Statistica Sinica Preprint No: SS-2025-0113	
Title	Multiple Testing of Local Extrema for Detection of Structural Breaks in Piecewise Linear Models
Manuscript ID	SS-2025-0113
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0113
Complete List of Authors	Zhibing He, Dan Cheng and Yunpeng Zhao
Corresponding Authors	Dan Cheng
E-mails	chengdan@asu.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Multiple Testing of Local Extrema for Detection of Structural Breaks in Piecewise Linear Models

Zhibing He, Dan Cheng and Yunpeng Zhao

Southeast University; Arizona State University; and Colorado State University

Abstract: In this paper, we propose a new generic method for detecting the number and locations of change points in piecewise linear models. Our method transforms the change point detection problem into identifying local extrema through kernel smoothing and differentiation of the data sequence. By computing p-values for all local extrema based on the derived peak height distributions of the derivatives of smooth Gaussian processes, we utilize the Benjamini-Hochberg procedure to identify significant local extrema as the detected change points. Our method effectively distinguishes between two types of change points: continuous breaks (Type I) and jumps (Type II). We study three scenarios of piecewise linear signals: pure Type I, pure Type II and a mixture of both. The results demonstrate that our proposed method ensures asymptotic control of the False Discovery Rate (FDR) and power consistency as the sequence length, slope changes, and jump size increase. Furthermore, compared to traditional change point detection methods based on recursive segmentation, our approach requires only one instance of multiple-testing across all candidate local extrema, thereby achieving the smallest computational complexity proportional to the data sequence length. Additionally, numerical studies illustrate that our method maintains FDR control and power consistency, even in non-asymptotic situations with moderate

Corresponding author: Dan Cheng. Email: chengdan@asu.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

slope changes or jumps. We have implemented our method in the R package "dSTEM".

Key words and phrases: structural breaks, change points, piecewise linear models, kernel smoothing, multiple testing, Gaussian processes, peak height distribution, and FDR.

1. Introduction

In this paper, we consider a canonical univariate statistical model:

$$y(t) = \mu(t) + z(t), \quad t \in \mathbb{R}, \tag{1.1}$$

where $z(t)$ is random noise and $\mu(t)$ is a piecewise linear signal of the form $\mu(t) = c_j + k_j t, \quad t \in (v_{j-1}, v_j]$, where $c_j, k_j \in \mathbb{R}$, $j = 1, 2, \ldots$ and $-\infty = v_0 < v_1 < v_2 < \cdots$. Assume that the structures of $\mu(t)$ are different in the neighboring v_j, i.e., $(c_j, k_j) \neq (c_{j+1}, k_{j+1})$, resulting in a continuous break or jump at v_j (see Figure 1). This v_j is called a change point or structural break.

The detection of structural breaks or change points plays a pivotal role in various fields, including statistics, econometrics, genomics, climatology, and medical imaging. These detection methods are broadly applied to different domains based on different types of signals. For example, in monitoring medical conditions (Khan et al. 2020), the presence of piecewise constant signals with jumps can indicate significant changes in patient health conditions. Accurate detection of these changes is essential for aiding in diagnosis, treatment planning, and object recognition tasks. On the other hand, piecewise linear signals

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

with continuous breaks are commonly encountered in climate change research (Tebaldi & Lobell 2008). These changes may represent gradual shifts in temperature patterns or human behavior. Furthermore, in financial markets, piecewise linear signals with non-continuous breaks may indicate sudden shifts in stock prices or market dynamics (Chang et al. 2008). Detecting these changes is crucial for making informed investment decisions.

To differentiate between the various ways of linear structural changes, we define the following two types of change points.

Definition 1. *A point v_j is called a Type* I *change point (structural break) if $c_j + k_j v_j = c_{j+1} + k_{j+1} v_j$ and $k_j \neq k_{j+1}$, and a point v_j is denoted as a Type* II *change point if $c_j + k_j v_j \neq c_{j+1} + k_{j+1} v_j$ for $j \geq 1$.*

At Type I change points v_j, the signals remain continuous, but the slopes experience a sudden change at v_j. Type II change points correspond to jumps in the signal, where there is a discontinuity between adjacent segments. Note that a special case of Type II change points is that $\mu(t)$ is piecewise constant, i.e., $k_j \equiv 0$ and $c_j \neq c_{j+1}$ for $j \geq 1$. In this paper, we specifically focus on three scenarios of the signal $\mu(t)$, each representing different characteristics of the change points. The scenarios are described as follows:

Scenario 1: Pure Type I Change Points. The signal $\mu(t)$ consists of continuous segments with slope changes at specific points v_j. Figure 1 (a) illustrates this

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

scenario, where the signal exhibits continuous behavior with distinct slopes at the change points.

Scenario 2: Pure Type II Change Points. The signal $\mu(t)$ contains only Type II change points (jumps). Figure 1 (d) illustrates this scenario, where the signal exhibits sudden changes or discontinuities at the change points.

Scenario 3: Mixture of Type I and Type II Change Points. The signal $\mu(t)$ combines both continuous breaks (Type I) and jumps (Type II). Figure 1 (g) illustrates this scenario, which allows a more general representation of signals with mixed characteristics.

Our primary objective is to detect both the number of change points and their locations simultaneously. To address this problem, we propose a novel approach that can handle an unknown number of multiple structural breaks occurring at unknown positions within the signal. Our proposed method for change point detection is illustrated and evaluated in the three aforementioned scenarios (pure Type I change points, pure Type II change points, and a mixture of Type I and Type II change points). Figure 1 shows the application of our method to each scenario, showcasing the mechanism behind our change point detection method.

The key idea behind our approach is based on the observation that a change point in the signal $\mu(t)$ will manifest as a local extremum (local maximum or local minimum) in either the first or second derivative of the smoothed signal.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

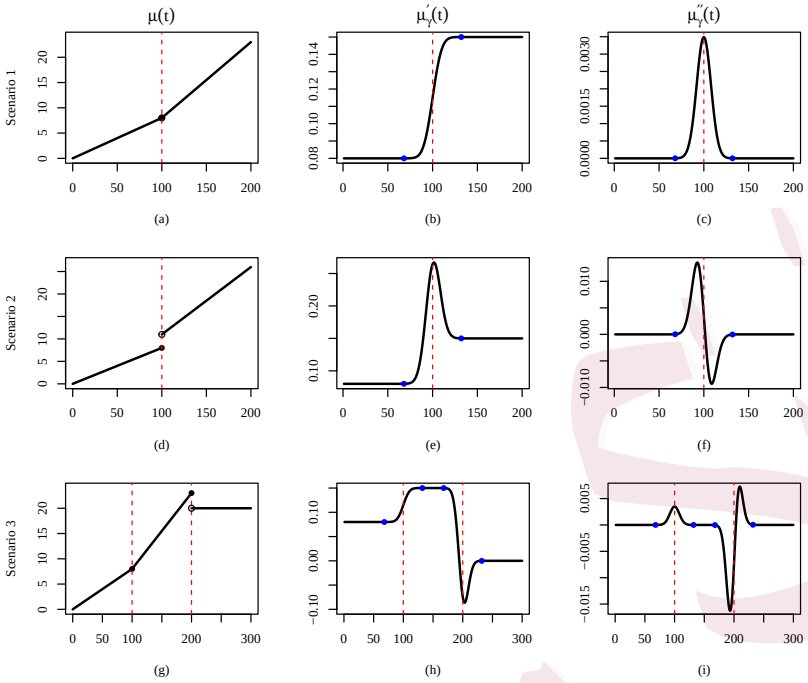

Figure 1: Illustration of change point detection. A change point in the piecewise linear signal $\mu(t)$ (left panel) becomes a local extremum in either the first derivative of the smoothed signal $\mu'_\gamma(t)$ (middle panel) or in the second derivative of the smoothed signal $\mu''_\gamma(t)$ (right panel). The red dashed lines indicate the locations of true change points, and the blue points represent the regions where the signal is smoothed using a Gaussian kernel. The top row shows a Type I change point v_j in $\mu(t)$ becomes a local extremum in $\mu''_\gamma(t)$ precisely at $t = v_j$. The middle row illustrates a Type II change point v_j in $\mu'_\gamma(t)$ becomes a local extremum in $\mu''_\gamma(t)$ around v_j. The bottom row shows that only Type II change points can generate local extrema in $\mu'_\gamma(t)$, whereas both Type I and Type II change points can generate local extrema in $\mu''_\gamma(t)$.

By kernel smoothing and differentiating the data sequence, we transform the problem of change point detection into identifying local extrema. Specifically, a

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Type I change point in $\mu(t)$ becomes a local extremum in the second derivative $\mu''_\gamma(t)$ (see Figure 1 (a) and (c)). A Type II change point in $\mu(t)$ becomes a local extremum in the first derivative $\mu'_\gamma(t)$ (see Figure 1 (d) and (e)). Therefore, in Scenario 1 and Scenario 2, the detection of change points relies on identifying local extrema in the second and first derivatives of the smoothed signal, respectively. However, in Scenario 3, a different approach is required due to the unique characteristics of each type. Type I change points do not generate local extrema in the first derivative $\mu'_\gamma(t)$ (see Figure 1 (b)). Therefore, in Scenario 3, the detection process begins by identifying all Type II change points as local extrema in $\mu'_\gamma(t)$. On the other hand, local extrema in the second derivative $\mu''_\gamma(t)$ can be generated by both Type I and Type II change points (see Figure 1 (i)). To detect Type I change points in Scenario 3, one needs to remove the local extrema in $\mu''_\gamma(t)$ generated by Type II change points. By combining these approaches, our proposed method can distinguish between Type I and Type II change points.

The detection of local extrema can be formulated as a peak detection problem, focusing on the local extrema of the first and second derivatives. In the literature, there have been notable contributions that address this issue. One such notable approach is the Smoothing and TEsting of Maxima (STEM) algorithm (Schwartzman et al. 2011). The STEM algorithm aims to identify local maxima of the derivative as candidate peaks, which correspond to potential

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

change points in the signal. By smoothing the signal with the kernel and applying multiple tests, the algorithm distinguishes between local extrema generated by true change points and those arising from random noise. Moreover, Cheng et al. (2020) introduced the differential STEM (dSTEM) method, specifically designed to detect change points in data sequences modeled as a piecewise constant signal plus noise. The dSTEM method leverages the principles of the STEM algorithm but adapts it to the piecewise constant signal model. This modification allows for more effective change point detection in scenarios where the signal exhibits piecewise constant behavior.

