Multiple testing of local extrema for detection of change points

Dan Cheng[*]**, Zhibing He**[*]

School of Mathematical and Statistical Sciences
Arizona State University
Tempe, AZ 85281, U.S.A.
e-mail: cheng.stats@gmail.com*;* zhibingh@asu.edu

and

Armin Schwartzman[†]

Division of Biostatistics and Bioinformatics and Halicioğlu Data Science Institute
University of California San Diego
La Jolla, CA 92093, U.S.A.
e-mail: armins@ucsd.edu

Abstract: A new approach to detect change points based on differential smoothing and multiple testing is presented for long data sequences modeled as piecewise constant functions plus stationary ergodic Gaussian noise. As an application of the STEM algorithm for peak detection developed in Schwartzman et al. [27] and Cheng and Schwartzman [5], the method detects change points as significant local maxima and minima after smoothing and differentiating the observed sequence. The algorithm, combined with the Benjamini-Hochberg procedure for thresholding p-values, provides asymptotic strong control of the False Discovery Rate (FDR) and power consistency, as the length of the sequence and the size of the jumps get large. Simulations show that FDR levels are maintained in non-asymptotic conditions and guide the choice of smoothing bandwidth. The methods are illustrated in magnetometer sensor data and genomic array-CGH data. An R package named "dSTEM" is available in R Cran.

Contents

1. Introduction

Detecting change points in the mean of an observed signal is a common statistical problem with applications in many research areas such as climatology [26], oceanography [17], finance [32] and medical imaging [23]. It often appears in the analysis of time series but it has more recently been found in the analysis of genomic sequences, see [10, 18, 22, 24, 30, 31] and the references therein. Given the large amounts of data present in modern applications, it is of interest to design a change point detection method that can operate over long sequences where the number and location of change points are unknown, and in such a way that the overall detection error rate is controlled.

Many different approaches have been proposed to find and estimate change points, such as kernel-based methods [1], Bayesian methods [3, 10], segmentation techniques [24, 31, 22], nonparametric tests [19] and L_1-penalty methods [9, 13, 30], including the PELT method [15, 17]. Though there is abundant literature on change points segmentation and detection, only a few papers address the FDR issue which treats the detection of change points as multiple hypothesis testing problems. Tibshirani and Wang [30] applied the fused lasso method to the hot-spot detection problem, and provided empirical evidence for the FDR control. Efron and Zhang [8] introduced an iterative local FDR based algorithm to explore copy number changes. Recently, Frick et al. [11] introduced a simultaneous multiscale change point estimator (SMUCE) for the change point

problem in exponential family regression, and proved the control of the probability of overestimating the true number of change points. Li et al. [20] improved the SMUCE method and proposed a multiscale segmentation method FDRSeg, which gives a non-asymptotic upper bound for its FDR in a Gaussian setting and is robust to the choice of parameter α. However, our proposed new approach is unique in the following two ways.

First, the noise is assumed to be a stationary Gaussian process, allowing the error terms to be correlated. This is an important departure from the standard assumption of white noise in most of the change-point literature. In fact, applied statisticians desiring to use change-point methods have sometimes abandoned this option in favor of other techniques simply because the white noise assumption does not hold [14]. This paper shows that change-point methods can be devised for correlated noise, expanding the domain of their applicability.

Second, we use the theory of Gaussian processes to compute p-values for all candidate change points, so that significant change points can be selected at a desired significance level. For concreteness, we adopt the Benjamini-Hochberg multiple testing procedure (abbreviated as *BH procedure* through this paper), enabling control of the false discovery rate (FDR) of detected change points when the data sequence is long and the number and location of change points are unknown. To the authors' knowledge, our work is the first article proposing a multiple testing method for controlling the FDR of detected change points. Moreover, the asymptotic properties of FDR and power are provided.

