$ times with the entry $\tilde{\sigma}^2(\mathbf{z}_2)$, $(n - \tilde{\tau})$ times with the entry $\tilde{\delta}^2(\mathbf{z}_2)$, and so on.

$$\Sigma = \begin{pmatrix} \Sigma^{(\mathbf{z}_1)} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \dots & \vdots \\ 0 & \dots & \Sigma^{(\mathbf{z}_j)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \Sigma^{(\mathbf{z}_m)} \end{pmatrix},$$

and

$$\Sigma^{(\mathbf{z}_j)} = \begin{pmatrix} \tilde{\sigma}_{(1)}^2(\mathbf{z}_j) & & & & \\ & \ddots & & & 0 \\ & & \tilde{\sigma}_{(\tilde{\tau})}^2(\mathbf{z}_j) & & \\ & & & \tilde{\delta}_{(\tilde{\tau}+1)}^2(\mathbf{z}_j) & \\ & 0 & & & \ddots \\ & & & & & \tilde{\delta}_{(n)}^2(\mathbf{z}_j) \end{pmatrix},$$

where $\tilde{\sigma}_{(1)}^2(\mathbf{z}_j) = \dots = \tilde{\sigma}_{(\tilde{\tau})}^2(\mathbf{z}_j) = \tilde{\sigma}^2(\mathbf{z}_j)$ and $\tilde{\delta}_{(\tilde{\tau}+1)}^2(\mathbf{z}_j) = \dots = \tilde{\delta}_{(n)}^2(\mathbf{z}_j) = \tilde{\delta}^2(\mathbf{z}_j)$. The mean

function f_0 is estimated by the tensor produce splines in Example 3.2.3 as the minimizer of the penalized weighted least squares (3.2).

Given a mean function estimate \tilde{f} , we extend the testing procedure in Gao et al. (2018) to find an estimate $\tilde{\tau}$ for the variance change point τ_0 . Then we can compute the maximum likelihood estimates $\tilde{\sigma}^2(\mathbf{z}_j) = \tilde{\tau}^{-1} \sum_{i=1}^{\tilde{\tau}} \{y_{ij} - \tilde{f}(t_i, \mathbf{z}_j)\}^2$ and $\tilde{\delta}^2 = (n - \tilde{\tau})^{-1} \sum_{i=\tilde{\tau}+1}^n \{y_{ij} - \tilde{f}(t_i, \mathbf{z}_j)\}^2$. For a potential change point position τ , we want to test the hypothesis

$$H_0 : \sigma_1^2(\mathbf{z}) = \cdots = \sigma_n^2(\mathbf{z}) \text{ versus } H_1 : \sigma_1^2(\mathbf{z}) = \cdots = \sigma_\tau^2(\mathbf{z}) \neq \sigma_{\tau+1}^2(\mathbf{z}) = \cdots = \sigma_n^2(\mathbf{z}), \quad (3.4)$$

Let

$$\ell(\tau) = \sum_{j=1}^m \tau \log \left[\frac{1}{\tau} \sum_{i=1}^{\tau} \{y_{ij} - \tilde{f}(t_i, \mathbf{z}_j)\}^2 \right] + (n - \tau) \log \left[\frac{1}{n - \tau} \sum_{i=\tau+1}^n \{y_{ij} - \tilde{f}(t_i, \mathbf{z}_j)\}^2 \right].$$

Note that $\ell(n) = -2L_0(\tilde{\sigma}^2) - mn - mn \log 2\pi$ and $\ell(\tau) = -2L_1(\tilde{\sigma}^2, \tilde{\delta}^2) - mn - mn \log 2\pi$, where L_0 and L_1 are respectively the log likelihood functions under the null and alternative hypotheses of (3.4). So we define the test statistic to be $\Delta_n^2 = \max_{1 < \tau < n} \{\ell(n) - \ell(\tau)\}$. The asymptotic null distribution for Δ_n^2 , together with the convergence rates of all the parameter estimates, are given in Gao et al. (2018).

3.2.5 Final Estimates

Although our numerical experiments suggest one subsample can often deliver decent estimates with properly chosen m and n , sometimes one may want to repeat the subsample multiple times. Suppose the subsampling is repeated K times. Denote the K sets of mean function and change point estimates obtained from each subsample as $\{(\hat{f}_1, \hat{\tau}_1), \dots, (\hat{f}_K, \hat{\tau}_K)\}$. Then our final mean and change point estimates are respectively $\hat{f} = \sum_{k=1}^K \hat{f}_k / K$ and $\hat{\tau} = \sum_{k=1}^K \hat{\tau}_k / K$.

To get the final estimate of the variance functions, we first use the mean function estimate \hat{f} to compute the residuals $\{\hat{\epsilon}_{ij} = y_{ij}^0 - \hat{f}(i/N, \mathbf{z}_j^0), i = 1, \dots, N; j = 1, \dots, M\}$. Use thin-plate splines to smooth the squared residuals $\{\hat{\epsilon}_{ij}^2, i = 1, \dots, \hat{\tau}; j = 1, \dots, M\}$ and $\{\hat{\epsilon}_{ij}^2, i = \hat{\tau} + 1, \dots, N; j = 1, \dots, M\}$ respectively to obtain the final variance function estimates $\hat{\sigma}^2(\mathbf{z})$ and $\hat{\delta}^2(\mathbf{z})$.

3.3 Simulations

In this section we report simulations conducted to evaluate the empirical performance of the proposed method. We considered two different mean functions $f_{01}(t, \mathbf{z}) = 20 + 12t(1 - t) + \sin(2\pi z_1) + \cos(2\pi z_2)$ and $f_{02}(t, \mathbf{z}) = \sin(t) + t^5 - 8t^3 + 10t + 6 + \sin(2\pi z_1) + \cos(2\pi z_2)$. The first function f_{01} had a surface similar to the estimated mean temperature function of the liver in the procurement study and the second function f_{02} represented a more complex smooth surface. The true variance functions before and after the change point τ_0 were respectively $\sigma_0^2(\mathbf{z}) = 0.18 + \frac{2}{3}\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}$ and $\delta_0^2(\mathbf{z}) = 0.03 + \frac{5}{7}\{(z_1 - 0.5)^4 + (z_2 - 0.5)^4\}$. The number of time points was fixed to be $N = 130$, same as the number of time points in the application. The variance function change point was always $\tau_0 = N/2 = 65$. The locations were a 20×20 regular grid on the square domain $[0, 1] \times [0, 1]$ and thus $M = 400$. When applying the proposed method, the size of the reduced time window was always $n = 60$ and the number of subsampling repetitions was fixed to be $K = 3$. All the simulations were run on the *Windows Server 2012 R2 64bit with Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz (2 processors) and 64GB RAM*.

3.3.1 Computational time comparison

We first conducted a small scale experiment to compare the computational time of the proposed method for various location subsample sizes with that of the method without any subsampling. One can think of the latter as having $m = M$ and $n = N$. Five data replicates with true mean function f_{01} , $M = 400$ and $N = 130$ were generated. For the method with subsampling, we considered four different subsample sizes of locations, namely $m = 10, 20, 30$, and 50. For the method without subsampling, all the simulations with $M = 400$ aborted due to memory overflow. Therefore, for a better comparison we reduced M to 100 so that the simulation without subsampling could finish without memory issues. The computational time

for each setting was obtained as the average running times on the 5 replicates. The result is summarized in Table 3.2. As we can see, even with a much reduced M the method without subsampling must take more than an hour to finish, comparing with the few minutes of the method with subsampling. Since the viability duration of an organ is generally in hours, the subsampling approach is clearly much more appealing here.

Method	With subsampling				Without subsampling	
n (# time points)	60				130	
m (# locations)	10	20	30	50	100	400
Running time (seconds)	100.97	120.37	132.57	218.54	3921.08	NA

Table 3.2: Computational time comparison. NA indicates aborted simulations due to memory overflow.

3.3.2 Choice of location subsample size

In this part we conducted simulations to study the choice of the location subsample size m . We first looked at $m = 30$ for both mean functions f_{01} and f_{02} . Their corresponding mean and variance function plots are in Figures 3.2 to 3.5. We can see the mean function estimates clearly matched well with the true mean functions. Although the variance function estimates didn't match with the true variance functions as well as the mean function estimates, all of them were in similar range as the true variance functions. This would be adequate for our purpose of identifying the variance function change point.

We then reduced m to 10 and conducted simulation with $f_0 = f_{01}$. The estimate plots are in Figures 3.6 and 3.7. As expected, both the mean and variance function estimates

deteriorated. This suggests to us that $m = 10$ might be too small for a proper estimation.