The literature on change point detection contains a large amount of statistical inference research, but most of it is specifically designed for the case of a single change point with an unknown location. For example, in the case of a single change point with an unknown location, Andrews (1993) proposed comprehensive treatments and testing methods for structural change. Perron (1989) utilized unit root tests to detect a one-time change in the level or slope of the trend function in univariate time series. Bai (1994) introduced the least squares method for estimating an unknown shift point (change point) in a piecewise constant model. In recent years, there has been extensive interest in multiple change point detection, particularly in the context of multiple testing-based methods. In the piecewise constant signal model, Yao & Au (1989) studied least squares

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

estimators for the locations and levels of the step function under both known and unknown numbers of jumps. Lavielle (2005) developed a penalized least squares method to estimate the number of change points and their locations. Binary segmentation (BS) (Vostrikova 1981) is another popular approach, Hyun et al. (2021) outlined similar post-selection tests for change point detection using Wild Binary Segmentation (WSB) (Fryzlewicz 2014) and circular Binary Segmentation (CBS) (Olshen et al. 2004). SMUCE (Frick et al. 2014) estimates the number of change points as the minimum among all candidate fits, where the empirical residuals pass a certain multi-scale test at a given significance level. Li et al. (2016) proposed an estimator that controls the False Discovery Rate (FDR) while allowing for a generous definition of a true discovery. FDR control was also demonstrated for the SaRa estimator (Hao et al. 2013) and the dSTEM estimator (Cheng et al. 2020) for multiple change point locations. For the continuous (Type I) change point, the underlying model corresponds to the classical bent–line (segmented) regression framework, where two linear segments meet at an unknown knot. A rich literature exists on this problem, including likelihood–based and grid–search approaches (Lerman 1980), as well as more recent computationally efficient estimation methods such as the iterative procedure of Muggeo (2003). Additionally, Bai & Perron (1998) provided asymptotic distribution results regarding the distance between estimated change points and

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

their true locations, assuming a known number of change points. Furthermore, methods such as narrowest-over-threshold (NOT) (Baranowski et al. 2019) and narrowest significance pursuit (NSP) (Fryzlewicz 2024) have been proposed to detect change points by focusing on localized regions that contain suspected features. These works represent a diverse range of approaches for change point detection, providing valuable insights and methodologies for addressing both single and multiple change point scenarios in various signal models.

In this paper, we propose a modified Smoothing and TEsting of Maxima (mSTEM) algorithm for change point detection in piecewise linear models. Our proposed approach is unique compared to the existing literature in the following ways: 1. Our method allows for the simultaneous estimation of both the number and locations of change points; 2. Our method can effectively distinguish between Type I and Type II change points; 3. Our proposed method achieves significantly lower computational complexity compared to traditional change point detection methods. By testing only the candidate peaks generated by true change points or random noise once, the method reduces the computation to the number of candidate peaks, which is proportional to the length of the data sequence; 4. We derive a general and straightforward form of the peak height distribution for derivatives of Gaussian processes, which depends on only one parameter. This generalizes and extends the distributions used in Schwartzman et al. (2011); 5.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Our method takes into account the presence of correlated noise by assuming that the noise follows a Gaussian process.

This paper is organized as follows. Section 2 introduces the framework of our change point detection method. Section 3 provides the change point detection algorithms in different scenarios and their asymptotic theories. Section 4 presents the simulation studies for various signal settings. Two real data studies are described in Section 5. Section 6 concludes with a brief discussion.

2. The mSTEM detection framework

2.1 The kernel smoothed signal

Consider the model (1.1). The jump size a_j at v_j is defined as

$$a_j = (c_{j+1} - c_j) + (k_{j+1} - k_j)v_j, \ j \geq 1. \tag{2.1}$$

By Definition 1, a change point v_j is a Type I change point if $a_j = 0$ and a Type II change point if $a_j \neq 0$. Assume $d = \inf_j(v_j - v_{j-1}) > 0$ that there is a minimal distance d between the neighboring change points. In addition, we assume $k = \inf_j |k_{j+1} - k_j| > 0$ at Type I change points and $a = \inf_j |a_j| > 0$ at Type II change points, respectively, so that slope changes and jump sizes do not become arbitrarily small.

Denote by $\phi(x)$ and $\Phi(x)$ the pdf and cdf of the standard normal distribution, respectively. Let $w_\gamma(t)$ be the Gaussian kernel with compact sup-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

port $[-c\gamma, c\gamma]$ (we let $c = 4$ throughout this paper) and bandwidth γ, i.e., $w_\gamma(t) = \frac{1}{\gamma}\phi\left(\frac{t}{\gamma}\right)\mathbf{1}\{-c\gamma \le t \le c\gamma\}$. Convolving the signal-plus-noise process (1.1) with the kernel $w_\gamma(t)$ leads to the generation of a smoothed random process:

$$y_\gamma(t) = w_\gamma(t) * y(t) = \int_{\mathbb{R}} w_\gamma(t-s)y(s)\,ds = \mu_\gamma(t) + z_\gamma(t), \tag{2.2}$$

where the smoothed signal and smoothed noise are defined, respectively, as $\mu_\gamma(t) = w_\gamma(t) * \mu(t)$ and $z_\gamma(t) = w_\gamma(t) * z(t)$. The smoothed noise $z_\gamma(t)$ is assumed to be zero-mean and four-times differentiable. To avoid the overlap of smoothing between two neighboring change points, we assume that $d = \inf_j(v_j - v_{j-1})$ is large enough, specifically, greater than $2c\gamma$.

2.2 Local extrema for derivatives of the smoothed signal

For a smooth function $f(t)$, denote by $f^{(\ell)}(t)$ its derivative ℓ, $\ell \ge 1$, and write by default $f'(t) = f^{(1)}(t)$ and $f''(t) = f^{(2)}(t)$, respectively. We have the following derivatives of the smoothed observed process (2.2),

$$y_\gamma^{(\ell)}(t) = w_\gamma^{(\ell)}(t) * y(t) = \int_{\mathbb{R}} w_\gamma^{(\ell)}(t-s)y(s)\,ds = \mu_\gamma^{(\ell)}(t) + z_\gamma^{(\ell)}(t), \tag{2.3}$$

where the derivatives of the smoothed signal and smoothed noise are $\mu_\gamma^{(\ell)}(t) = w_\gamma^{(\ell)}(t) * \mu(t)$ and $z_\gamma^{(\ell)}(t) = w_\gamma^{(\ell)}(t) * z(t)$ for $\ell \ge 1$, respectively. The following lemmas on the first and second derivatives of the smoothed signal $\mu_\gamma(t)$ are useful for characterizing the local extrema (see Figure 1).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Lemma 1. *The first and second derivatives of* $\mu_\gamma(t)$ *over* $(v_{j-1}+c\gamma,\ v_{j+1}-c\gamma)$ *are given, respectively, by*

$$\mu'_\gamma(t)=\begin{cases} k_j[2\Phi(c)-1], & t\in(v_{j-1}+c\gamma,\ v_j-c\gamma),\\ \dfrac{a_j}{\gamma}\phi\left(\dfrac{v_j-t}{\gamma}\right)+(k_j-k_{j+1})\Phi\left(\dfrac{v_j-t}{\gamma}\right) & \\ \qquad +(k_j+k_{j+1})\Phi(c)-k_j, & t\in(v_j-c\gamma,\ v_j+c\gamma),\\ k_{j+1}[2\Phi(c)-1], & t\in(v_j+c\gamma,\ v_{j+1}-c\gamma); \end{cases}$$

$$\mu''_\gamma(t)=\begin{cases} \dfrac{a_j(v_j-t)+(k_{j+1}-k_j)\gamma^2}{\gamma^3}\phi\left(\dfrac{v_j-t}{\gamma}\right), & t\in(v_j-c\gamma,\ v_j+c\gamma),\\ 0, & \textit{otherwise}. \end{cases} \tag{2.4}$$

where a_j *is the jump size defined by* (2.1).

Lemma 2. *The local extremum of* $\mu'_\gamma(t)$ *over* $(v_j-c\gamma, v_j+c\gamma)$ *is achieved at*

$$t=\begin{cases} v_j+\gamma^2 q_j, & \textit{if } a_j\neq 0, \qquad (2.5a)\\ \textit{does not exist}, & \textit{if } a_j=0; \qquad (2.5b) \end{cases}$$

while the local extremum of $\mu''_\gamma(t)$ *over* $(v_j-c\gamma, v_j+c\gamma)$ *is achieved at*

$$t=\begin{cases} v_j+\dfrac{1}{2}\left(\gamma^2 q_j\pm\gamma\sqrt{\gamma^2q_j^2+4}\right), & \textit{if } a_j\neq 0, \qquad (2.6a)\\ v_j, & \textit{if } a_j=0, \qquad (2.6b) \end{cases}$$

where $q_j=\frac{k_{j+1}-k_j}{a_j}$ *for* $a_j\neq 0$.

Recall that a_j is the jump size at the change point v_j; therefore, v_j is a Type I change point if $a_j=0$, and a Type II change point if $a_j\neq 0$. Throughout this

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

paper, derivatives are taken with respect to the smoothed signal $\mu_\gamma(t)$, which ensures that derivatives are well defined even in the presence of Type II change points, where the underlying signal is not continuous. Note that q_j is the ratio between the slope change and the jump size. The Lemma 2 characterizes the relationship between the peak location of the differentiated smoothed signal and the original location v_j, producing the following result.

Proposition 1. *A Type* I *change point* v_j *becomes a local extremum in* $\mu''_\gamma(t)$ *precisely at* v_j *(see* (2.6b)*); while a Type* II *change point* v_j *results in a local extremum in* $\mu'_\gamma(t)$ *at* $v_j + \gamma^2 q_j$ *(see* (2.5a)*), which tends to* v_j *as* $q_j \to 0$.

More specifically, a Type I change point v_j becomes a local maximum in $\mu''_\gamma(t)$ if $k_{j+1} - k_j > 0$ (a local minimum if $k_{j+1} - k_j < 0$); and a Type II change point v_j results in a local maximum in $\mu'_\gamma(t)$ near v_j if $a_j > 0$ (a local minimum if $a_j < 0$), regardless of the sign of $k_{j+1} - k_j$. Basically, Proposition 1 shows that the change points would be transformed into peaks in $\mu'_\gamma(t)$ or $\mu''_\gamma(t)$, providing the main idea for the detection in the pure cases of Scenarios 1 and 2.

Remark 1. Based on the results in the lemma 2, we discuss the idea of detecting the change points of Types I and II separately in Scenario 3, where both types of change points. This will be shown in more detail in Algorithm 3 below.

First, note that a Type I change point does not generate any local extremum in $\mu'_\gamma(t)$ (see (2.5b)). This property is crucial for detecting Type II change points

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

by identifying peaks in $\mu'_\gamma(t)$ as stated by Proposition 1 (see step 1 in Algorithm 3). On the other hand, as mentioned by (2.6a), a Type II change point v_j will generate a pair of local maximum and minimum in $\mu''_\gamma(t)$, with locations tending to $v_j \pm \gamma$ when q_j is sufficiently small. This will help us remove those local extrema in $\mu''_\gamma(t)$ generated from the detected Type II change points in the previous step and thus detect the Type I change points by identifying peaks in $\mu''_\gamma(t)$ as outlined by Proposition 1 (see step 2 in Algorithm 3).

2.3 Main ideas

Figure 1 demonstrates the key concept of the proposed change point detection method. The main idea is to transform the problem of change point detection into testing local extrema in the derivatives of the smoothed signal. A candidate peak can arise from either a true change point in the signal region or pure random noise. Multiple testing based on the peak height distribution of the first and second derivatives is then used to identify the significant peaks as the true change points. Although this main idea (or Figure 1) is illustrated in signal $\mu(t)$, it is also applicable to signal-plus-noise $y(t)$ under certain asymptotic conditions such that the signal strength (the size of jump or slope change) and the length of the data sequence are large; see conditions (C1) and (C2) below.