In this paper, we consider a signal-plus-noise model where the true signal is a piecewise constant function and the change points are defined as the points of discontinuity. Inspired by the method for detecting peaks in [27] and [5], we modify the STEM algorithm therein to detect change points. The central idea is the observation that the true signal has zero derivative everywhere except at the change points, where the derivative is infinite. Thus, in the presence of noise and under temporal or spatial sampling, change points can be seen as positive or negative peaks in the derivative of the smoothed signal. Note that because of the time sampling, derivatives cannot be observed directly and can only be estimated. The focus on the derivative of the smoothed signal effectively transforms the change point detection problem into a peak detection problem. As in the STEM algorithm, the resulting peak detection problem is then solved by identifying local maxima and minima of the derivative as candidate peaks and applying a multiple testing procedure to the list of candidates.

The "differential Smoothing and TEsting of Maxima/Minima" (dSTEM) algorithm for change point detection introduced in the present work is illustrated by a toy example in Figure 1 and consists of the following steps:

1. *Differential kernel smoothing*: to transform change points to local maxima or minima, and to increase the SNR. The principle of this step is illustrated in Figure 2.
2. *Candidate peaks*: find local maxima and minima of the differentiated smoothed process.

3. *P-values*: computed at each local maximum and minimum under the the null hypothesis of no signal in a local neighborhood.
4. *Multiple testing*: apply a multiple testing procedure to the set of local maxima and minima; declare as detected change points those local maxima and minima whose p-values are significant.
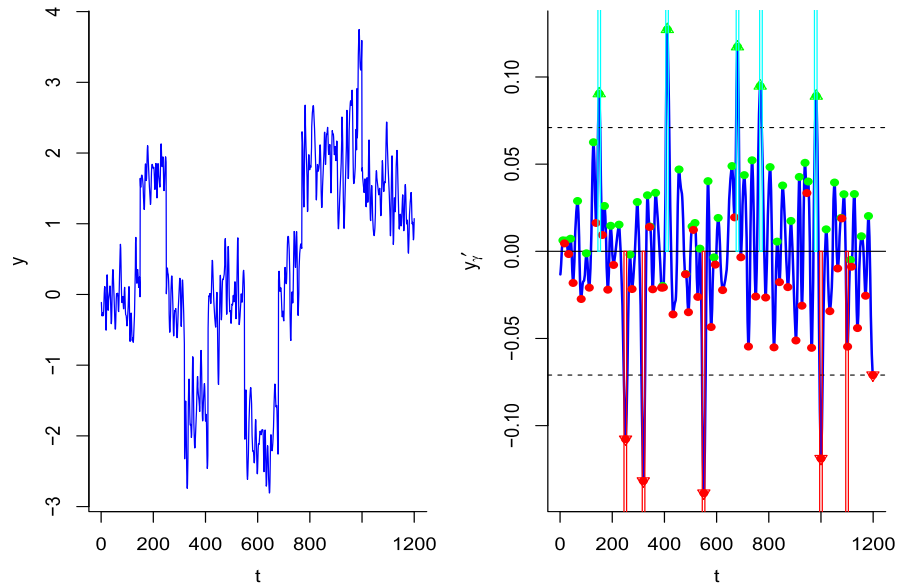

Fig 1. *Following the notation in §2.1 and Example 3.1, the left panel is the observed signal-plus-noise model $y(t)$ containing ten true change points with varying a_i and noise $z(t)$ given by (3.5) with $\sigma = 1$ and $\nu = 2$. The right panel illustrates the dSTEM algorithm. The blue curve is $y'_\gamma(t)$, obtained with a Gaussian smoothing kernel with standard deviation $\gamma = 6$. Local maxima (green solid dots) and local minima (red solid dots) are declared as significant (marked with solid triangles) at FDR level $\alpha = 0.1$ if their heights are beyond the dotted line thresholds. The cyan and pink bars indicate the location tolerance intervals $(v_i - b, v_i + b)$ with $b = 5$ for increasing and decreasing change points respectively. At this tolerance, there are nine true discoveries and one false discovery.*