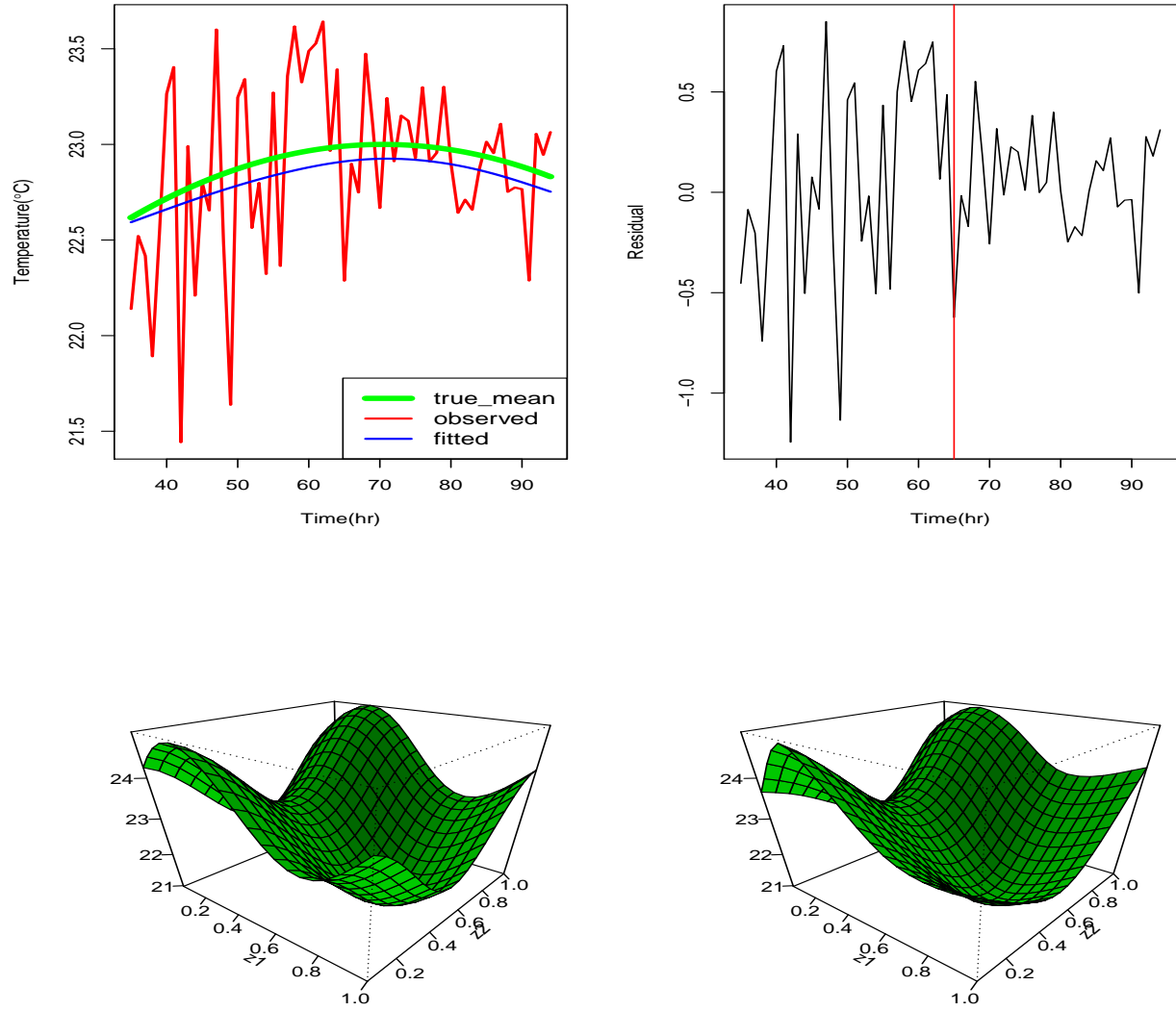


Figure 3.2: Mean function estimate plots with $f_0 = f_{01}$ and $m = 30$. Top left: observation (red), true mean (green), and estimated mean (blue) at a fixed location. Top right: de-trended data with estimated variance change point (red). Bottom panels: true mean surface (left) and estimated mean surface (right) at a fixed time point.

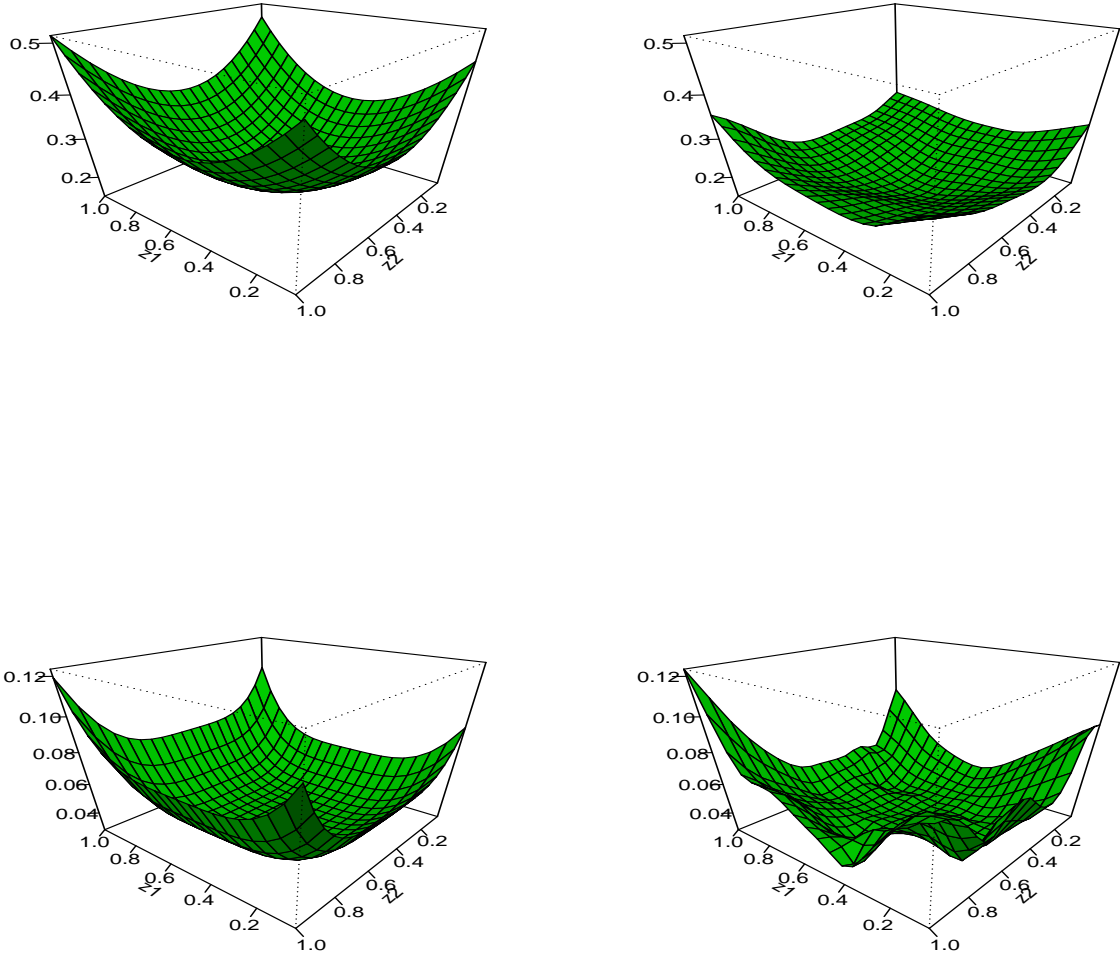


Figure 3.3: Variance function estimate plots with $f_0 = f_{01}$ and $m = 30$. Top panels: true function σ_0^2 (left) and its estimate (right). Bottom panels: true function δ^2 (left) and its estimate (right).