To further illustrate the main idea, especially the proposed algorithm, toy examples are presented in Figures 2 and 3. In Scenario 1 (pure Type I change

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

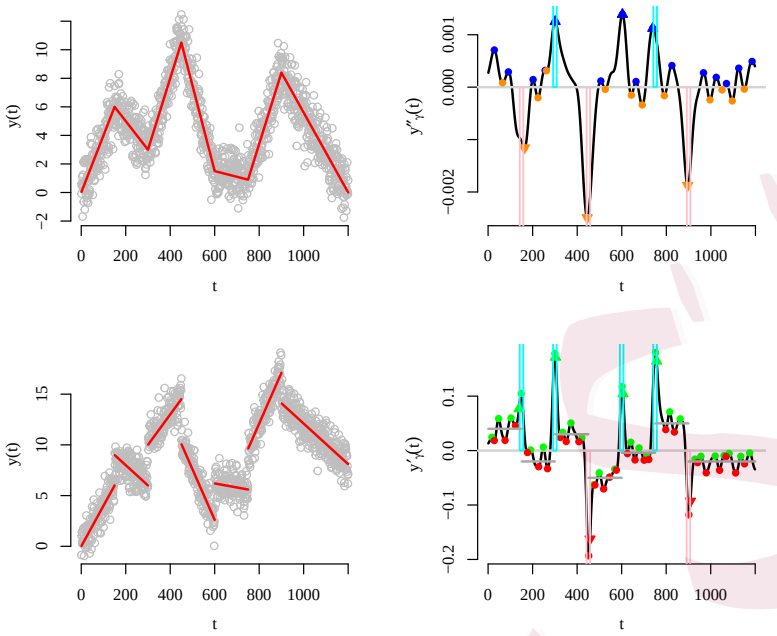

Figure 2: Process of Type I and Type II change point detection. The two plots on the left panel show the data (grey circles) with six Type I change points and the data with six Type II change points respectively. The red lines indicate piecewise linear signals $\mu(t)$. The two plots on the right panels show the second and first derivatives of the smoothed data ($\gamma = 10$), respectively. In $y'_\gamma(t)$, local maxima and local minima are represented by green and red dots respectively. In $y''_\gamma(t)$, they are represented by blue and orange dots respectively. Triangles indicate the significant local extrema at significant level $\alpha = 0.05$. The cyan and pink bars show the location tolerance intervals $(v_j - b, v_j + b)$ with $b = 5$ for the true change points. The grey horizontal lines (lower right plot) indicate piecewise slopes of the signal, serving as baselines in the testing of $y'_\gamma(t)$. In these two testings, all change points of both Type I and Type II are detected correctly without any false discovery.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

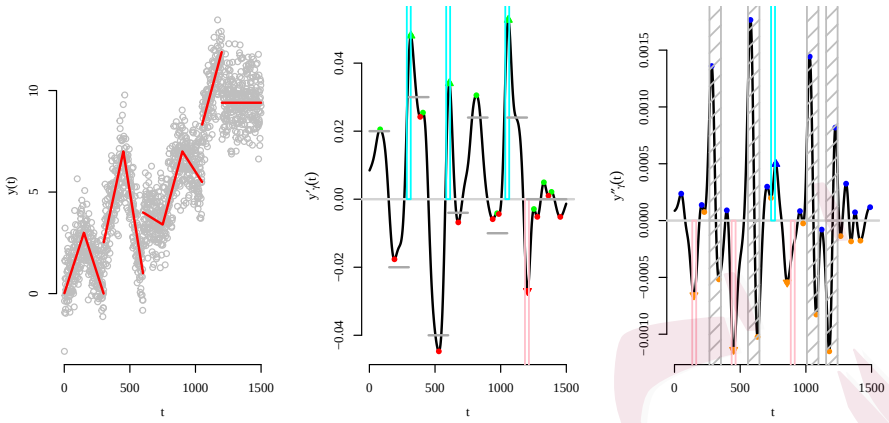

Figure 3: Process of mixture of Type I and Type II change point detection. The same graphical symbols and colors as in Figure 2 are utilized here. The left plot (grey circles) shows the data with four Type I and four Type II change points. The middle and right plots show the detection of Type II change points in $y'_\gamma(t)$ and the detection of Type I change points in $y''_\gamma(t)$, respectively. Type II change points are first detected in the first derivative $y'_\gamma(t)$, then the local extrema in the second derivative $y''_\gamma(t)$ generated by the detected Type II change points are removed (represented by grey shaded bars and each bar is $[v_j - 2\gamma, v_j + 2\gamma]$). After this filtering process, the remaining significant local extrema in $y''_\gamma(t)$ are identified as Type I change points. In this testing, all Type I and Type II change points are detected correctly without any false discovery.

points), a Type I change point generates a peak exactly at its location in the second derivative $\mu''_\gamma(t)$, as indicated in (2.6b). In Scenario 2 (pure Type II change points), a Type II change point produces a peak in the first derivative $\mu'_\gamma(t)$ around its location, as described in (2.5a). By subtracting the corresponding baseline (piecewise slope at v_j) from the peaks in $y'_\gamma(t)$, the problem reduces

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

to a standard peak detection problem. In Scenario 3, a combination of Type I and Type II change points is considered. The Type II change points are first identified based on the peaks in $y'_\gamma(t)$, and then the Type I change points are detected as significant peaks in $y''_\gamma(t)$ after cutting off the Type II change points.

Our proposed method, called mSTEM (**m**odified **S**moothing and **TE**sting of **M**axima/**M**inima), consists of the following steps. 1. Differential kernel smoothing: to transform change points into local maxima or minima (illustrated in Figure 1), and can meanwhile increase SNR; 2. Candidate peaks: to find local extrema of either the first or second derivative of the smoothed observed data; 3. P-values: to compute the p-value of each local maximum or minimum under the null hypothesis of no change point (no signal) in a local neighborhood; 4. Multiple testing: to apply a multiple testing procedure (Benjamini-Hochberg procedure) to the set of local maxima and minima; a change point is claimed to be detected if the p-value is significant.

3. Multiple change point detection for linear models

Suppose that we observe $y(t)$ defined by (1.1) with J change points in the line of length L centered at the origin, which is denoted by $U(L) = (-L/2, L/2)$. We define *signal region* as $\mathbb{S}_1 = \cup_{j=1}^{J} S_j = \cup_{j=1}^{J}(v_j - b, v_j + b)$, which is the region of interest where true change points are expected to occur. It is defined as

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

the union of intervals centered around each true change point v_j. The *null region* is the complement of the signal region within $U(L)$ given by $\mathbb{S}_0 = U(L) \setminus \mathbb{S}_1$, where true change points are not expected to occur.

3.1 Pure Type I change point detection

We first consider the case of pure type I change points, which do not generate any peaks in $\mu'_\gamma(t)$ by (2.5b). Applying Proposition 1, it is appropriate to detect the Type I change points by detecting peaks in $y''_\gamma(t)$. Following the mSTEM procedure proposed for the detection of change points in Section 2.3, we present the mSTEM algorithm for detecting Type I change points.

Algorithm 1 (mSTEM algorithm for Type I change point detection)**.**

1. Differential kernel smoothing*: Obtain the process $y''_\gamma(t)$ in (2.3) by convolution of $y(t)$ with the kernel derivative $w''_\gamma(t)$.*

2. Candidate peaks*: Find the set of local maxima and minima of $y''_\gamma(t)$ in $U(L)$, denoted by $\tilde{T}_{\mathrm{I}} = \tilde{T}_{\mathrm{I}}^{+} \cup \tilde{T}_{\mathrm{I}}^{-}$, where $\tilde{T}_{\mathrm{I}}^{+} = \{t \in U(L) : y_\gamma^{(3)}(t) = 0,\ y_\gamma^{(4)}(t) < 0\}$, $\tilde{T}_{\mathrm{I}}^{-} = \{t \in U(L) : y_\gamma^{(3)}(t) = 0,\ y_\gamma^{(4)}(t) > 0\}$. Note that $\tilde{T}_{\mathrm{I}}^{+}$ and $\tilde{T}_{\mathrm{I}}^{-}$ are the sets of local maxima and minima in $y''_\gamma(t)$.*

3. P-values*: For each $t \in \tilde{T}_{\mathrm{I}}^{+}$, compute the p-value $p_{\mathrm{I}}(t)$ for testing the hypotheses $\mathcal{H}_0(t) : \{\mu''(s) = 0$ for all $s \in (t-b, t+b)\}$ vs. $\mathcal{H}_A(t) : \{\mu'(s+) > \mu'(s-)$ for some $s \in (t-b, t+b)\}$; and for each $t \in \tilde{T}_{\mathrm{I}}^{-}$, compute the p-value*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

$p_{\mathrm{I}}(t)$ *for the hypotheses* $\mathcal{H}_0(t) : \{\mu''(s) = 0$ *for all* $s \in (t-b, t+b)\}$ vs. $\mathcal{H}_A(t) : \{\mu'(s+) < \mu'(s-)$ *for some* $s \in (t-b, t+b)\}$, *where* $\mu'(s+) = \lim_{x \to s+} \mu'(x)$, $\mu'(s-) = \lim_{x \to s-} \mu'(x)$ *and* $b > 0$ *is an appropriate location tolerance parameter.*

4. Multiple testing*: Apply the Benjamini-Hochberg (BH) multiple testing procedure on the set of p-values* $\{p_{\mathrm{I}}(t),\, t \in \tilde{T}_{\mathrm{I}}\}$, *defined in Section 3.1.1 below, and declare all significant local extrema whose p values are lower than the significance threshold.*

3.1.1 P-value calculation

The p-value in step 3 of Algorithm 1 is given by

$$p_{\mathrm{I}}(t) = \begin{cases} F_{z''_\gamma}(y''_\gamma(t)) & t \in \tilde{T}_{\mathrm{I}}^{+}, \\ F_{z''_\gamma}(-y''_\gamma(t)) & t \in \tilde{T}_{\mathrm{I}}^{-}, \end{cases} \tag{3.1}$$

where $F_{z''_\gamma}(u)$ (defined in Section 3.4 below) is the conditional right tail probability of $z''_\gamma(t)$ at the local maximum $t \in \tilde{T}_{\mathrm{I}}^{+}$, evaluated under the null model $\mu''_\gamma(s) = 0, \forall s \in (t-b, t+b)$, i.e.,

$$F_{z''_\gamma}(u) = P\left(z''_\gamma(t) > u \middle|\, t \text{ is a local maximum of } z''_\gamma(t)\right). \tag{3.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

The second line in (3.1) follows from the symmetry property:

$$\begin{aligned}&\mathbb{P}\left(z''_\gamma(t) < u \mid t \text{ is a local minimum of } z''_\gamma\right)\\ &\quad = \mathbb{P}\left(-z''_\gamma(t) > -u \mid t \text{ is a local maximum of } -z''_\gamma\right) = F_{z''_\gamma}(-u),\end{aligned}$$

since $z''_\gamma(t)$ and $-z''_\gamma(t)$ have the same distribution.

3.1.2 Error and power definitions

We define

$$\widetilde{T}_I^{\mathrm{I}}(u) = \widetilde{T}_I^{+,\mathrm{I}}(u) \cup \widetilde{T}_I^{-,\mathrm{I}}(u),$$

where

$$\begin{aligned}\widetilde{T}_I^{+,\mathrm{I}}(u) &= \left\{t \in U(L) : y''_\gamma(t) > u,\ y^{(3)}_\gamma(t) = 0,\ y^{(4)}_\gamma(t) < 0\right\},\\ \widetilde{T}_I^{-,\mathrm{I}}(u) &= \left\{t \in U(L) : y''_\gamma(t) < -u,\ y^{(3)}_\gamma(t) = 0,\ y^{(4)}_\gamma(t) > 0\right\}.\end{aligned}$$

It is seen that $\tilde{T}_{\mathrm{I}}^{+}(u)$ and $\tilde{T}_{1}^{-}(u)$ are the sets of local maxima of $y''_\gamma(t)$ above u and the sets of local minima of $y''_\gamma(t)$ below $-u$, respectively. The number of totally and falsely detected change points at threshold u is defined, respectively, as $R_{\mathrm{I}}(u) = \#\{t \in \tilde{T}_{\mathrm{I}}(u)\}$ and $V_{\mathrm{I}}(u) = \#\{t \in \tilde{T}_{\mathrm{I}}(u) \cap \mathbb{S}_0\}$. Both are defined as zero if $\tilde{T}_{\mathrm{I}}(u)$ is empty. The FDR at threshold u is defined as the expected proportion of falsely detected change points $\mathrm{FDR}_{\mathrm{I}}(u) = \mathrm{E}\left(\frac{V_{\mathrm{I}}(u)}{R_{\mathrm{I}}(u) \vee 1}\right)$.