The proposed dSTEM algorithm above differs from the algorithms in [27] and [5] in that peaks are sought in the derivative of the smoothed signal rather than the smoothed signal itself, and that both positive and negative peaks are considered. In addition, an important consideration for the proper definition of error in change point detection is that, as opposed to the peak detection problems considered in [27] and [5] where signal peaks had compact support, a true single change point over a continuous domain at $t = v$ has Lebesgue measure zero. Thus in the presence of noise, it can hardly be detected exactly at $t = v$. Therefore we introduce a *location tolerance b* that defines the precision within which a change point should be detected. Specifically, given b, a detected change point is regarded as a true discovery if it falls in the interval $(v - b, v + b)$. Conversely, if a significant change point is found more than a distance b
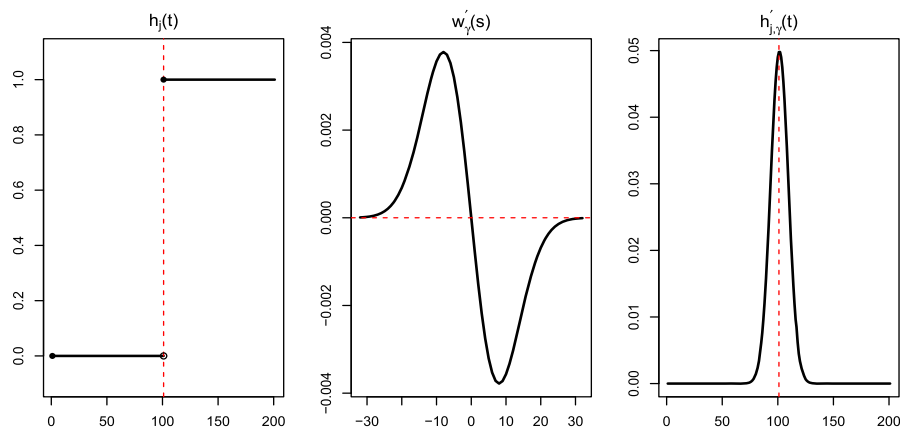
FIG 2. *Following the notation in §2.1 and §2.2, the left panel is the change point indicator function $h_j(t)$ with $v_j = 101$, that is, there exists only one change point at the location $t = v_j = 101$. The middle panel is the differential Gaussian kernel $w'_\gamma(s)$ on $[-4\gamma, 4\gamma]$ with $\gamma = 8$. The right panel is $h'_{j,\gamma}(t)$, which is the differential kernel smoothing (under Gaussian kernel) of $h_j(t)$ obtained by convolution, as shown in (2.7). By such transformation, the change point becomes a local maxima of a smooth function with compact support.*

from any true change point, it is considered a false discovery. The quantity b is not used in the dSTEM algorithm itself but is needed for proper error definition.

Technically, the main contributions of our results include the following: (1) we use the novel idea of employing kernel smoothing and differentiation to transform change points into peaks; (2) we combine local maxima and local minimum together as candidate peaks to compute p-values and perform the multiple hypothesis testing, improving the peak detection techniques in our previous work [5]; (3) we introduce the concept of location tolerance parameter for new definitions of FDR and power, which are more applicable in general models compared to those in [5]; (4) by extending the techniques of the proofs in [27] and [5], we show that the dSTEM algorithm exhibits asymptotic FDR control and power consistency as the length of the sequence and the size of the jumps at the change points increase.

Simulations for varying levels of smoothing bandwidth γ, smoothing degree of noise ν and jump size a are used to study the behavior of the algorithm under non-asymptotic conditions. The simulation results help guide the choice of smoothing bandwidth γ with respect to ν and the desired location tolerance. In general, power increases with bandwidth to a limit dictated by the distance between the change points, so admitting a higher tolerance generally allows a higher bandwidth and higher power.

The methods are illustrated in a genomic sequence of array-CGH data in a breast-cancer tissue sample [21, 12]. The goal of the analysis is to find genomic segments with copy-number alterations. These are found by detecting change

points in the copy number genomic sequence. Another application is magnetometer sensor readings, aiming at finding the start and end points of hand gesture motion, which is the critical step and foundation of establishing a secret key based on hand gestures.