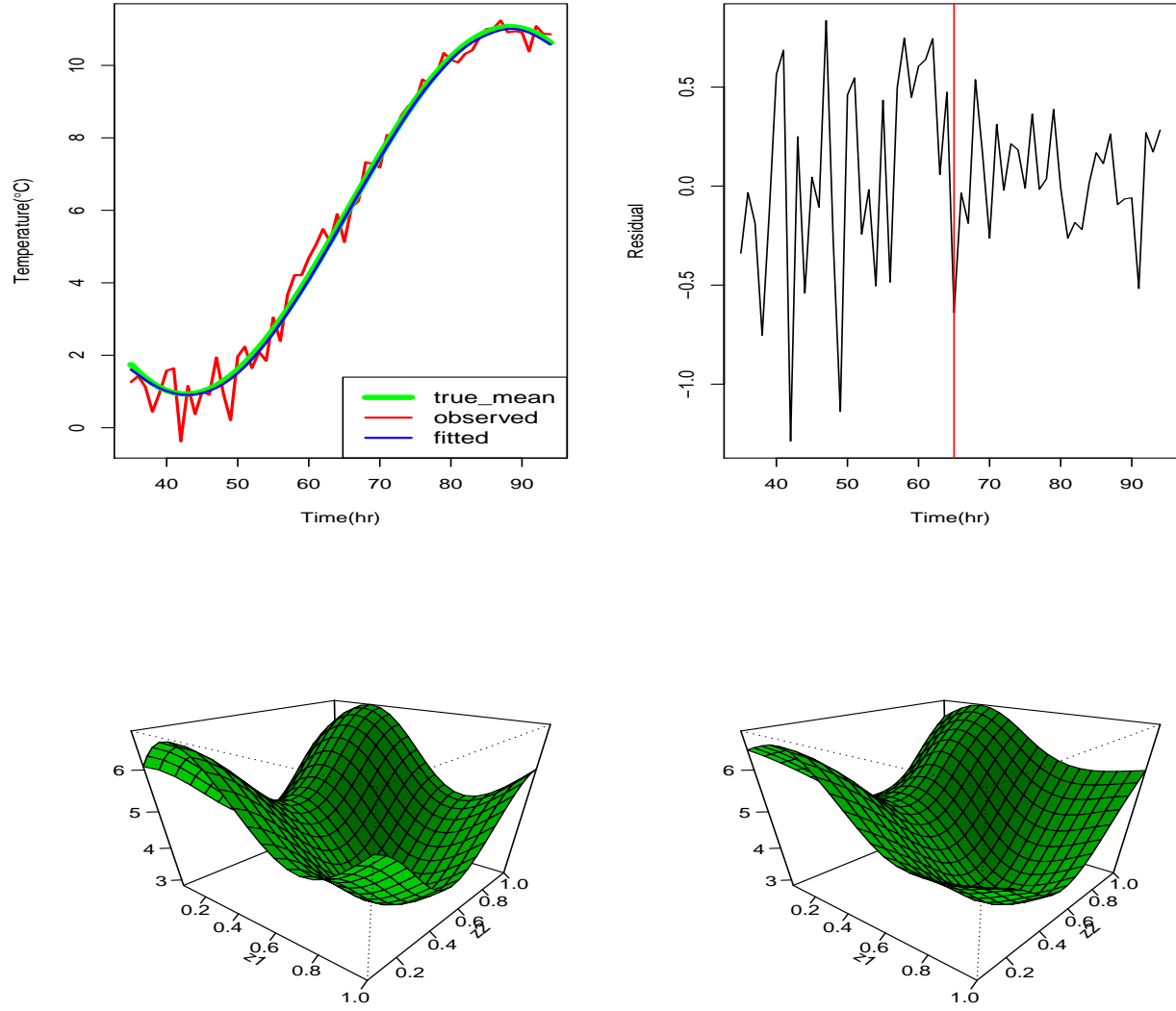


Figure 3.4: Mean function estimate plots with $f_0 = f_{02}$ and $m = 30$. Top left: observation (red), true mean (green), and estimated mean (blue) at a fixed location. Top right: de-trended data with estimated variance change point (red). Bottom panels: true mean surface (left) and estimated mean surface (right) at a fixed time point.

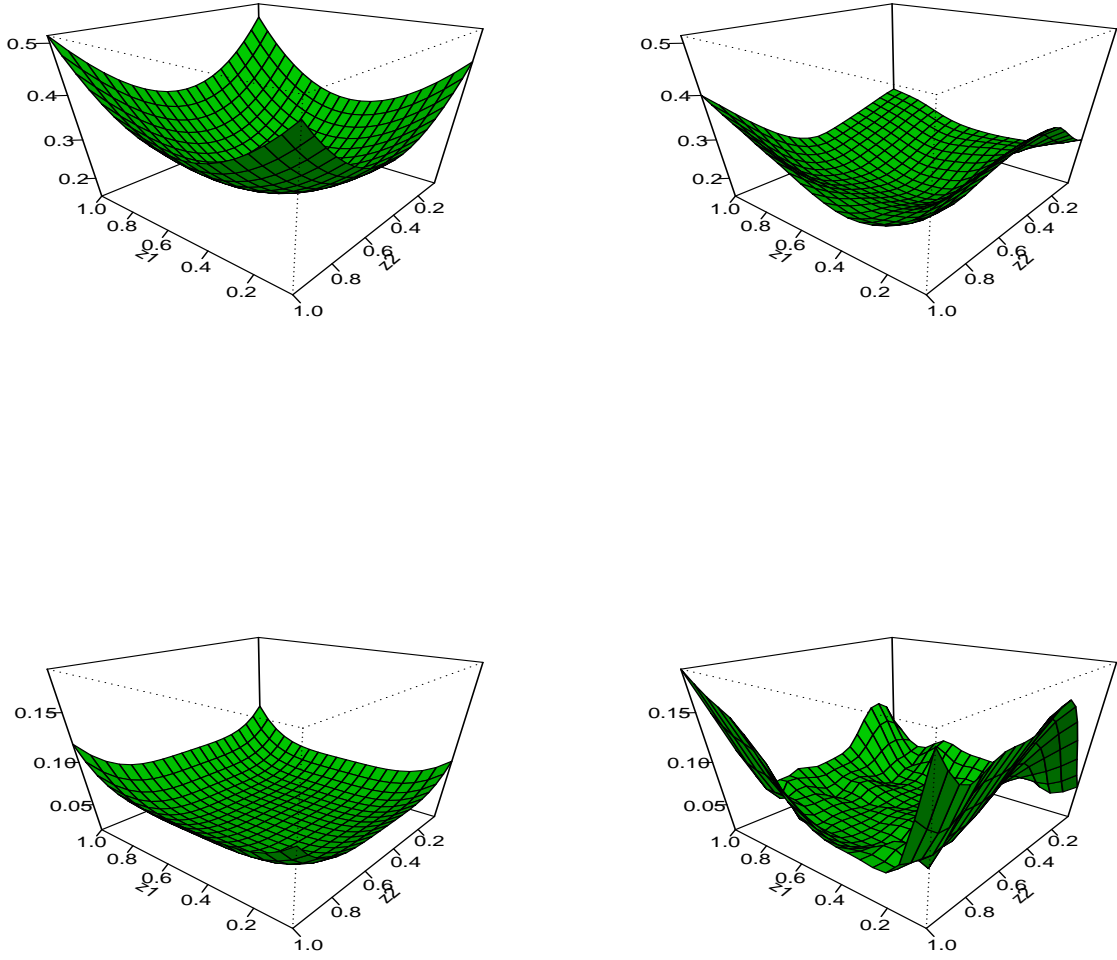


Figure 3.5: Variance function estimate plots with $f_0 = f_{02}$ and $m = 30$. Top panels: true function σ_0^2 (left) and its estimate (right). Bottom panels: true function δ^2 (left) and its estimate (right).