Denote by $\mathcal{I}_{\mathrm{I}}^{+}$ and $\mathcal{I}_{\mathrm{I}}^{-}$ the collections of indices j corresponding to the increasing and decreasing (in slope change) Type I change points, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

The power is defined as

$$\begin{aligned}\text{Power}_{\text{I}}(u) &= \frac{1}{J}\sum_{j=1}^{J} P\left(\tilde{T}_{\text{I}}(u) \cap S_j \neq \emptyset\right) \\ &= \frac{1}{J}\left\{ \sum_{j\in\mathcal{I}_{\text{I}}^{+}} P\left(\tilde{T}_{\text{I}}^{+}(u) \cap S_j \neq \emptyset\right) + \sum_{j\in\mathcal{I}_{\text{I}}^{-}} P\left(\tilde{T}_{\text{I}}^{-}(u) \cap S_j \neq \emptyset\right) \right\},\end{aligned} \tag{3.3}$$

where $P(\cdot)$ is the probability of detecting the change point v_j within the signal region $S_j = (v_j - b, v_j + b)$. Throughout this paper, when considering the local extrema of the second derivative of the process over the signal region S_j or the smoothed signal region $S_{j,\gamma} = (v_j - c\gamma, v_j + c\gamma)$ around j, the local extrema are regarded as local maxima if $j \in \mathcal{I}_{\text{I}}^{+}$ and as local minima if $j \in \mathcal{I}_{\text{I}}^{-}$.

3.1.3 Asymptotic FDR control and power consistency

Denote by $E[\tilde{m}_{z''_\gamma}(U(1))]$ and $E[\tilde{m}_{z''_\gamma}(U(1), u)]$ the expected number of local extrema, and the expected number of both local maxima above the level u and local minima below $-u$ of $z''_\gamma(t)$ in the unit interval $U(1) = (-1/2, 1/2)$, respectively. Note that, by symmetry, the expected number of local minima of $z''_\gamma(t)$ below the level $-u$ is equal to the expected number of local maxima of $z''_\gamma(t)$ above u, which is $E[\tilde{m}_{z''_\gamma}(U(1), u)]/2$. Similarly, we define $E[\tilde{m}_{z'_\gamma}(U(1))]$ and $E[\tilde{m}_{z'_\gamma}(U(1), u)]$ for $z'_\gamma(t)$ on $U(1)$.

Recall that the BH procedure applied in step 4 of Algorithm 1 is defined as

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

follows. Let m_{I} be the size of the set $\tilde{T}_{\mathrm{I}}$. For a fixed $\alpha \in (0,1)$, let ℓ be the largest index for which the ith smallest p-value is less than $i\alpha/m_{\mathrm{I}}$. Then the null hypothesis $\mathcal{H}_0(t)$ at $t \in \tilde{T}_{\mathrm{I}}$ is rejected if

$$p_{\mathrm{I}}(t) < \frac{\ell\alpha}{m_{\mathrm{I}}} \iff \begin{cases} y''_\gamma(t) > \tilde{u}_{\mathrm{I}} = F^{-1}_{z''_\gamma}\left(\frac{\ell\alpha}{m_{\mathrm{I}}}\right) & \text{if } t \in \tilde{T}^+_{\mathrm{I}}, \\ y''_\gamma(t) < -\tilde{u}_{\mathrm{I}} = -F^{-1}_{z''_\gamma}\left(\frac{\ell\alpha}{m_{\mathrm{I}}}\right) & \text{if } t \in \tilde{T}^-_{\mathrm{I}}, \end{cases} \tag{3.4}$$

where $\ell\alpha/m_{\mathrm{I}}$ is defined as 1 if $m_{\mathrm{I}} = 0$. Note that $\tilde{T}_{\mathrm{I}}$ and therefore m_{I} and $\tilde{u}_{\mathrm{I}}$ are also random, which is different from the usual BH procedure. We define the FDR in this BH procedure as $\mathrm{FDR}_{\mathrm{I,BH}} = E\left\{\frac{V_{\mathrm{I}}(\tilde{u}_{\mathrm{I}})}{R_{\mathrm{I}}(\tilde{u}_{\mathrm{I}})\vee 1}\right\}$. The expectation is taken over all possible realizations of the random threshold $\tilde{u}_{\mathrm{I}}$. In contrast to the standard BH procedure, the number of p-values, m_{I}, is random. Similarly, the corresponding power, $\mathrm{Power}_{\mathrm{I,BH}}$, is defined as (3.3) with u replaced by $\tilde{u}_{\mathrm{I}}$.

For the asymptotic theories of Type I change point detection, we make the assumption:

(C1). $L \to \infty$, $k = \inf_j |k_{j+1} - k_j| \to \infty$ and $(\log L)/k^2 \to 0$.

Theorem 1. *Suppose that $y(t)$ contains only Type* I *change points, condition (C1) holds and $J/L \to A > 0$ as $L \to \infty$.*

(i) If Algorithm 1 is applied with a fixed threshold u, then

$$\begin{aligned} \mathrm{FDR}_{\mathrm{I}}(u) &\to \frac{E[\tilde{m}_{z'_\gamma}(U(1),u)](1-2c\gamma A)}{E[\tilde{m}_{z'_\gamma}(U(1),u)](1-2c\gamma A)+A}, \\ \mathrm{Power}_{\mathrm{I}}(u) &\to 1. \end{aligned} \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

(ii) If Algorithm 1 is applied with the random threshold $\tilde{u}_{\mathrm{I}}$ *in* (3.4)*, then*

$$\begin{aligned} \mathrm{FDR}_{\mathrm{I,BH}} &\to \alpha \frac{E[\tilde{m}_{z'_\gamma}(U(1))](1-2c\gamma A)}{E[\tilde{m}_{z'_\gamma}(U(1))](1-2c\gamma A)+A}, \\ \mathrm{Power}_{\mathrm{I,BH}} &\to 1. \end{aligned} \tag{3.6}$$

Under the asymptotic condition (C1), the BH procedure with the random threshold $\tilde{u}_{\mathrm{I}}$ in (3.4) has asymptotically the same error control properties as if the threshold were deterministic and given by $u^*_{\mathrm{I}} = F^{-1}_{z''_\gamma}(\frac{\alpha A}{A+E[\tilde{m}_{z''_\gamma}(U(1))](1-2c\gamma A)(1-\alpha)})$ (see Section S6 in Supplementary Materials).

By the definition of d, we have $Jd < L$ or $d < L/J \to 1/A$. Meanwhile, it is assumed $d = \inf_j(v_j - v_{j-1}) > 2c\gamma$, thus we have $1/A > 2c\gamma$ and hence $1 - 2c\gamma A > 0$. Note that $2c\gamma J$ represents the length of smoothed signal regions and $(L - 2c\gamma J)$ is the length of smoothed null regions, which are the complements of smoothed signal regions. Part (i) of Theorem 1 shows that FDR converges to the ratio of the number of local extrema generated by random noise to the total number of local extrema, which includes those induced by random noise and the true change points. Part (ii) guaranties that FDR is asymptotically controlled at a prespecified general significant level α.

Remark 2. *Condition (C1) assumes that the signal strength* k *increases with the search space* L *to remove excess error caused by the smoothed signal spreading into neighboring null regions, which are the complements of signal regions,*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

thereby enabling asymptotic error control. This assumption is not restrictive, as the search space grows exponentially faster than the signal strength. Viewing the data as pointwise test statistics, this assumption is analogous to the conditions required for the consistency of model selection in high-dimensional regression. Specifically, $(\log p)/n \to 0$ *where* p *is the number of features and* n *is the sample size. Additionally, condition (C1) is easy to state but is much stronger than necessary. It suffices for* $(\log L)/k^2$ *to be bounded by a constant that depends on the third or fourth derivatives of the smoothed noise (see Remark S1 in the Supplementary Materials). The numerical studies in Section 4 further demonstrate that FDR control and power consistency are maintained even in non-asymptotic settings with moderate slope changes.*

3.2 Pure Type II change point detection

In the detection of Type I breaks, taking the second derivative, the piecewise linear parts of $\mu(t)$ become 0 in $\mu''_\gamma(t)$ over $\mathbb{R} \setminus \mathbb{S}_{1,\gamma}$, where $\mathbb{S}_{1,\gamma} = \cup_{j=1}^{J}(v_j - c\gamma, v_j + c\gamma)$ is the *smoothed signal region* (see the intervals indicated by the blue points in Figure 1). However, in the detection of Type II breaks, the piecewise linear parts become piecewise constants (the slopes) in $\mu'_\gamma(t)$ over $\mathbb{R} \setminus \mathbb{S}_{1,\gamma}$. To detect Type II breaks, the null hypothesis (no signal/jump) $\mu'_\gamma(t) = k(t)$, where $k(t)$ is the corresponding piecewise slope at t, will be tested. Hence, it is crucial to estimate the piecewise slopes around the change points v_j.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

3.2.1 Piecewise slopes estimate

The main idea of estimating piecewise slopes is to divide the data sequence into segments and estimate the slopes within each segment using linear regression. These segments are determined based on the presence of pairwise local maxima and minima, which are located near $v_j \pm \gamma$ around the change points in the second derivative $\mu''_\gamma(t)$ (see Figure 1 (f)).

To ensure that all true Type II breaks are captured and to provide a better estimate of the piecewise slopes, Algorithm 1 is applied with a larger significance level, such as 0.1. This helps in obtaining non-conservative initial estimators for Type II change points based on the peaks detected in $y''_\gamma(t)$. Then, within each segment, the piecewise slope is estimated using a robust regression model (Huber 2004) to mitigate the influence of the change points.

Once the piecewise slopes are obtained, we will apply the following algorithm, which is based on the local extrema of the first derivative of the smoothed data, to detect Type II change points.

Algorithm 2 (mSTEM algorithm for Type II change point detection)**.**

1. Differential kernel smoothing*: Obtain the process* $y'_\gamma(t)$ *in* (2.3) *by convolution of* $y(t)$ *with the kernel derivative* $w'_\gamma(t)$.