2. The multiple testing scheme

2.1. The model

We consider a continuous time model, although the algorithm is designed for data discretely sampled in time. Consider the signal-plus-noise model

$$y(t) = \mu(t) + z(t), \qquad t \in \mathbb{R}, \tag{2.1}$$

where the signal $\mu(t)$ is a piecewise constant function of the form

$$\mu(t) = \sum_{j=0}^{\infty} a_j h_j(t), \qquad a_j \in \mathbb{R} \setminus \{0\},$$

with $h_j(t) = \mathbb{1}(t \geq v_j)$ for $v_j \in \mathbb{R}$. We are interested in finding the *change points* v_j. For the asymptotic analysis, we assume

$$a = \inf_j |a_j| > 0 \quad \text{and} \quad d = \inf_j |v_j - v_{j-1}| > 0, \tag{2.2}$$

so that the change points do not become arbitrarily small in size nor arbitrarily close to each other.

Let $w_\gamma(t) = w(t/\gamma)/\gamma$, where $\gamma > 0$ is the bandwidth parameter and $w(t) \geq 0$ is a unimodal symmetric kernel with compact connected support $[-c, c]$ and unit action. Convolving the process (2.1) with the kernel $w_\gamma(t)$ results in the smoothed random process

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

where the smoothed signal and smoothed noise are defined respectively as

$$\mu_\gamma(t) = w_\gamma(t) * \mu(t) = \sum_{j=0}^{\infty} a_j h_{j,\gamma}(t) \quad \text{and} \quad z_\gamma(t) = w_\gamma(t) * z(t), \tag{2.4}$$

and where the smoothed change point takes the form

$$h_{j,\gamma}(t) = w_\gamma(t) * h_j(t). \tag{2.5}$$

The smoothed noise $z_\gamma(t)$ defined by (2.4) is assumed to be a zero-mean four-times differentiable stationary ergodic Gaussian process.

2.2. Change point detection as peak detection of the derivative

Consider now the derivative of the smoothed observed process (2.3)

$$y'_\gamma(t) = w'_\gamma(t) * y(t) = \int_{\mathbb{R}^N} w'_\gamma(t-s)y(s)\,ds = \mu'_\gamma(t) + z'_\gamma(t), \qquad (2.6)$$

where the derivatives of the smoothed signal and smoothed noise are respectively

$$\mu'_\gamma(t) = w'_\gamma(t) * \mu(t) = \sum_{j=0}^{\infty} a_j h'_{j,\gamma}(t) \quad \text{and} \quad z'_\gamma(t) = w'_\gamma(t) * z(t).$$

A key observation from (2.5) is that

$$
\begin{aligned}
h'_{j,\gamma}(t) &= \int_{\mathbb{R}} w'_\gamma(t-s)h_j(s)\,ds = \int_{\mathbb{R}} w'_\gamma(s)h_j(t-s)\,ds \\
&= \int_{\mathbb{R}} w'_\gamma(s)\mathbb{1}(t-s \geq v_j)\,ds = \int_{-\infty}^{t-v_j} w'_\gamma(s)\,ds = w_\gamma(t-v_j),
\end{aligned}
\qquad (2.7)
$$

as illustrated in Figure 2. Thus (2.6) represents a signal-plus-noise model where the smoothed signal

$$\mu'_\gamma(t) = \sum_{j=0}^{\infty} a_j h'_{j,\gamma}(t) = \sum_{j=0}^{\infty} a_j w_\gamma(t-v_j) \qquad (2.8)$$

is a sequence of unimodal peaks with the same shape as that of w_γ and located at locations v_j. The problem of finding change points in $y_\gamma(t)$ is thus reduced to finding (positive or negative) peaks in $y'_\gamma(t)$.

For simplicity, the theoretical results below assume that the compact supports $S_{j,\gamma}$ of the smoothed peak shape $h'_{j,\gamma}(t) = w_\gamma(t-v_j)$ do not overlap. In practice, we have found in simulations that our method still works well when the smoothed supports $S_{j,\gamma}$ have certain small overlaps. However, when the overlaps are too large, they will affect the detection results. This is explored via simulations in Section 4.