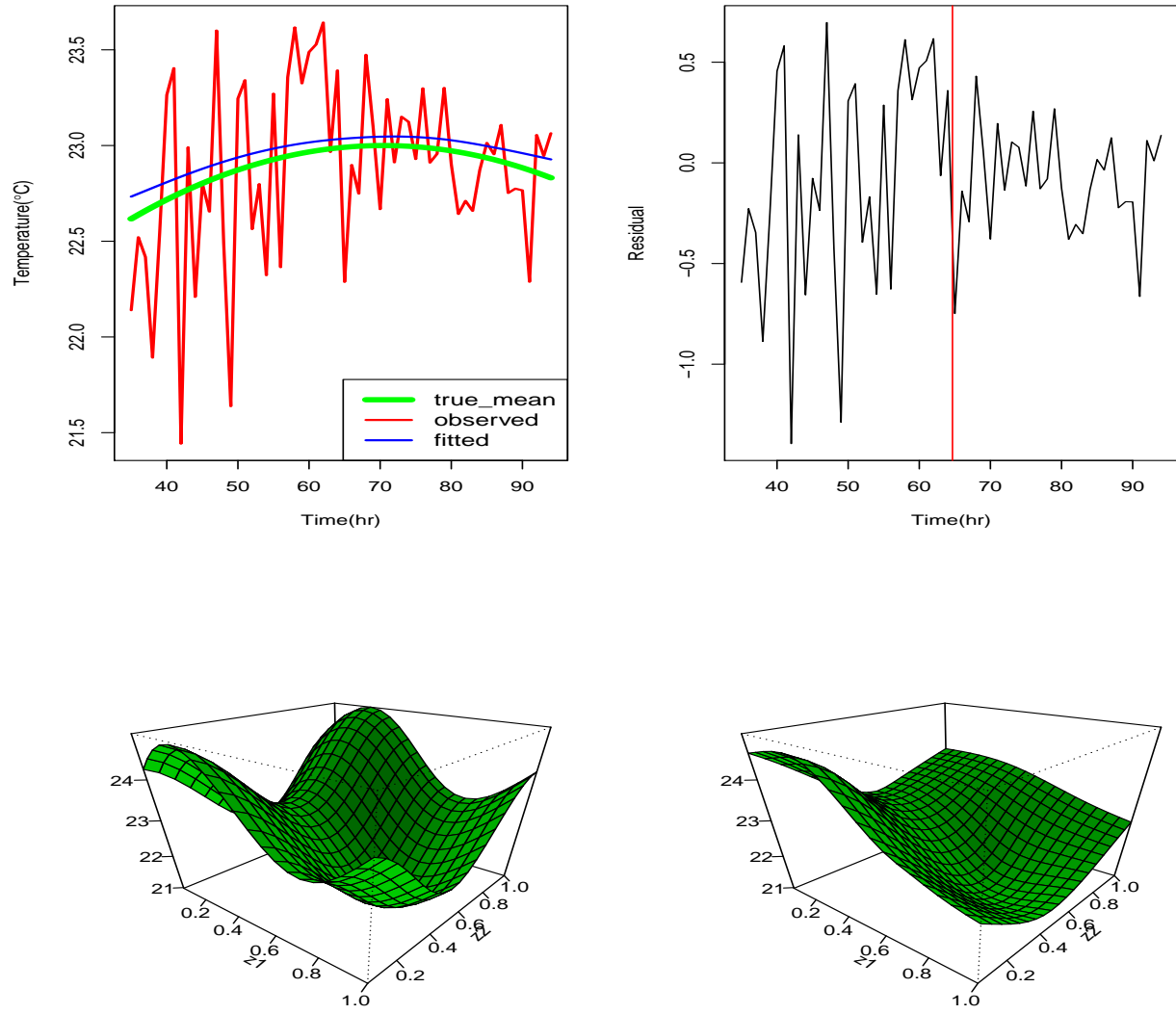


Figure 3.6: Mean function estimate plots with $f_0 = f_{01}$ and $m = 10$. Top left: observation (red), true mean (green), and estimated mean (blue) at a fixed location. Top right: de-trended data with estimated variance change point (red). Bottom panels: true mean surface (left) and estimated mean surface (right) at a fixed time point.

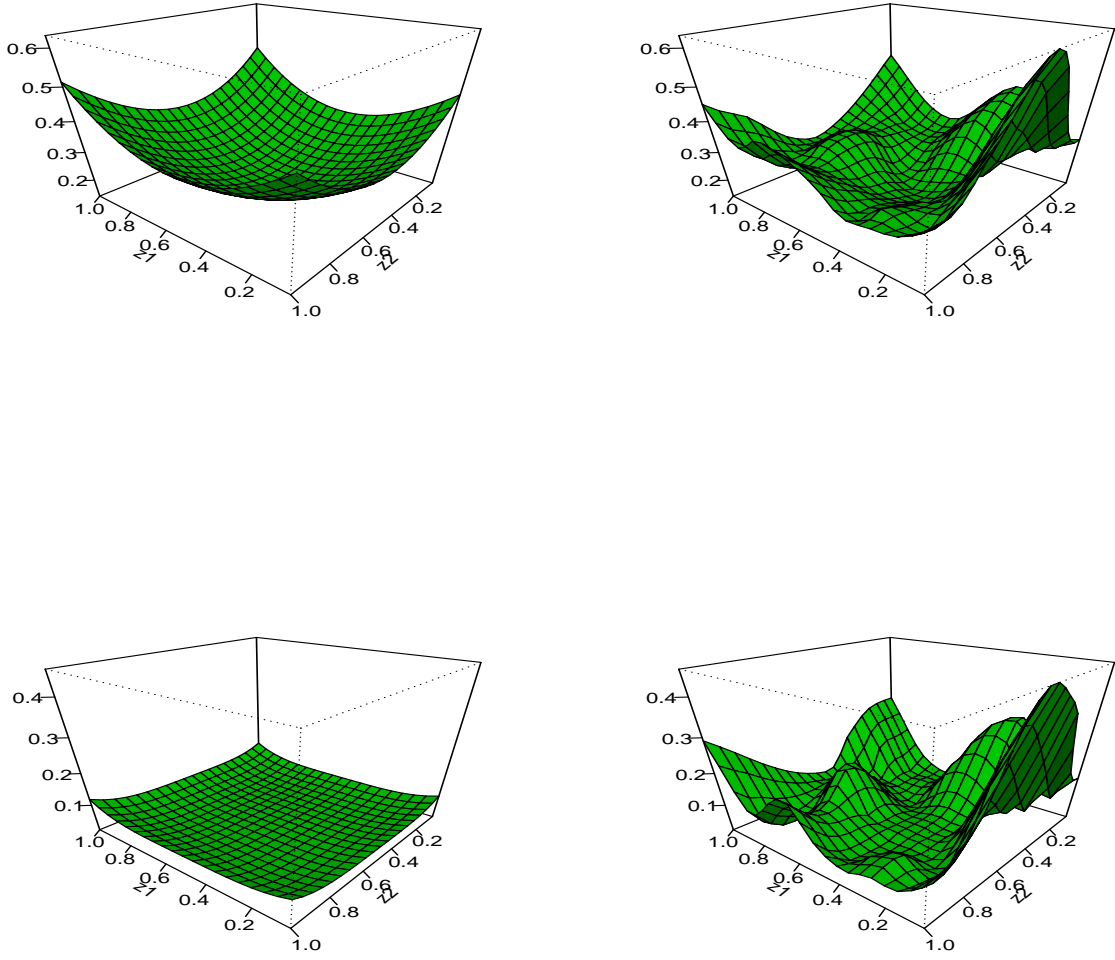


Figure 3.7: Variance function estimate plots with $f_0 = f_{01}$ and $m = 10$. Top panels: true function σ_0^2 (left) and its estimate (right). Bottom panels: true function δ^2 (left) and its estimate (right).

3.3.3 Estimation performance

We now increased the simulation scale based on the findings from the previous numerical experiments. We used $m = 30$ for the location subsample size and repeated the simulation with 1000 data replicates for both true mean functions f_{01} and f_{02} . The boxplots of change point estimates and MSEs of mean function estimates are respectively in Figures 3.8 and 3.9. We can see that our areal method can provide very accurate variance change point estimates with very small mean square errors of the mean function estimates.

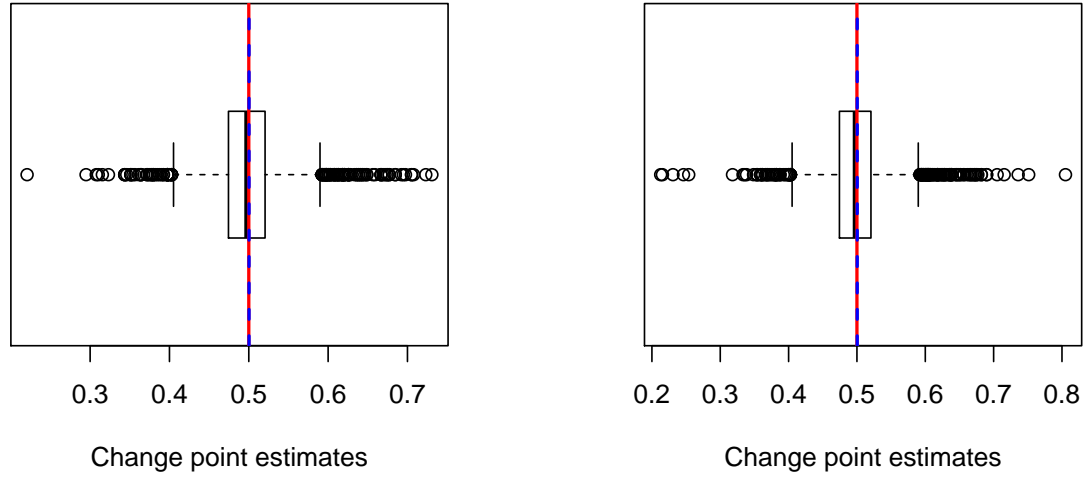


Figure 3.8: Boxplots of change point estimates. Left: Simulations with the true mean function $= f_{01}$; Right panels: Simulations with the true mean function $= f_{02}$. Red solid line is the true change point $\tau_0/N = 0.5$. Blue dash line is the mean of the estimated change point.