2. Candidate peaks*: Find the set of local maxima and minima of* $y'_\gamma(t)$ *in* $U(L)$,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

denoted by $\tilde{T}_{\text{II}} = \tilde{T}_{\text{II}}^{+} \cup \tilde{T}_{\text{II}}^{-}$, *where*

$$\begin{aligned} \tilde{T}_{II}^{+,\text{II}} &= \left\{ t \in U(L) : y'_\gamma(t) = 0,\ y^{(3)}_\gamma(t) < 0 \right\}, \\ \tilde{T}_{II}^{-,\text{II}} &= \left\{ t \in U(L) : y'_\gamma(t) = 0,\ y^{(3)}_\gamma(t) > 0 \right\}. \end{aligned} \tag{3.7}$$

3. P-values*: For each* $t \in \tilde{T}_{\text{II}}^{+}$, *compute the p-value* $p_{\text{II}}(t)$ *for testing the hypotheses*

$$\mathcal{H}_0(t) : \{\mu'(s) = k(s) \quad \textit{for all} \quad s \in (t-b, t+b)\} \quad \text{vs.}$$

$$\mathcal{H}_A(t) : \{\mu(s+) > \mu(s-) \quad \textit{for some} \quad s \in (t-b, t+b)\},$$

where $k(s)$ *is the piecewise slope at* s, $\mu(s+) = \lim_{x \to s+} \mu(x)$, $\mu(s-) = \lim_{x \to s-} \mu(x)$ *and* $b > 0$ *is an appropriate location tolerance parameter. For each* $t \in \tilde{T}_{\text{II}}^{-}$, *compute the p-value* $p_{\text{II}}(t)$ *for testing the hypotheses* $\mathcal{H}_0(t)$: $\{\mu'(s) = k(s)$ *for all* $s \in (t-b, t+b)\}$ vs. $\mathcal{H}_A(t)$: $\{\mu(s+) < \mu(s-)$ *for some* $s \in (t-b, t+b)\}$.

4. Multiple testing*: Apply the BH procedure on the set of p-values* $\{p_{\text{II}}(t),\ t \in \tilde{T}_{\text{II}}\}$, *and declare significant all local extrema whose p-values are smaller than the significance threshold.*

Similarly to (3.1) and (3.2), the p-value in step 3 of Algorithm 2 is given by

$$p_{\text{II}}(t) = \begin{cases} F_{z'_\gamma}(y'_\gamma(t) - k(t)) & t \in \tilde{T}_{\text{II}}^{+}, \\ F_{z'_\gamma}(-(y'_\gamma(t) - k(t))) & t \in \tilde{T}_{\text{II}}^{-}. \end{cases}$$

Here $F_{z'_\gamma}(u)$ is the right tail probability of $z'_\gamma(t)$ at the local maximum $t \in \tilde{T}_{\text{II}}^{+}$, as defined in Section 3.4 below, which is $F_{z'_\gamma}(u) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

$P\left(z'_\gamma(t) > u \middle|\ t \text{ is a local maximum of } z'_\gamma(t)\right)$.

3.2.2 Error and power definitions

We define $\tilde{T}_{\text{II}}(u) = \tilde{T}^+_{\text{II}}(u) \cup \tilde{T}^-_{\text{II}}(u)$, where

$$\tilde{T}^+_{\text{II}}(u) = \left\{t \in U(L) : y'_\gamma(t) - k(t) > u,\ y''_\gamma(t) = 0,\ y^{(3)}_\gamma(t) < 0\right\},$$

$$\tilde{T}^-_{\text{II}}(u) = \left\{t \in U(L) : y'_\gamma(t) - k(t) < -u,\ y''_\gamma(t) = 0,\ y^{(3)}_\gamma(t) > 0\right\}.$$

The number of totally and falsely detected change points and FDR at threshold u are defined as $R_{\text{II}}(u) = \#\{t \in \tilde{T}_{\text{II}}(u)\}$, $V_{\text{II}}(u) = \#\{t \in \tilde{T}_{\text{II}}(u) \cap \mathbb{S}_0\}$, $\text{FDR}_{\text{II}}(u) = \text{E}\left\{\frac{V_{\text{II}}(u)}{R_{\text{II}}(u)\vee 1}\right\}$, respectively.

Denote by $\mathcal{I}^+_{\text{II}}$ and $\mathcal{I}^-_{\text{II}}$ the sets of indices j corresponding to increasing and decreasing (in jump) Type II change points v_j, respectively. The power is defined as

$$\begin{aligned}\text{Power}_{\text{II}}(u) &= \frac{1}{J}\sum_{j=1}^{J} P\left(\tilde{T}_{\text{II}}(u) \cap S_j \neq \emptyset\right) \\ &= \frac{1}{J}\left[\sum_{j\in\mathcal{I}^+_{\text{II}}} P\left(\tilde{T}^+_{\text{II}}(u) \cap S_j \neq \emptyset\right) + \sum_{j\in\mathcal{I}^-_{\text{II}}} P\left(\tilde{T}^-_{\text{II}}(u) \cap S_j \neq \emptyset\right)\right].\end{aligned} \tag{3.8}$$

Throughout this paper, when considering local extrema over the signal region $S_j = (v_j - b, v_j + b)$ or the smoothed signal region $S_{j,\gamma} = (v_j - c\gamma, v_j + c\gamma)$ around j, the local extrema of the derivative of the process are regarded as local maxima if $j \in \mathcal{I}^+_{\text{II}}$ and as local minima if $j \in \mathcal{I}^-_{\text{II}}$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

3.2.3 Asymptotic FDR control and power consistency

Suppose the BH procedure is applied in step 4 of Algorithm 2 as follows. Define m_{II} as the size of set $\tilde{T}_{\text{II}}$. For a fixed $\alpha \in (0,1)$, let ℓ be the largest index for which the ith smallest p-value is less than $i\alpha/m_{\text{II}}$. Then the null hypothesis $\mathcal{H}_0(t)$ at $t \in \tilde{T}_{\text{II}}$ is rejected if

$$p_{\text{II}}(t) < \frac{\ell\alpha}{m_{\text{II}}} \iff \begin{cases} y'_\gamma(t) - k(t) > \tilde{u}_{\text{II}} = F^{-1}_{z'_\gamma}\left(\frac{\ell\alpha}{m_{\text{II}}}\right) & \text{if } t \in \tilde{T}^+_{\text{II}}, \\ y'_\gamma(t) - k(t) < -\tilde{u}_{\text{II}} = -F^{-1}_{z'_\gamma}\left(\frac{\ell\alpha}{m_{\text{II}}}\right) & \text{if } t \in \tilde{T}^-_{\text{II}}, \end{cases} \tag{3.9}$$

where $F_{z'_\gamma}$ is defined in (3.2) with z''_γ replaced by z'_γ, and $\ell\alpha/m_{\text{II}}$ is defined as 1 if $m_{\text{II}} = 0$. Since $\tilde{u}_{\text{II}}$ is random, we define the FDR in such a BH procedure as $\text{FDR}_{\text{II,BH}} = E\left\{\frac{V_{\text{II}}(\tilde{u}_{\text{II}})}{R_{\text{II}}(\tilde{u}_{\text{II}})\vee 1}\right\}$. Similarly, the corresponding power, denoted by $\text{Power}_{\text{II,BH}}$, is defined as (3.3) with u replaced by $\tilde{u}_{\text{II}}$.

To establish asymptotic theories for Type II change point detection, we make the assumption:

(C2). $a = \inf_j |a_j| \to \infty$, $q = \sup_j |\frac{k_{j+1}-k_j}{a_j}| \to 0$ and $\frac{\log L}{a^2} \to 0$ as $L \to \infty$.

Theorem 2. *Suppose $y(t)$ contains only Type* II *change points, the condition (C2) holds and $J/L \to A > 0$ as $L \to \infty$.*

(i) If Algorithm 2 is applied with a fixed threshold u, then

$$\text{FDR}_{\text{II}}(u) \to \frac{E[\tilde{m}_{z'_\gamma}(U(1),u)](1-2c\gamma A)}{E[\tilde{m}_{z'_\gamma}(U(1),u)](1-2c\gamma A)+A} \quad \textit{and} \quad \text{Power}_{\text{II}}(u) \to 1.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

(ii) If Algorithm 2 is applied with a random threshold $\tilde{u}_{\mathrm{II}}$ (3.9)*, then*

$$\mathrm{FDR}_{\mathrm{II,BH}} \to \alpha \frac{E[\tilde{m}_{z'_\gamma}(U(1))](1-2c\gamma A)}{E[\tilde{m}_{z'_\gamma}(U(1))](1-2c\gamma A)+A} \quad \textit{and} \quad \mathrm{Power}_{\mathrm{II,BH}} \to 1.$$

Condition (C2) is similar to condition (C1) but focuses on the jump size. It requires the jump size to dominate the slope changes, ensuring a clear distinction between Type I and Type II change points. Similarly, the limit of $(\log L)/a^2$ needs only to be bounded by a constant that depends on the second and third derivatives of the smoothed signal (see Lemma S3 and Remark S1 in Supplementary Materials).

3.3 Mixture of Type I and Type II change point detection

We have shown that pure Type I and Type II change points can be detected through peak detection in the second derivative $y''_\gamma(t)$ (see Algorithm 1) and the first derivative $y'_\gamma(t)$ (see Algorithm 2), respectively. However, it is very common for a real signal to contain both Type I and Type II change points. Distinguishing between Type I and Type II change points is a key challenge when managing signals comprising a combination of these types. Our strategy involves initially identifying Type II change points via peak detection within the first derivative $y'_\gamma(t)$, and then detecting Type I change points by excluding the locations of Type II change points from consideration in the second derivative $y''_\gamma(t)$. The rationale behind this approach is as follows: Type II change points generate peaks in the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

first derivative $y'_\gamma(t)$, while Type I change points do not. By detecting significant peaks in $y'_\gamma(t)$ using Algorithm 2, we can identify the locations of Type II change points. Once the Type II change points are identified, we can focus on detecting Type I change points. Note that both Type I and Type II change points can produce peaks in the second derivative $y''_\gamma(t)$. However, some of these peaks may be generated by Type II change points, and we want to exclude them from our Type I detection.

To distinguish between Type I and Type II change points in the mixture case, we remove the Type II change points detected in the first step. This means that any peaks in the second derivative $y''_\gamma(t)$ that coincide with the locations of Type II change points are excluded from the set of potential Type I change points. The remaining peaks in $y''_\gamma(t)$, which are not associated with Type II change points, are more likely to be generated by Type I change points.

Algorithm 3 (mSTEM algorithm for mixture of Type I and II change point detection)**.**

1. Detect Type II breaks*: Perform Algorithm 2 to obtain estimates of Type* II *breaks, denoted as* $\hat{v}_i \in \tilde{T}_{\text{II}}$ *for* $i = 1, \ldots, R_{\text{II}}$*. A larger* γ *in this step is suggested to achieve a better estimate of Type* II *change points and fewer false discoveries. Because a larger* γ *generally results in a higher SNR (see Section 3.5).*

2. Candidate Type I peaks*: Find the set of local maxima and minima of* $y''_\gamma(t)$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

in $U(L)$*, denoted by* $\tilde{T}_{\mathrm{I}} = \tilde{T}_{\mathrm{I}}^{+} \cup \tilde{T}_{\mathrm{I}}^{-}$*. As* $\tilde{T}_{\mathrm{I}}$ *contains local extrema generated by both Type* I *and Type* II *breaks, it is necessary to remove the peaks generated by Type* II *breaks that are detected in step 1. Let* $\mathbb{S}_{\mathrm{II}}^{*} = \cup_{i=1}^{R_{\mathrm{II}}}(\hat{v}_i - 2\gamma, \hat{v}_i + 2\gamma)$ *be the removed region, then the set of candidate peaks of Type* I *breaks is defined as* $\tilde{T}_{\mathrm{I}\backslash\mathrm{II}} = \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}^{+} \cup \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}^{-}$*, where* $\tilde{T}_{\mathrm{I}\backslash\mathrm{II}}^{+} = \tilde{T}_{\mathrm{I}}^{+} \setminus \mathbb{S}_{\mathrm{II}}^{*}$ *and* $\tilde{T}_{\mathrm{I}\backslash\mathrm{II}}^{-} = \tilde{T}_{\mathrm{I}}^{-} \setminus \mathbb{S}_{\mathrm{II}}^{*}$.