2.3. The dSTEM algorithm for change point detection

Suppose we observe $y(t)$ with J_L jumps defined by (2.1) in the line of length L centered at the origin, denoted by $U(L) = (-L/2, L/2)$. The following dSTEM (differential Smoothing and TEsting of Maxima/Minima) is a modified version of the STEM algorithm of [27] and [5] for detecting change points.

Algorithm 2.1 (dSTEM algorithm for change point detection)**.**

　　1. Differential kernel smoothing*: Obtain the process* (2.6) *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 $\tilde{T}_\gamma = \tilde{T}_\gamma^+ \cup \tilde{T}_\gamma^-$, where*

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

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

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

where $\mu(s+) = \lim_{x \to s+} \mu(x)$ and $\mu(s-) = \lim_{x \to s-} \mu(x)$ are the right limit and left limit of μ at s respectively; and for each $t \in \tilde{T}_\gamma^-$, compute the p-value $p_\gamma(t)$ for testing the (conditional) hypotheses

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

where $b > 0$ is an appropriate location tolerance.
4. Multiple testing: *Let $\tilde{m}_\gamma = \#\{ t \in \tilde{T}_\gamma \}$ be the number of tested hypotheses. Apply a multiple testing procedure on the set of $\tilde{m}_\gamma$ p-values $\{ p_\gamma(t), t \in \tilde{T}_\gamma \}$, and declare significant all local extrema whose p-values are smaller than the significance threshold.*

2.4. P-values

Given the observed heights $y'_\gamma(t)$ at the local maxima or minima $t \in \tilde{T}_\gamma = \tilde{T}_\gamma^+ \cup \tilde{T}_\gamma^-$, p-values in step (3) of Algorithm 2.1 are computed as

$$p_\gamma(t) = \begin{cases} F_\gamma(y'_\gamma(t)), & t \in \tilde{T}_\gamma^+, \\ F_\gamma(-y'_\gamma(t)), & t \in \tilde{T}_\gamma^-, \end{cases} \qquad (2.9)$$

where $F_\gamma(u)$ denotes the right tail probability of $z'_\gamma(t)$ at the local maximum $t \in \tilde{T}_\gamma^+$, evaluated under the null model $\mu'(s) = 0, \forall s \in (t - b, t + b)$, that is,

$$F_\gamma(u) = \mathrm{P}\left(z'_\gamma(t) > u \,\middle|\, t \text{ is a local maximum of } z'_\gamma(t) \right). \qquad (2.10)$$

The second line in (2.9) is obtained by noting that, by (2.10),

$$\mathrm{P}\left(z'_\gamma(t) < u \,\middle|\, t \text{ is a local minimum of } z'_\gamma(t) \right)$$
$$= \mathrm{P}\left(-z'_\gamma(t) > -u \,\middle|\, t \text{ is a local maximum of } -z'_\gamma(t) \right) = F_\gamma(-u),$$

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

The distribution (2.10) has a closed-form expression, which can be obtained as in [27, 4, 6] or [7]. More specifically, the distribution (2.10) is given by

$$F_\gamma(u) = 1 - \Phi\left(u\sqrt{\frac{\lambda_{6,\gamma}}{\Delta}}\right) + \sqrt{\frac{2\pi\lambda_{4,\gamma}^2}{\lambda_{6,\gamma}\sigma_\gamma'^2}}\phi\left(\frac{u}{\sigma_\gamma'}\right)\Phi\left(u\sqrt{\frac{\lambda_{4,\gamma}^2}{\Delta\sigma_\gamma'^2}}\right), \quad (2.11)$$