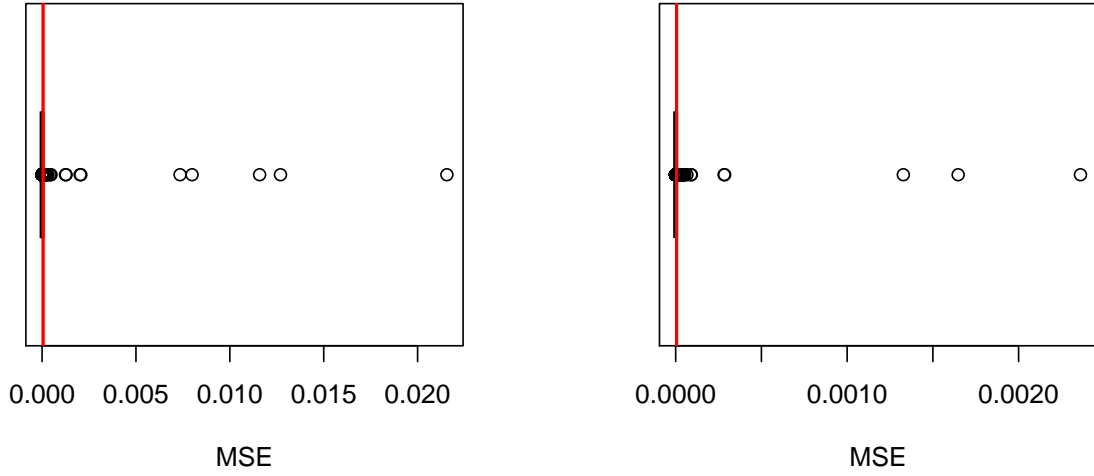


Figure 3.9: Boxplots of MSEs. Left panels: Simulations with the true mean function $= f_{01}$; Right panels: Simulations with the true mean function $= f_{02}$. Red line is the mean of the MSEs.

3.4 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 paper, a lobe of porcine liver, as shown in Figure 2.5, 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 selected five distinct areas (#1 to #5) from the liver surface, as shown in Figure 3.10. The numbers of locations in the areas were respectively 2383, 6177, 6563, 6480 and 13466 temperature profiles. We applied the proposed method with $m = 30$, $n = 60$, and $K = 3$ to each of these five pieces of liver. The computational times for the analysis of these areas were respectively 377.67s, 399.95s, 386.07s, 304.15s and 335.55s. The estimated variance change point $\hat{\tau}$ were 12.83, 12.1, 12.3, 13.13 and 11.95 hours for area #1 to #5 respectively. Figures 3.11 to 3.15 compared the estimated change point from the newly proposed method with the heatmaps and boxplots of the change point estimates from spot-wise analysis for areas #1 to #5. As we can see, these single change points can represent the variance change time of the corresponding chunk of liver very well. We selected these areas based on the uniformity of the estimated variance change points obtained from the spot-wise method. Area #2 is the most uniform area where all the spot-wise variance change points are similar to each other as showed in Figure 3.11. Areas #1 and #3 are fairly uniformly distributed areas. As we can see from Figures 3.11 to 3.13, the detected variance change points for these three areas using the areal detection method are earlier than the mean/median of the spot-wise detection results. The areal detection results are more likely to represent the viability status of the selected areas, because the simple summaries of the spot-wise analysis results provide poor references for the viability status of the corresponding organ piece. For example, the mean can be severely contaminated by a few noisy extreme values and the median indicates that half of the organ piece has already deteriorated. Therefore the areal detection results are more reliable for the purpose of the application. Also, from the statistical perspective, any

summary statistic based on spot-wise variance change points essentially treats these change points as sufficient statistics for the rather complicated experimental data. In contrast, the areal method incorporates much more comprehensive information from the data. Areas #4 and #5 in Figures 3.14 and 3.15 are the most heterogeneous areas on the liver, and even the areal method cannot be trusted due to the complexity of the viability status on these areas, especially for the area #5. Based on our studies, those areas are not recommended for transplant due to its unstable viability conditions.

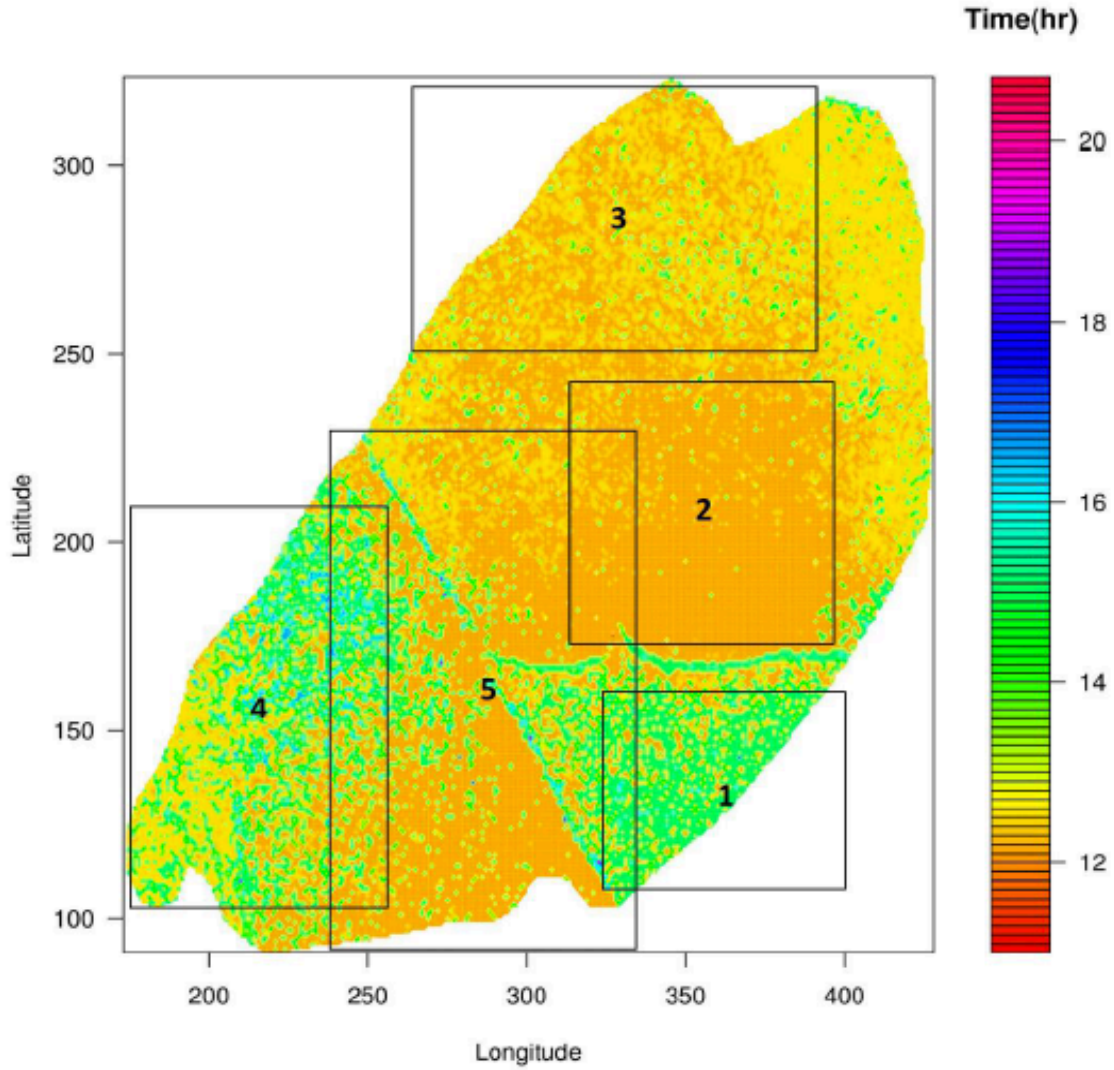


Figure 3.10: The heat map of estimated variance change points of temperatures on the organ surface in the liver procurement experiment with 5 selected areas. Area #1: latitude [109, 160], longitude [232, 400]; Area #2: latitude [175, 245], longitude [312, 398]; Area #3: latitude [251, 320], longitude [261, 394]; Area #4: latitude [104, 213], longitude [175, 255]; Area #5: latitude [91, 238], longitude [237, 355].