3. P-values*: For each* $t \in \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}^{+}$*, compute the p-value* $p_{\mathrm{I}\backslash\mathrm{II}}(t)$ *for testing the hypotheses*

$$\mathcal{H}_0(t) : \{\mu''(s) = 0 \quad \textit{for all} \quad s \in (t-b, t+b)\} \quad \text{vs.}$$

$$\mathcal{H}_A(t) : \{\mu'(s+) > \mu'(s-) \quad \textit{for some} \quad s \in (t-b, t+b)\};$$

and for each $t \in \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}^{-}$*, compute the p-value* $p_{\mathrm{I}}(t)$ *for testing the hypotheses*

$$\mathcal{H}_0(t) : \{\mu''(s) = 0 \quad \textit{for all} \quad s \in (t-b, t+b)\} \quad \text{vs.}$$

$$\mathcal{H}_A(t) : \{\mu'(s+) < \mu'(s-) \quad \textit{for some} \quad s \in (t-b, t+b)\}.$$

4. Multiple testing*: Apply a multiple testing procedure on the set of p-values* $\{p_{\mathrm{I}\backslash\mathrm{II}}(t),\, t \in \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}\}$*, and declare significant all local extrema whose p-values are smaller than the significance threshold.*

3.3.1 Asymptotic FDR control and power consistency

Let $\mathbb{S}_{1,\mathrm{I}\backslash\mathrm{II}} = \cup_{j=1}^{J_1}(v_j - b, v_j + b) \setminus \mathbb{S}_{\mathrm{II}}^{*}$ be the signal region of Type I change points, and $\mathbb{S}_{0,\mathrm{I}\backslash\mathrm{II}} = U(L) \setminus \mathbb{S}_{1,\mathrm{I}\backslash\mathrm{II}}$ be the null region of Type I change points. Then, the number of totally and falsely detected Type I change points at threshold u

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

is defined respectively as $R_{\mathrm{I}\backslash\mathrm{II}}(u) = \#\{t \in \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}(u)\}$ and $V_{\mathrm{I}\backslash\mathrm{II}}(u) = \#\{t \in \tilde{T}_{\mathrm{I}\backslash\mathrm{II}}(u) \cap \mathbb{S}_{0,\mathrm{I}\backslash\mathrm{II}}\}$.

Given fixed thresholds u_1 and u_2 for Type I and Type II change point detection, respectively, FDR is defined as $\mathrm{FDR}_{\mathrm{III}}(u_1, u_2) = \mathrm{E}\left\{\frac{V_{\mathrm{I}\backslash\mathrm{II}}(u_1)+V_{\mathrm{II}}(u_2)}{[R_{\mathrm{I}\backslash\mathrm{II}}(u_1)+R_{\mathrm{II}}(u_2)]\vee 1}\right\}$, and the power is defined as

$$\mathrm{Power}_{\mathrm{III}}(u_1, u_2) = \frac{1}{J}\Bigg\{ \sum_{j\in\mathcal{I}_{\mathrm{I}}^{+}} P\left(\tilde{T}^{+}_{\mathrm{I}\backslash\mathrm{II}}(u_1) \cap S_j \neq \emptyset\right) + \sum_{j\in\mathcal{I}_{\mathrm{I}}^{-}} P\left(\tilde{T}^{-}_{\mathrm{I}\backslash\mathrm{II}}(u_1) \cap S_j \neq \emptyset\right) + \sum_{j\in\mathcal{I}_{\mathrm{II}}^{+}} P\left(\tilde{T}^{+}_{\mathrm{II}}(u_2) \cap S_j \neq \emptyset\right) + \sum_{j\in\mathcal{I}_{\mathrm{II}}^{-}} P\left(\tilde{T}^{-}_{\mathrm{II}}(u_2) \cap S_j \neq \emptyset\right) \Bigg\}. \tag{3.10}$$

For random thresholds $\tilde{u}_{\mathrm{I}}$ and $\tilde{u}_{\mathrm{II}}$ defined in (3.4) and (3.9) respectively, $\mathrm{FDR}_{\mathrm{III,BH}}$ is defined as the FDR with u_1 and u_2 replaced by $\tilde{u}_{\mathrm{I}}$ and $\tilde{u}_{\mathrm{II}}$ respectively. Similarly, the corresponding power, denoted by $\mathrm{Power}_{\mathrm{III,BH}}$, is defined as (3.10) with u_1 and u_2 replaced by $\tilde{u}_{\mathrm{I}}$ and $\tilde{u}_{\mathrm{II}}$, respectively.

Theorem 3. *Suppose $y(t)$ contains J_1 Type* I *change points and J_2 Type* II *change points ($J = J_1 + J_2$), conditions (C1) and (C2) hold, and $J_1/L \to A_1 > 0$ and $J_2/L \to A_2 > 0$ as $L \to \infty$.*

(i) If Algorithm 3 is applied with fixed thresholds u_1 and u_2 for Type I *and Type* II *change points respectively, then $\mathrm{Power}_{\mathrm{III}}(u_1, u_2) \to 1$ and $\limsup \mathrm{FDR}_{\mathrm{III}}(u_1, u_2)$ are bounded above by*

$$\frac{E[\tilde{m}_{z''_\gamma}(U(1), u_1)](1 - 2c\gamma A_1) + E[\tilde{m}_{z'_\gamma}(U(1), u_2)](1 - 2c\gamma A_2)}{E[\tilde{m}_{z''_\gamma}(U(1), u_1)](1 - 2c\gamma A_1) + E[\tilde{m}_{z''_\gamma}(U(1), u_2)](1 - 2c\gamma A_2) + A}. \tag{3.11}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

(ii) If Algorithm 3 is applied with random thresholds $\tilde{u}_{\mathrm{I}}$ *and* $\tilde{u}_{\mathrm{II}}$ *for Type* I *and Type* II *change points respectively, then* $\limsup \mathrm{FDR}_{\mathrm{III,BH}} \leq \alpha$ *and* $\mathrm{Power}_{\mathrm{III,BH}} \to 1$.

3.4 Gaussian auto-correlation noise and related peak height distributions

Let $X(t)$ be a centered smooth stationary Gaussian process with variance σ^2. Note that, due to stationarity, $E(X(t)X''(t)) = -\mathrm{Var}(X'(t))$. Define $\eta = -\mathrm{Corr}(X(t), X''(t)) = \frac{\mathrm{Var}(X'(t))}{\sqrt{\mathrm{Var}(X(t))\mathrm{Var}(X''(t))}}$. Let $F_X(x) = P(X(t) > x \mid t \text{ is a local maximum of } X)$ denote the peak height distribution of the Gaussian process $X(t)$. Using the Kac-Rice formula, we obtain

$$F_X(x) = 1 - \Phi\left(\frac{x}{\sigma\sqrt{1-\eta^2}}\right) + \sqrt{2\pi}\eta\phi\left(\frac{x}{\sigma}\right)\Phi\left(\frac{\eta x}{\sigma\sqrt{1-\eta^2}}\right). \tag{3.12}$$

Note that (3.12) is a general version of the peak height distribution derived in Schwartzman et al. (2011) and Cheng & Schwartzman (2015). An important and attractive characteristic of (3.12) is that, for a unit-variance process, it only depends on a single parameter η, the negative correlation between the process and its second derivative.

To implement the simulations below, we consider a specific example of $X(t)$ and derive the peak height distributions. Let the noise $z(t)$ in (1.1) be $z(t) = \int_{\mathbb{R}} \frac{1}{\nu}\phi\left(\frac{t-s}{\nu}\right) dB(s)$ with $\nu > 0$, where $dB(s)$ is Gaussian white noise ($z(t)$ is regarded by convention as Gaussian white noise when $\nu = 0$). Convolving

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

with a Gaussian kernel $w_\gamma(t) = (1/\gamma)\phi(t/\gamma)$ with $\gamma > 0$ produces a zero-mean infinitely differentiable stationary Gaussian process

$$z_\gamma(t) = \int_{\mathbb{R}} w_\gamma(t-x)z(x)dx = \int_{\mathbb{R}} \frac{1}{\xi}\phi\left(\frac{t-s}{\xi}\right) dB(s), \quad \xi = \sqrt{\gamma^2+\nu^2}. \tag{3.13}$$

Lemma 3. *Let $z_\gamma(t)$ be defined in (3.13), the variances of its derivatives are given by*

$$\mathrm{Var}(z'_\gamma(t)) = \frac{1}{4\sqrt{\pi}\xi^3}, \ \mathrm{Var}(z''_\gamma(t)) = \frac{3}{8\sqrt{\pi}\xi^5},$$
$$\mathrm{Var}(z^{(3)}_\gamma(t)) = \frac{15}{16\sqrt{\pi}\xi^7}, \ \mathrm{Var}(z^{(4)}_\gamma(t)) = \frac{105}{32\sqrt{\pi}\xi^9}.$$

Combining Lemma 3 and (3.12), we immediately obtain the following peak height distributions for the first and second derivatives of the smoothed process.

Proposition 2. *Let $z_\gamma(t)$ be defined in (3.13). Then , the peak height distributions of $z'_\gamma(t) = z^{(1)}_\gamma(t)$ and $z''_\gamma(t) = z^{(2)}_\gamma(t)$ are given respectively by*

$$F_{z^{(\ell)}_\gamma}(x) = 1-\Phi\left(\frac{x}{\sigma_\ell\sqrt{1-\eta_\ell^2}}\right)+\sqrt{2\pi}\eta_\ell\phi\left(\frac{x}{\sigma_\ell}\right)\Phi\left(\frac{\eta_\ell x}{\sigma_\ell\sqrt{1-\eta_\ell^2}}\right), \quad \ell = 1,2,$$

where $\sigma_1^2 = \frac{1}{4\sqrt{\pi}\xi^3}$, $\eta_1 = \frac{\sqrt{3}}{\sqrt{5}}$, $\sigma_2^2 = \frac{3}{8\sqrt{\pi}\xi^5}$ and $\eta_2 = \frac{\sqrt{5}}{\sqrt{7}}$.

The p-values in the proposed mSTEM algorithms can be calculated using Proposition 2. Additionally, given that $z_\gamma(t)$ is defined in (3.13), applying the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Kac-Rice formula, the expected number of local extrema in the smoothed Gaussian processes $z'_\gamma(t)$ and $z''_\gamma(t)$ in the unit interval can be calculated by

$$E[\tilde{m}_{z'_\gamma}(U(1))] = \frac{1}{\pi}\sqrt{\frac{\mathrm{Var}(z^{(3)}_\gamma(t))}{\mathrm{Var}(z''_\gamma(t))}} = \frac{\sqrt{10}}{2\pi\xi};$$
$$E[\tilde{m}_{z''_\gamma}(U(1))] = \frac{1}{\pi}\sqrt{\frac{\mathrm{Var}(z^{(4)}_\gamma(t))}{\mathrm{Var}(z^{(3)}_\gamma(t))}} = \frac{\sqrt{14}}{2\pi\xi}.$$

Remark 3. *In this example, Gaussian noise is assumed because it enables a closed formula for the peak height distribution and the expected number of local extrema. For non-Gaussian noise, the p-values associated with the peak heights can be computed via Monte Carlo simulation.*

3.5 SNR and optimal bandwidth γ

Intuitively, a higher SNR facilitates the detection of change points, as it enhances the contrast between signal and noise. On the other hand, SNR is not only associated with the size of the slope change and jump, but is also a function of bandwidth γ. Thus, it is important to study the SNR when choosing the best bandwidth. The following result, which is a direct consequence of Lemmas **??** and 3, provides definitions and properties for the SNR of both types of change points.