where $\sigma_\gamma'^2 = \mathrm{Var}(z_\gamma'(t))$, $\lambda_{4,\gamma} = \mathrm{Var}(z_\gamma''(t))$, $\lambda_{6,\gamma} = \mathrm{Var}(z_\gamma'''(t))$, $\Delta = \sigma_\gamma'^2\lambda_{6,\gamma} - \lambda_{4,\gamma}^2$, and $\phi(x)$, $\Phi(x)$ are the standard normal density and cumulative distribution function, respectively. The quantities $\sigma_\gamma'^2$, $\lambda_{4,\gamma}$ and $\lambda_{6,\gamma}$ depend on the kernel $w_\gamma(t)$ and the autocorrelation function of the original noise process $z(t)$. Explicit expressions may be obtained, for instance, for the Gaussian autocorrelation model in Example 3.1 below, which we use later in the simulations.

2.5. Error definitions

Assuming the model of §2.1, define the *signal region* $\mathbb{S}_1^b = \cup_{j=1}^{J_L}(v_j - b, v_j + b)$ and *null region* $\mathbb{S}_0^b = U(L) \setminus \mathbb{S}_1^b$. For $u > 0$, let $\tilde{T}_\gamma(u) = \tilde{T}_\gamma^+(u) \cup \tilde{T}_\gamma^-(u)$, where

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

indicating that $\tilde{T}_\gamma^+(u)$ and $\tilde{T}_\gamma^-(u)$ are respectively the set of local maxima of $y_\gamma'(t)$ above u and the set of local minima of $y_\gamma'(t)$ below $-u$. The number of totally and falsely detected change points at threshold u are defined respectively as

$$R_\gamma(u) = \#\{t \in \tilde{T}_\gamma^+(u)\} + \#\{t \in \tilde{T}_\gamma^-(u)\},$$
$$V_\gamma(u;b) = \#\{t \in \tilde{T}_\gamma^+(u) \cap \mathbb{S}_0^b\} + \#\{t \in \tilde{T}_\gamma^-(u) \cap \mathbb{S}_0^b\}. \quad (2.12)$$

Both are defined as zero if $\tilde{T}_\gamma(u)$ is empty. The FDR at threshold u is defined as the expected proportion of falsely detected jumps

$$\mathrm{FDR}_\gamma(u;b) = \mathrm{E}\left\{\frac{V_\gamma(u;b)}{R_\gamma(u) \vee 1}\right\}. \quad (2.13)$$

Note that when γ and u are fixed, $V_\gamma(u;b)$ and hence $\mathrm{FDR}_\gamma(u;b)$ are decreasing in b.

Following the notation in [5], define the *smoothed signal region* $\mathbb{S}_{1,\gamma}$ to be the support of $\mu_\gamma'(t)$ and *smoothed null region* $\mathbb{S}_{0,\gamma} = U(L) \setminus \mathbb{S}_{1,\gamma}$. We call the difference between the expanded signal support due to smoothing and the true signal support the *transition region* $\mathbb{T}_\gamma = \mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1^b = \mathbb{S}_0^b \setminus \mathbb{S}_{0,\gamma}$.

2.6. Power

Denote by I^+ and I^- the collections of indices j corresponding to increasing and decreasing change points v_j, respectively. We define the power of Algorithm 2.1

as the expected fraction of true discovered change points

$$\text{Power}_\gamma(u; b) = \frac{1}{J_L} \sum_{j=1}^{J_L} \text{Power}_{j,\gamma}(u; b)$$

$$= \text{E}\left[\frac{1}{J_L} \left(\sum_{j \in I^+} \mathbb{1}\left(\tilde{T}_\gamma^+(u) \cap (v_j - b, v_j + b) \neq \emptyset \right) \right. \right. \tag{2.14}$$

$$\left. \left. + \sum_{j \in I^-} \mathbb{1}\left(\tilde{T}_\gamma^-(u) \cap (v_j - b, v_j + b) \neq \emptyset \right) \right) \right],$$

where $\text{Power}_{j,\gamma}(u; b)$ is the probability of detecting jump v_j within a distance b,

$$\text{Power}_{j,\gamma}(u; b) = \begin{cases} \text{P}\left(\tilde{T}_\gamma^+(u) \cap (v_j - b, v_j + b) \neq \emptyset \right), & \text{if } j \in I^+, \\ \text{P}\left(\tilde{T}_\gamma^-(u) \cap (v_j - b, v_j + b) \neq \emptyset \right), & \text{if } j \in I^-. \end{cases} \tag{2.15}$$

The indicator function in (2.14) ensures that only one significant local extremum is counted within a distance b of a change point, so power is not inflated. Note that when γ and u are fixed, $\text{Power}_\gamma(u; b)$ and $\text{Power}_{j,\gamma}(u; b)$ are increasing in b.