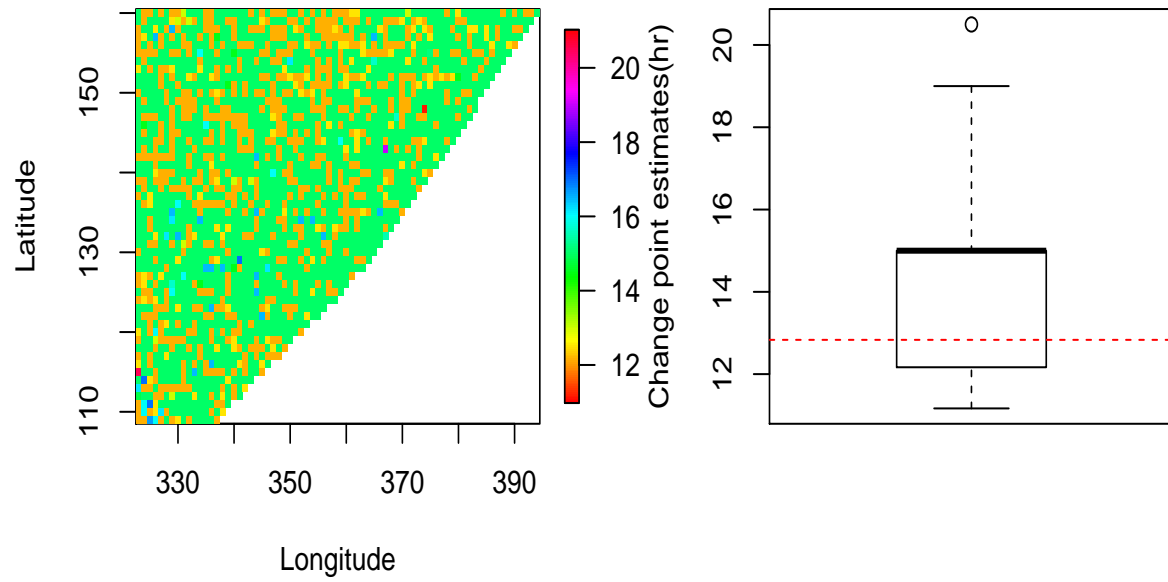


Figure 3.11: Variance change point estimates for area #1. Left: heat map of the estimated variance change points from spot-wise analysis. Right: boxplot of the estimated change points from spot-wise analysis for area #1. The estimated change point from applying the new method to area #1 was 12.83 hours.

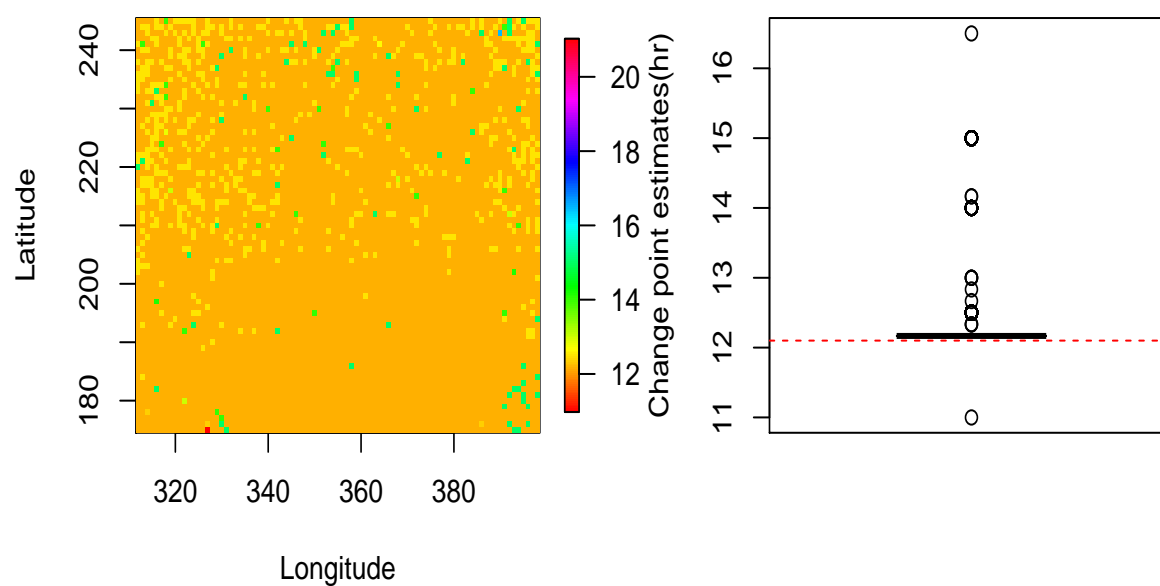


Figure 3.12: Variance change point estimates for area #2. Left: heat map of the estimated variance change points from spot-wise analysis. Right: boxplot of the estimated change points from spot-wise analysis for area #2. The estimated change point from applying the new method to area #2 was 12.1 hours.

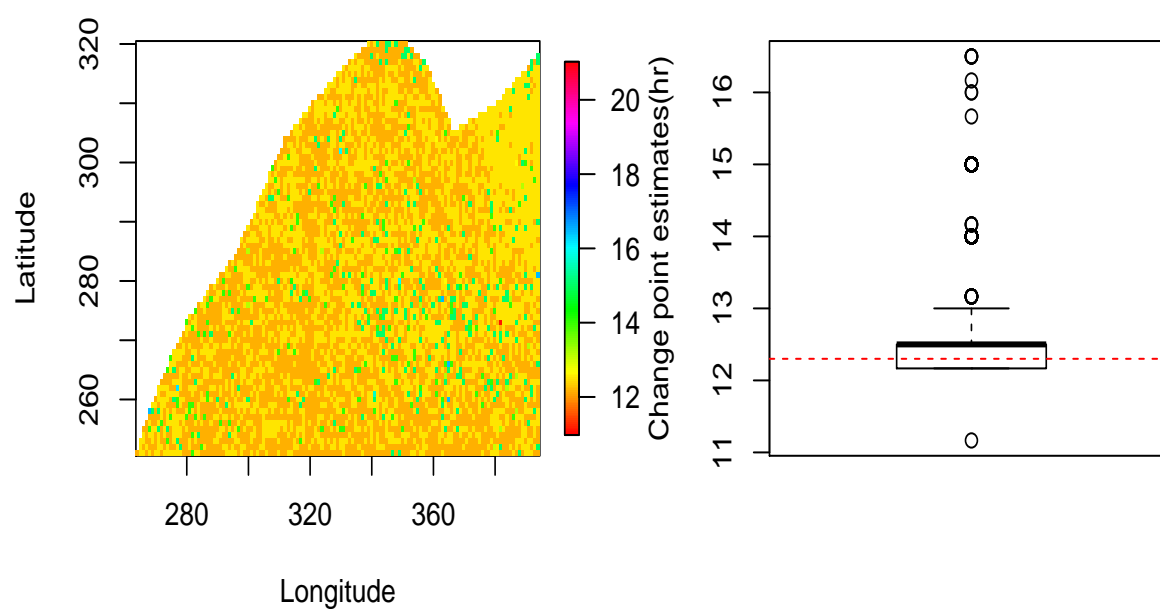


Figure 3.13: Variance change point estimates for area #3. Left: heat map of the estimated variance change points from spot-wise analysis. Right: boxplot of the estimated change points from spot-wise analysis for area #3. The estimated change point from applying the new method to area #3 was 12.3 hours.