Lemma 4. *Suppose that $z_\gamma(t)$ is the smoothed noise defined in (3.13), we let $\nu = m\gamma$ ($m = 0$ when $z(t)$ is Gaussian white noise). For a Type* I *change point*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

v_j, the SNR *is calculated as*

$$\text{SNR}_{\text{I}}(v_j) := \frac{|\mu''_\gamma(v_j)|}{\sqrt{\text{Var}(z''_\gamma(v_j))}} = \frac{|\frac{k_{j+1}-k_j}{\sqrt{2\pi}\gamma}|}{\sqrt{\frac{3}{8\sqrt{\pi}\xi^5}}} = \frac{2\gamma^{3/2}}{\sqrt{3}\pi^{1/4}}(1+m^2)^{5/4}|k_{j+1}-k_j|.$$

For a Type II *change point v_j, it asymptotically becomes a local extremum at v_j under condition (C2), and the corresponding* SNR *is calculated as*

$$\text{SNR}_{\text{II}}(v_j) = C_\gamma \left[\frac{\sqrt{2}a_j}{\sqrt{\gamma}} + R_j\gamma^{3/2}\right], \tag{3.14}$$

where

$$C_\gamma = \pi^{1/4}(1+m^2)^{3/4},$$

$$R_j = (k_{j+1}-k_j) + 2(k_j+k_{j+1})(\Phi(c)-1).$$

Due to condition (C2) and $\Phi(c) \approx 1$, the SNR for both Type I and II change points is increasing in the sizes of slope change and jump, respectively. In addition, Lemma 4 shows that the SNR increases in γ for both Type I and Type II change points. Therefore, if the minimal distance between the neighboring change points d is large enough, a large γ is preferable (see Figure 4). However, for a fixed d, to avoid the overlap of two neighboring smoothed signal segments $(v_j - c\gamma, v_j + c\gamma)$ and $(v_{j+1} - c\gamma, v_{j+1} + c\gamma)$, γ should not be too large (such as $2c\gamma \leq d$).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

4. Simulation Study

4.1 Simulation settings

In this section, we will assess the performance of our method in various signal scenarios. The signals follow the form $\mu(t) = c_j + k_j t$, where $t = 1, \ldots, L$. True change points are defined as $v_j = jd$ for $j = 1, \ldots, \lfloor L/d \rfloor - 1$ with $d = 150$ representing the separation between adjacent change points. We consider four distinct signal scenarios with corresponding signal coefficients: (1). piecewise linear mean with continuous change points (Type I change points): $c_1 = k_1 = 0$ and $k_{j+1} - k_j = 0.1$ for $j \geq 1$; (2). piecewise constant mean with jumps (special case of Type II change points): $c_1 = k_1 = 0$ and $a_j = c_{j+1} - c_j = 10$ for $j \geq 1$; (3). piecewise linear mean with discontinuous change points (Type II change points): $c_1 = k_1 = 0$, $a_j = 10$, $k_{j+1} - k_j = 0.05$ for odd j and $k_{j+1} - k_j = -0.05$ for even j; (4). mixture of Type I and Type II change points: This scenario consists of the signal from scenario (1) followed by the signal from scenario (3). For the first three scenarios, let $L = 1500$ and thus $J = 9$ for each scenario. For the mixture case, we extend the signal length to $L = 3000$ with $J_1 = J_2 = 9$.

The noise is generated from a stationary ergodic Gaussian process with zero mean and $\nu = 1$. To control the False Discovery Rate (FDR), we implement

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

the BH procedure at a significant level of $\alpha = 0.05$. Our simulation results are based on the average of 1000 replications.

4.2 Performance of our method

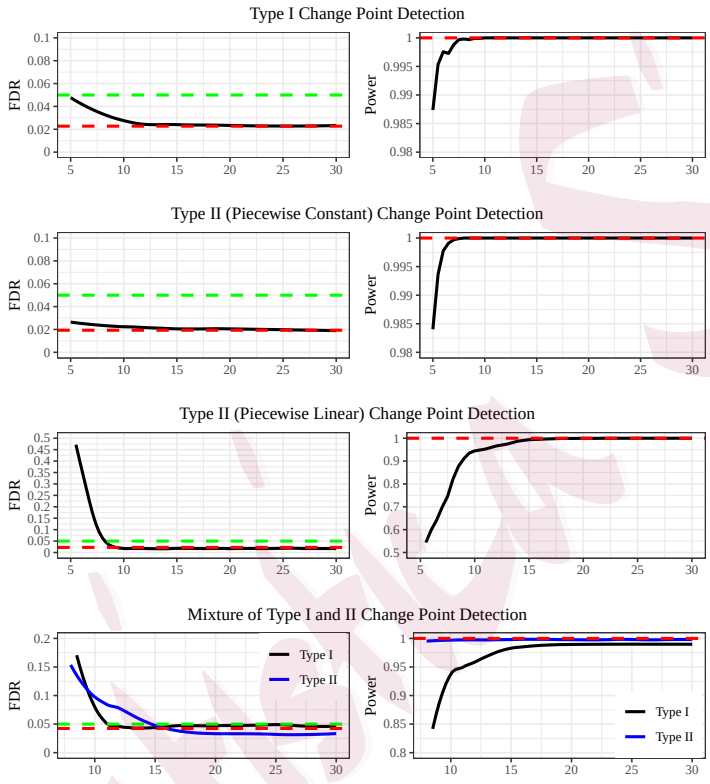

Figure 4: FDR (left panel) and power (right panel) versus SNR for detecting change points in various signal types. This example employs a kernel bandwidth of $\gamma = 10$ and a location tolerance of $b = 10$. The red dashed lines mark the asymptotic limits of FDR and power, while the green dashed lines indicate the significant level $\alpha = 0.05$.

In Section 3, we presented the theoretical results of the False Discovery Rate (FDR) and the power of our method. In this section, we validate these theoretical

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

properties through numerical studies.

Figure 4 shows the performance of our method in four different signal scenarios. Notably, as SNR increases, the FDR converges to its asymptotic limit. This limit remains consistently below the FDR control level of $\alpha = 0.05$. Meanwhile, the statistical power approaches 1. For a fixed bandwidth γ, a higher SNR implies a more substantial slope change or jump size, aligning with conditions (C1) and (C2), respectively. It is worth noting that while our theoretical analysis assumes an infinite SNR, our numerical results show that even with a moderate SNR, both the FDR and power have already converged to their asymptotic limits.

4.3 Comparison with other methods

To compare with other change point detection methods, we evaluated the precision and computational efficiency of change point detection. Short-term (see section 4.1) and long-term data sequences were considered. The extended long-term data sequence was generated by replicating the short-term data tenfold within our simulation, guaranteeing that the underlying patterns remained consistent between the two datasets.

We compared our method with three multiple testing-based change point detection methods: B&P (Bai 1997), NOT (Baranowski et al. 2019), and NSP (Fryzlewicz 2024). Note that we excluded the results of the B&P method from

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

the long-term data comparison due to the extremely long computing time. For the NSP method, the change points were estimated as confidence intervals, and we used the midpoints of these intervals as point estimates for comparison with other methods for the sake of consistency. Furthermore, the scenario of a mixture of change points of Type I and Type II was not considered because the other three methods cannot distinguish between change points of Type I and Type II.

Suppose that the number of true change points is J, and the number of detected change points is $\hat{J}$. For each detected change point $\hat{v}_i$, $i = 1, 2, \ldots, \hat{J}$, we define it as the estimate of the true change point v_j such that $j = \text{argmin}_{j=1,2,\ldots,J} |\hat{v}_i - v_j|$. Thus, we let $v_i^0 = v_j$ be the true change point corresponding to $\hat{v}_i$. Note that a single true change point may correspond to multiple estimated change points. The detection accuracy of $\hat{v}_i$ is measured based on the distance $|\hat{v}_i - v_i^0|$ for $i = 1, 2, \ldots, \hat{J}$, and the capture rate of an interval I is defined as $\sum_{i=1}^{\hat{J}} \mathbf{1}(|\hat{v}_i - v_i^0| \in I)/J$. A high capture rate may be achieved by detecting an excessive number of change points, which, in turn, leads to a high FDR. Therefore, a good change point detection method should achieve high detection accuracy while maintaining a low FDR.

In Table 1, the results for the short-term data illustrate the performance of our proposed method, NOT and NSP in terms of FDR control and power in all scenarios. Both NOT and our method demonstrate the capability to achieve high

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

accuracy in change point detection within a single bandwidth γ, and our method is the fastest among the algorithms considered. In Table 2, it is seen that for the long-term data sequence, NOT and NSP do not perform well with respect to FDR control, power, and detection accuracy. In contrast, our method consistently maintains excellent performance, retaining the advantage of the shortest computing time. Note that the better performance of our proposed method in terms of FDR and power with long-term data can be attributed to asymptotic conditions $L \to \infty$ and $J/L \to A > 0$. Additionally, the minimal computing time required by our method in both short-term and long-term data sequences demonstrates its superior computational complexity.

5. Data examples

In this section, we applied our method to real applications and performed a comparative analysis with NOT and NSP. We evaluated the performance of all three methods on the dataset of global temperature records.

5.1 Global temperature records

Climate change research plays a pivotal role in addressing the impacts of global warming and guiding policy decisions related to mitigation and adaptation strategies. Analyzing temperature changes is crucial for understanding the patterns and shifts in climate over time. In this paper, we study the data concerning global

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Table 1: Detection accuracy of change points for the short-term data sequence ($\gamma = b = 10$).

Signal Type	Method	Capture rate of interval					FDR	Power	Time (s)
		$[0, \frac{1}{3}\gamma)$	$[\frac{1}{3}\gamma, \gamma)$	$[\gamma, 2\gamma)$	$[2\gamma, 4\gamma)$	$\geq 4\gamma$			
Type I	mSTEM	0.8400	0.1533	0.0217	0.0267	0.0483	0.0125	0.9933	0.1370
	NOT	0.9883	0.0117	0.0000	0.0017	0.0000	0.0572	1.0000	1.3550
	NSP	0.6183	0.1900	0.1617	0.0300	0.0000	0.0458	0.9999	3.2417
	B&P	0.4700	0.5083	0.0217	0.0000	0.0000	0.1632	0.9783	87.4562
Type II Piecewise Constant	mSTEM	0.9617	0.0383	0.0000	0.0367	0.0333	0.0227	1.0000	0.0290
	NOT	0.9833	0.0183	0.0017	0.0000	0.0017	0.0558	1.0000	0.2863
	NSP	0.6767	0.3133	0.2117	0.0867	0.0100	0.0517	0.9001	1.9430
	B&P	0.4117	0.0233	0.0467	0.2050	0.0000	0.1260	0.4350	71.0924
Type II Piecewise Linear	mSTEM	0.9983	0.0017	0.0000	0.0067	0.0233	0.0348	1.0000	0.0839
	NOT	0.8833	0.0933	0.0233	0.0000	0.0000	0.0727	0.9766	0.4524
	NSP	0.6967	0.2417	0.0333	0.0283	0.0000	0.0626	0.9384	2.8013
	B&P	0.3333	0.0000	0.3015	0.3333	0.0745	0.1633	0.3333	68.3345
Type II Piecewise Linear	mSTEM	0.9983	0.0017	0.0000	0.0067	0.0233	0.0348	1.0000	0.0839
	NOT	0.8833	0.0933	0.0233	0.0000	0.0000	0.0727	0.9766	0.4524
	NSP	0.6967	0.2417	0.0333	0.0283	0.0000	0.0626	0.9384	2.8013

mean land-ocean temperature deviations (measured in degrees Celsius) from the period 1880 to 2015, in comparison to the average temperature between 1951 and 1980. This dataset is available at https://data.giss.nasa.gov/gistemp/graphs.