3. Asymptotic FDR control and power consistency

Suppose the BH procedure is applied in step 4 of Algorithm 2.1 as follows. For a fixed $\alpha \in (0, 1)$, let k be the largest index for which the ith smallest p-value is less than $i\alpha/\tilde{m}_\gamma$. Then the null hypothesis $\mathcal{H}_0(t)$ at $t \in \tilde{T}_\gamma$ is rejected if

$$p_\gamma(t) < \frac{k\alpha}{\tilde{m}_\gamma} \iff \begin{cases} y_\gamma'(t) > \tilde{u}_{\text{BH}} = F_\gamma^{-1}\left(\frac{k\alpha}{\tilde{m}_\gamma} \right) & \text{if } t \in \tilde{T}_\gamma^+, \\ y_\gamma'(t) < -\tilde{u}_{\text{BH}} = -F_\gamma^{-1}\left(\frac{k\alpha}{\tilde{m}_\gamma} \right) & \text{if } t \in \tilde{T}_\gamma^-, \end{cases} \tag{3.1}$$

where $k\alpha/\tilde{m}_\gamma$ is defined as 1 if $\tilde{m}_\gamma = 0$. Since $\tilde{u}_{\text{BH}}$ is random, we define FDR in such BH procedure as

$$\text{FDR}_{\text{BH},\gamma}(b) = \text{E}\left\{ \frac{V_\gamma(\tilde{u}_{\text{BH}}; b)}{R_\gamma(\tilde{u}_{\text{BH}}) \vee 1} \right\},$$

where $R_\gamma(\cdot)$ and $V_\gamma(\cdot; b)$ are defined in (2.12) and the expectation is taken over all possible realizations of the random threshold $\tilde{u}_{\text{BH}}$. We will make use of the following conditions:

(C1) The assumptions of §2.1 hold.
(C2) $L \to \infty$ and $a = \inf_j |a_j| \to \infty$, such that $(\log L)/a^2 \to 0$, $J_L/L = A_1 + O(a^{-2} + L^{-1/2})$ with $A_1 > 0$.

In condition (C2), we assume that the length of the search space L increases. In order for the detection procedure to have good power while the error is

controlled, the signal strength a should also increase. This assumption is not restrictive since $(\log L)/a^2 \to 0$ implies the search space may grow exponentially faster than the signal strength. These conditions are realistic in applications. In data with repeated observations and sample size n, the signal-to-noise ratio (SNR), equivalent to our signal strength a here, is proportional to $\sqrt{n}$, which is large when the sample size is large in the classical asymptotic sense. Since L can be treated as the dimensionality of the problem, adopting the traditional notation $p = L$, the condition $(\log L)/a^2 \to 0$ becomes $(\log p)/n \to 0$, which is similar to the condition required for consistent model selection in high dimensional regression with p covariates.

Let $\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))]$ and $\mathrm{E}[\tilde{m}_{0,\gamma}(U(1), u)]$ be the expected number of local maxima and local maxima above level u of $z'_\gamma(t)$ on the unit interval $U(1) = (-1/2, 1/2)$, respectively. In particular, applying the Kac-Rice formula, we have the following explicit result [27],

$$\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))] = \frac{1}{2\pi}\sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}}. \tag{3.2}$$

Note that, by symmetry, the expected number of local minina below level u of $z'_\gamma(t)$ on the unit interval $U(1)$ is given by $\mathrm{E}[\tilde{m}_{0,\gamma}(U(1), -u)]$.