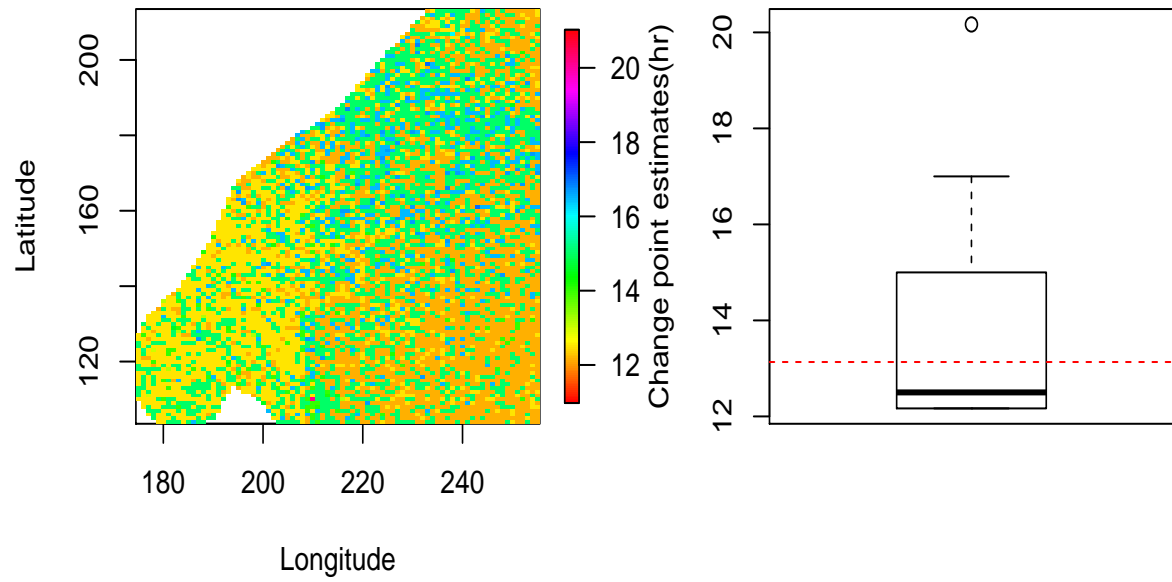


Figure 3.14: Variance change point estimates for area #4. Left: heat map of the estimated variance change points from spot-wise analysis. Right: boxplot of the estimated change points from spot-wise analysis for area #4. The estimated change point from applying the new method to area #4 was 13.13 hours.

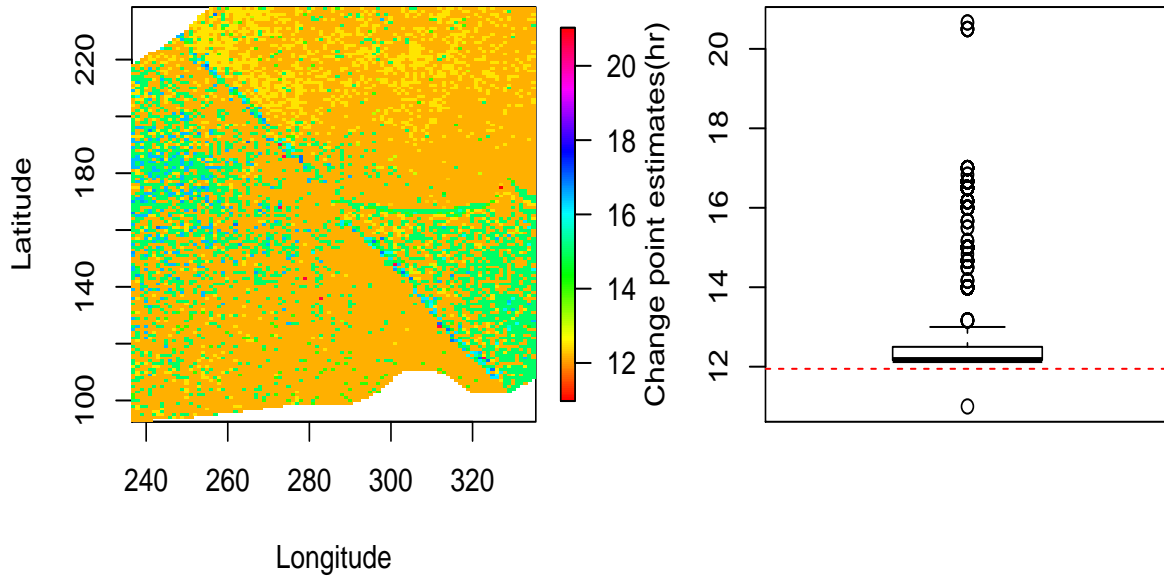


Figure 3.15: Variance change point estimates for area #5. Left: heat map of the estimated variance change points from spot-wise analysis. Right: boxplot of the estimated change points from spot-wise analysis for area #5. The estimated change point from applying the new method to area #5 was 11.95 hours.

3.5 Conclusion

In this chapter, we have expanded and applied our proposed variance change point detection method to multi-dimensional spaces with the location (latitude and longitude) and time involved. To overcome the computational difficulties of such an extension, we have introduced a very efficient three-stage subsampling strategy and dramatically reduced the computation time while producing decent estimates for all the unknown parameters and functions. The comparison simulations show that one can apply our method under different levels of subsampling for different research targets. For the purpose of estimating the variance change point only, we can choose smaller subsample size without any repeated subsampling, which can dramatically reduce the computation time. For the purpose of estimating the mean surface and variances structures, we can use larger subsample size with repeated subsampling. The application of our method to the liver procurement experiment gives accurate estimation of change points for chunks of the organ.

In this dissertation, we have provided two different variance change point detection approaches, which are the spot-wise method in Chapter 2 and the areal method in Chapter 3. Each of these methods has its pros and cons with respect to different research questions. The spot-wise method can provide a general viability structure of the whole organ surface, which can be used as a reference for selecting the target area for the later transplant, while the areal analysis can give you a distinct viability condition of the selected organ piece. We can combine these two methods together to provide more accurate guidance for viability assessment and organ transplant.

Bibliography

- Albert, P. S. and Hunsberger, S. A. (2004), “Identifying multiple changepoints in heterogeneous binary data with an application to molecular genetics,” *Biostatistics (Oxford)*, 5, 515–529.
- Barry, D. and Hartigan, J. A. (1992), “Product Partition Models for Change Point Problems,” *The Annals of Stat.*, 20, 260–279.
- Benjamini, Y. and Hochberg, Y. (1995), “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *J. Roy. Statist. Soc. Ser. B*, 57, 289–300.
- Benjamini, Y. and Yekutieli, D. (2001), “The control of the false discovery rate in multiple testing under dependency,” *The Annals of Stat.*, 29, 1165–1188.
- Chen, J. and Gupta, A. K. (1997), “Testing and locating variance changepoints with application to stock prices,” *J. Amer. Statist. Assoc.*, 92, 739–747.
- Chen, J. and Gupta, A. K. (2012), *Parametric statistical change point analysis (2nd Ed.)*, Basel; Cambridge, MA: Birkhäuser Verlag.
- Chen, K.-m., Cohen, A., and Sackrowitz, H. (2011), “Consistent multiple testing for change points,” *J. Multivariate Anal.*, 102, 1339–1343.
- Chernoff, H. and Zacks, S. (1964), “Estimating the current mean of a normal distribution which is subjected to changes in time,” *The Annals of Math. Stat.*, 35, 999–1018.
- Davis, W. W. (1979), “Robust methods for detection of shifts of the innovation variance of a time series,” *Technometrics*, 21, 313–320.
- Drineas, P., Mahoney, M. W., Muthukrishnan, S., and Sarlós, T. (2011), “Faster least squares approximation,” *Numerische mathematik*, 117, 219–249.

- Duchon, J. (1977), “Splines minimizing rotation-invariant semi-norms in Sobolev spaces,” *Constructive theory of functions of several variables*, 85–100.
- Eggermont, P. and LaRiccia, V. (2009), *Maximum penalized likelihood estimation*, vol. II, Springer.
- Farley, J. U., Hinich, M., and McGuire, T. W. (1975), “Some comparisons of tests for a shift in the slopes of a multivariate linear time series model,” *J. Econometrics*, 3, 297–318.
- Fithian, W. and Hastie, T. (2014), “Local case-control sampling: Efficient subsampling in imbalanced data sets,” *Annals of statistics*, 42, 1693.
- Gao, Z., Shang, Z., Du, P., and Robertson, J. L. (2018), “Variance change point detection under a smoothly-changing mean trend with application to liver procurement,” *J. Amer. Statist. Assoc.*, accepted.
- Gardner, L. A. (1969), “On detecting changes in the mean of normal variates,” *The Annals of Math. Stat.*, 40, 116–126.
- Good, I. J. and Gaskins, R. A. (1971), “Nonparametric roughness penalties for probability densities,” *Biometrika*, 58, 255–277.
- Grégoire, G. and Hamrouni, Z. (2002), “Change point estimation by local linear smoothing,” *J. Multivariate Anal.*, 83, 56–83.
- Gu, C. (1990), “Adaptive spline smoothing in non Gaussian regression models,” *J. Amer. Statist. Assoc.*, 85, 801–807.
- Gu, C. (1996), “Penalized likelihood hazard estimation: A general procedure,” *Statist. Sin.*, 6, 861–876.
- Gu, C. (2013), *Smoothing Spline ANOVA Models (2nd Ed.)*, New York: Springer-Verlag.
- Gu, C. and Qiu, C. (1993), “Smoothing spline density estimation: Theory,” *The Annals of Stat.*, 21, 217–234.