Figure 5 illustrates the results of applying all three methods to global temperature data. Our method successfully detected two Type II change points in

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Table 2: Detection accuracy of change points for the long-term data sequence ($\gamma = b = 10$).

Signal Type	Method†	Capture rate of interval					FDR	Power	Time (s)
		$[0, \frac{1}{3}\gamma)$	$[\frac{1}{3}\gamma, \gamma)$	$[\gamma, 2\gamma)$	$[2\gamma, 4\gamma)$	$\geq 4\gamma$			
Type I	mSTEM	0.7616	0.2241	0.0295	0.0244	0.0135	0.0127	0.9963	0.2469
	NOT	0.1358	0.1712	0.2475	0.5028	0.2369	0.0853	0.8732	112.5007
	NSP	0.3512	0.4607	0.1692	0.0086	0.0000	0.0792	0.8362	433.6193
Type II Piecewise Constant	mSTEM	0.9702	0.0000	0.000	0.0154	0.0161	0.01463	1.0000	0.1188
	NOT	0.8633	0.0037	0.007	0.0177	0.0090	0.0735	0.9164	48.4487
	NSP	0.6398	0.3202	0.030	0.0000	0.0000	0.08011	0.9228	213.0398
Type II Piecewise Linear	mSTEM	0.9899	0.0000	0.0000	0.0065	0.0133	0.0237	0.9992	1.0627
	NOT	0.8748	0.0002	0.0009	0.0032	0.0025	0.1217	0.8012	10.6064
	NSP	0.6059	0.3543	0.0298	0.0000	0.0000	0.1383	0.8255	400.6633

† B&P method was not included due to its substantial computational time requirements.

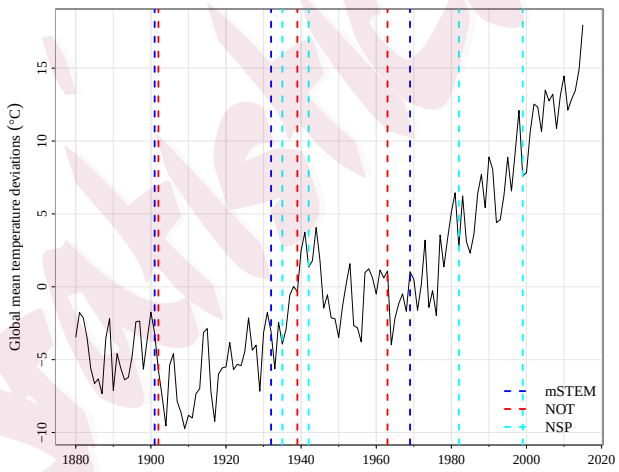

Figure 5: Change point detection of global mean land-ocean temperature deviations.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

1902 and 1934, along with one Type I change point in 1968. These findings closely align with the results presented in a previous study by Yu & Ruggieri (2019), where they conducted an analysis of five time series and identified three distinct periods of significant temperature change in the 20th century. Specifically, their research highlighted change points between 1902-1917, 1936-1945, and 1963-1976. The result of our detection method aligns with the visual assessment of the data sequence. The observed pattern suggests that the signal remained relatively constant during the first period (1880 - 1902), followed by a linear trend in the second period (1903 - 1934), then another constant signal in the third period (1935 - 1971), and finally a linear trend in the last period.

6. Discussion

6.1 Choice of smoothing bandwidth γ

The choice of the smoothing bandwidth γ is, in fact, an important aspect of change point detection. It plays a role in balancing the trade-off between the overlap of the smoothed signal supports and the accuracy of detecting change points. Thus, a small γ (if the noise is highly autocorrelated) or a relatively large γ (if the noise is less autocorrelated) is preferred to increase power, but only to the extent that the smoothed signal supports have little overlap. Considering the Gaussian kernel with support domain $\pm c\gamma$, an initial guideline to choose γ can

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

be set as $d/(2c)$, where d is the minimum distance between any two neighboring change points. This guideline ensures that the supports of the smoothed signals do not overlap excessively. In practice, a data-driven approximation of d is to use an initial estimator from some classic change point detection methods, such as NOT or NSP, and the resulting $\widehat{d}$ provides a practical anchor for bandwidth selection. A reasonable choice is then $\gamma \approx \widehat{d}/(2c)$, which helps prevent excessive overlap of smoothed signal supports.

6.2 Type II change points with moderate $|q_j|$

For a Type II change point v_j, we have studied in this paper the case of small $|q_j|$, and we showed in Section S1 of the Supplementary Materials the case of large $|q_j|$, which behaves similarly to a Type I change point. But when a Type II change point has a moderate $|q_j|$, for example, $|q_j|$ is close to c/γ, the first derivative $\mu'_\gamma(t)$ can produce a local extremum at $t = v_j \pm \gamma^2|q_j| \approx v_j \pm c\gamma$, which is near the endpoints of the smoothed signal region and outside of $v_j \pm b$. This may lead to false detections in the noise region while failing to detect v_j in the signal region, resulting in an overestimation of FDR and an underestimation of power. Addressing this limitation of detecting Type II change points with moderate $|q_j|$ would be an interesting question for future research.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

6.3 Nonstationary Gaussian noise

In this paper, the Gaussian noise is assumed to be stationary, which allows the error terms to be correlated. However, in real applications, it is common for the data to contain nonstationary Gaussian noise (Heinonen et al. 2016). Thus, it is valuable to detect change points under nonstationary Gaussian noise. To achieve this, we will study the peak height distribution for nonstationary Gaussian noise and show the FDR control and power consistency under certain assumptions, which would be promising but comes with specific challenges.

Supplementary Material

The online Supplementary Material contains the theoretical proofs, technical derivations, and additional numerical results that support the findings of this paper. Specifically, it provides the rigorous mathematical proofs for the main theorems and technical lemmas, including the asymptotic distributions of the test statistics and the error rate controls. It also outlines the detailed implementation of the proposed multiple testing procedure for structural break detection, alongside extended simulation studies that evaluate the finite-sample performance, power, and robustness of the methodology under various noise structures and signal intensities.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Acknowledgments

The authors thank Prof. Domenico Marinucci for motivating this research problem and Prof. Armin Schwartzman for helpful discussions and valuable suggestions.

Funding

Dan Cheng acknowledges support from the National Science Foundation grant DMS-2220523 and the Simons Foundation Collaboration Grant 854127.

References

Andrews, D. W. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica: Journal of the Econometric Society* pp. 821–856.

Bai, J. (1994), 'Least squares estimation of a shift in linear processes', *Journal of Time Series Analysis* **15**(5), 453–472.

Bai, J. (1997), 'Estimation of a change point in multiple regression models', *Review of Economics and Statistics* **79**(4), 551–563.

Bai, J. & Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Baranowski, R., Chen, Y. & Fryzlewicz, P. (2019), 'Narrowest-over-threshold detection of multiple change points and change-point-like features', *Journal of the Royal Statistical Society: Series B* **81**(3), 649–672.

Chang, P.-C., Fan, C.-Y. & Liu, C.-H. (2008), 'Integrating a piecewise linear representation method and a neural network model for stock trading points prediction', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **39**(1), 80–92.

Cheng, D., He, Z. & Schwartzman, A. (2020), 'Multiple testing of local extrema for detection of change points', *Electronic Journal of Statistics* **14**(2), 3705–3729.

Cheng, D. & Schwartzman, A. (2015), 'Distribution of the height of local maxima of gaussian random fields', *Extremes* **18**(2), 213–240.

Frick, K., Munk, A. & Sieling, H. (2014), 'Multiscale change point inference', *Journal of the Royal Statistical Society: Series B* **76**(3), 495–580.

Fryzlewicz, P. (2014), 'Wild binary segmentation for multiple change-point detection', *The Annals of Statistics* **42**(6), 2243–2281.

Fryzlewicz, P. (2024), 'Narrowest significance pursuit: inference for multiple

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

change-points in linear models', *Journal of the American Statistical Association* **119**(546), 1633–1646.

Hao, N., Niu, Y. S. & Zhang, H. (2013), 'Multiple change-point detection via a screening and ranking algorithm', *Statistica Sinica* **23**(4), 1553.

Heinonen, M., Mannerström, H., Rousu, J., Kaski, S. & Lähdesmäki, H. (2016), Non-stationary gaussian process regression with hamiltonian monte carlo, *in* 'Artificial Intelligence and Statistics', PMLR, pp. 732–740.

Huber, P. J. (2004), *Robust statistics*, Vol. 523, John Wiley & Sons.

Hyun, S., Lin, K. Z., G'Sell, M. & Tibshirani, R. J. (2021), 'Post-selection inference for changepoint detection algorithms with application to copy number variation data', *Biometrics* **77**(3), 1037–1049.

Khan, F., Ghaffar, A., Khan, N. & Cho, S. H. (2020), 'An overview of signal processing techniques for remote health monitoring using impulse radio uwb transceiver', *Sensors* **20**(9), 2479.

Lavielle, M. (2005), 'Using penalized contrasts for the change-point problem', *Signal processing* **85**(8), 1501–1510.

Lerman, P. (1980), 'Fitting segmented regression models by grid search', *Journal of the Royal Statistical Society Series C: Applied Statistics* **29**(1), 77–84.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Li, H., Munk, A. & Sieling, H. (2016), 'Fdr-control in multiscale change-point segmentation', *Electronic Journal of Statistics* **10**(1), 918–959.

Muggeo, V. M. (2003), 'Estimating regression models with unknown break-points', *Statistics in medicine* **22**(19), 3055–3071.

Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. (2004), 'Circular binary segmentation for the analysis of array-based dna copy number data', *Biostatistics* **5**(4), 557–572.

Perron, P. (1989), 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica: journal of the Econometric Society* pp. 1361–1401.

Schwartzman, A., Gavrilov, Y. & Adler, R. J. (2011), 'Multiple testing of local maxima for detection of peaks in 1d', *Annals of statistics* **39**(6), 3290.

Tebaldi, C. & Lobell, D. (2008), 'Towards probabilistic projections of climate change impacts on global crop yields', *Geophysical Research Letters* **35**(8).

Vostrikova, L. Y. (1981), Detecting "disorder" in multidimensional random processes, *in* 'Doklady Akademii Nauk', Vol. 259, Russian Academy of Sciences, pp. 270–274.

Yao, Y.-C. & Au, S.-T. (1989), 'Least-squares estimation of a step function', *Sankhyā: The Indian Journal of Statistics, Series A* pp. 370–381.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0113

Yu, M. & Ruggieri, E. (2019), 'Change point analysis of global temperature records', *International Journal of Climatology* **39**(8), 3679–3688.

Zhibing He

School of Statistics and Data Science, Southeast University, China

E-mail: zhibing.he@seu.edu.cn

Dan Cheng

School of Mathematical and Statistical Sciences, Arizona State University, USA

E-mail: chengdan@asu.edu

Yunpeng Zhao

Department of Statistics, Colorado State University, USA

E-mail: yunpeng.zhao@colostate.edu