- Gu, C. and Xiang, D. (2001), “Cross-validating non-gaussian data: Generalized approximate cross-validation revisited,” *J. Comput. Graph. Statist.*, 10, 581–591.
- Halpern, A. L. (1999), “Minimally selected P and other tests for a single abrupt changepoint in a binary sequence,” *Biometrics*, 55, 1044–1050.
- Halpern, A. L. (2000), “Multiple-changepoint testing for an alternating segments model of a binary sequence,” *Biometrics*, 56, 903–908.
- Hariz, S. B., Wylie, J. J., and Zhang, Q. (2007), “Optimal rate of convergence for non-parametric change-point estimators for nonstationary sequences,” *The Annals of Stat.*, 35, 1802–1826.
- Hawkins, D. M. (1977), “Testing a sequence of observations for a shift in location,” *J. Amer. Statist. Assoc.*, 72, 180–186.
- Hinkley, D. V. (1970), “Inference About the Change-Point in a Sequence of Random Variables,” *Biometrika*, 57, 1–17.
- Hinkley, D. V. (1971), “Inference about the change-point from cumulative sum tests,” *Biometrika*, 58, 509–523.
- Horváth, L. (1993), “The maximum likelihood method for testing changes in the parameters of normal observations,” *The Annals of Stat.*, 21, 671–680.
- Horvath, L. (1993), “The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations,” *The Annals of Stat.*, 21, 671–680.
- Hsu, D. (1979), “Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis,” *J. Amer. Statist. Assoc.*, 74, 31–40.
- Hsu, D. A. (1977), “Tests for variance shift at an unknown time point,” *Applied Statistics*, 26, 279–284.

- Hušková, M. and Kirch, C. (2008), “Bootstrapping confidence intervals for the change-point of time series,” *J. Time Series Anal.*, 29, 947–972.
- Inclán, C. (1993), “Detection of multiple changes of variance using posterior odds,” *J. Bus. Econom. Statist.*, 11, 289–300.
- Inclán, C. and Tiao, G. C. (1994), “Use of cumulative sums of squares for retrospective detection of changes of variance,” *J. Amer. Statist. Assoc.*, 89, 913–923.
- Johnson, R. A. and Bhattacharyya, G. K. (1968), “Nonparametric Tests for Shift at an Unknown Time Point,” *The Annals of Math. Stat.*, 39, 1731–1743.
- Kim, Y.-J. and Gu, C. (2004), “Smoothing Spline Gaussian Regression: More Scalable Computation Via Efficient Approximation,” *J. Roy. Statist. Soc. Ser. B*, 66, 337–356.
- Kimeldorf, G. and Wahba, G. (1970a), “A correspondence between Bayesian estimation of stochastic processes and smoothing by splines,” *The Annals of Math. Stat.*, 41, 495–502.
- Kimeldorf, G. and Wahba, G. (1970b), “Spline functions and stochastic processes,” *Sankhya Ser. A*, 32, 173–180.
- Kimeldorf, G. and Wahba, G. (1971), “Some results on Tchebycheffian spline functions,” *J. Math. Anal. Applic.*, 33, 82–85.
- Krishnaiah, P. R., Miao, B., and Zhao, L. (1990), “Local likelihood method in the problems related to change point,” *Chinese Annals of Mathematics Series B*, 11, 363–375.
- Loader, C. R. (1996), “Change point estimation using nonparametric regression,” *The Annals of Stat.*, 24, 1667–1678.
- Loschi, R. and Cruz, F. (2002), “An analysis of the influence of some prior specifications in the identification of change points via product partition model,” *Comput. Statist. Data Anal.*, 39, 477–501.

- Loschi, R. and Cruz, F. (2005), “Extension to the product partition model: computing the probability of a change,” *Comput. Statist. Data Anal.*, 48, 255–268.
- Ma, P., Mahoney, M. W., and Yu, B. (2015), “A Statistical Perspective on Algorithmic Leveraging,” *Journal of Machine Learning Research*, 16, 861–911.
- Ma, P. and Sun, X. (2015), “Leveraging for big data regression,” *Wiley Interdisciplinary Reviews: Computational Statistics*, 7, 70–76.
- Matteson, D. S. and James, N. A. (2014a), “A nonparametric approach for multiple change point analysis of multivariate data,” *J. Amer. Statist. Assoc.*, 109, 334–345.
- Matteson, D. S. and James, N. A. (2014b), “A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data,” *J. Amer. Statist. Assoc.*, 109, 334–345.
- Meinguet, J. (1979), “Multivariate interpolation at arbitrary points made simple,” *Zeitschrift für angewandte Mathematik und Physik (ZAMP)*, 30, 292–304.
- Niu, Y. S. and Zhang, H. (2012), “The screening and ranking algorithm to detect DNA copy number variations,” 6, 1306–1326.
- Page, E. S. (1957), “On Problems in which a Change in a Parameter Occurs at an Unknown Point,” *Biometrika*, 44, 248–252.
- Pan, J. and Chen, J. (2006), “Application of modified information criterion to multiple change point problems,” *J. Multivariate Anal.*, 97, 2221 – 2241.
- Pettitt, A. (1979), “A non-parametric approach to the change-point problem,” *Applied Statistics*, 126–135.
- Qiu, P., Zou, C., and Wang, Z. (2010), “Nonparametric profile monitoring by mixed effects modeling (with discussion),” *Technometrics*, 52, 265–277.

- Rokhlin, V. and Tygert, M. (2008), “A fast randomized algorithm for overdetermined linear least-squares regression,” *Proceedings of the National Academy of Sciences*, 105, 13212–13217.
- Rozenholc, Y. (2001), “Nonparametric tests of change-points with tapered data,” *J. Time Series Anal.*, 22, 13–43.
- Schwarz, G. (1978), “Estimating the dimension of a model,” *The Annals of Stat.*, 6, 461–464.
- Sen, A. and Srivastava, M. S. (1975a), “On tests for detecting change in mean,” *The Annals of Stat.*, 3, 98–108.
- Sen, A. and Srivastava, M. S. (1975b), “Some one-sided tests for change in level,” *Technometrics*, 17, 61–64.
- Shang, Z. and Cheng, G. (2013), “Local and global asymptotic inference in smoothing spline models,” *The Annals of Stat.*, 41, 2608–2638.
- Shang, Z. and Cheng, G. (2017), “Computational limits of a distributed algorithm for smoothing spline,” *J. Mach. Learn. Res.*, 18, 1–37.
- Talwar, P. P. (1983), “Detecting a shift in location: Some robust tests,” *J. Econometrics*, 23, 353–367.
- Vershynin, R. (2012), “Introduction to the non-asymptotic analysis of random matrices,” in *Compressed Sensing: Theory and Applications*, eds. Eldar, Y. and Kutyniok, G., Cambridge: Cambridge University Press, chap. 5, pp. 210–268.
- Wahba, G. (1986), “Partial and interaction spline models for the semiparametric estimation of functions of several variables,” in *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*, pp. 75–80.
- Wahba, G. (1990), *Spline Models for Observational Data*, vol. 59 of CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia: SIAM.

- Wahba, G., Lin, Y., and Leng, C. (2001), “Penalized log likelihood density estimation, via smoothing spline ANOVA and ranGACV,” Tech. Rep. 1048, Department of Statistics, University of Wisconsin, Madison, WI.
- Wahba, G. and Wendelberger, J. (1980), “Some new mathematical methods for variational objective analysis using splines and cross validation,” *Monthly weather review*, 108, 1122–1143.
- Wang, H., Yang, M., and Stufken, J. (2018), “Information-based optimal subdata selection for big data linear regression,” *J. Amer. Statist. Assoc.*, forthcoming.
- Wang, H., Zhu, R., and Ma, P. (2017), “Optimal Subsampling for Large Sample Logistic Regression,” *Journal of the American Statistical Association*.
- Wichern, D. W., Miller, R. B., and Hsu, D.-A. (1976), “Changes of variance in first-order autoregressive time series models with an application,” *Applied Statistics*, 25, 248–256.
- Worsley, K. J. (1979), “On the likelihood ratio test for a shift in location of normal populations,” *J. Amer. Statist. Assoc.*, 74, 365–367.
- Xu, D. and Wang, Y. (2018), “Divide and Recombine Approaches for Fitting Smoothing Spline Models with Large Datasets,” *J. Comput. Graph. Statist.*, forthcoming.
- Zhang, N. R. and Siegmund, D. O. (2007), “A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data,” *Biometrics*, 63, 22–32.
- Zhang, Y., Duchi, J., and Wainwright, M. (2015), “Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates,” *J. Mach. Learn. Res.*, 16, 3299–3340.