Theorem 3.1. *Let conditions (C1) and (C2) hold.*

(i) Suppose Algorithm 2.1 is applied with a fixed threshold $u > 0$. Then

$$\mathrm{FDR}_\gamma(u; b) \le \frac{2\mathrm{E}[\tilde{m}_{0,\gamma}(U(1), u)](1 - 2c\gamma A_1)}{2\mathrm{E}[\tilde{m}_{0,\gamma}(U(1), u)](1 - 2c\gamma A_1) + A_1} + O(a^{-2} + L^{-1/2}).$$

(ii) Suppose Algorithm 2.1 is applied with the random threshold $\tilde{u}_{\mathrm{BH}}$ (3.1). Then

$$\mathrm{FDR}_{\mathrm{BH},\gamma}(b) \le \alpha \frac{2\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))](1 - 2c\gamma A_1)}{2\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))](1 - 2c\gamma A_1) + A_1} + O(a^{-1} + L^{-1/4}).$$

Proof. Since $w_\gamma(t)$ has compact support $[-c\gamma, c\gamma]$, by (2.7), the support $\mathbb{S}_{1,\gamma}$ of $\mu'_\gamma(t)$ in (2.8) is composed of the support segments $[v_j - c\gamma, v_j + c\gamma]$ of $h'_{j,\gamma}(t)$. By condition (C2), $|\mathbb{S}_{1,\gamma}|/L = 2c\gamma A_1 + O(a^{-2} + L^{-1/2})$, which implies $|\mathbb{S}_{0,\gamma}|/L = 1 - 2c\gamma A_1 + O(a^{-2} + L^{-1/2})$.

Notice that, on the null region $\mathbb{S}_{0,\gamma}$, the expected number of local extrema, including both local maxima and minima, equals $2|\mathbb{S}_{0,\gamma}|\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))]$. On the other hand, following the proof of Theorem 3 in [5], the expected number of local extrema on the signal region $\mathbb{S}_{1,\gamma}$ is asymptotically equivalent to J_L. This is because, for each $j \in I^+$ and $b > 0$, as $a \to \infty$, asymptotically, there is no local maximum of $y'_\gamma(t)$ in $(v_j - c\gamma, v_j - b) \cup (v_j + b, v_j + c\gamma)$, and there is only one local maximum of $y'_\gamma(t)$ in $(v_j - b, v_j + b)$. The reasoning for the case of minima is similar.

The result then follows from the same arguments for proving Theorem 3 in [5] with $N = 1$, $A_{2,\gamma} = 2c\gamma A_1$, $z_\gamma(t)$ replaced by $z'_\gamma(t)$ and $\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))]$ replaced by $2\mathrm{E}[\tilde{m}_{0,\gamma}(U(1))]$. $\quad\square$

Lemma 3.2. *Let conditions (C1) and (C2) hold. As $|a_j| \to \infty$, the power for peak j and fixed u (2.15) can be approximated by*

$$\text{Power}_{j,\gamma}(u;b) = \Phi\left(\frac{|a_j|w_\gamma(0) - u}{\sigma'_\gamma}\right)(1 + O(|a_j|^{-2})). \tag{3.3}$$

Proof. By (2.7), $h'_{j,\gamma}(v_j) = w_\gamma(0)$ is the maximum of $h'_{j,\gamma}(t)$ over $t \in \mathbb{R}$. The result then follows from Lemma 4 in [5] with $z_\gamma(t)$ replaced by $z'_\gamma(t)$. □

By similar arguments in [5] (see equation (20) therein), one can show that the random threshold $\tilde{u}_{\text{BH}}$ converges asymptotically to the deterministic threshold

$$u^*_{\text{BH}} = F_\gamma^{-1}\left(\frac{\alpha A_1}{A_1 + 2\text{E}[\tilde{m}_{0,\gamma}(U(1))](1 - 2c\gamma A_1)(1 - \alpha)}\right), \tag{3.4}$$

where $\text{E}[\tilde{m}_{0,\gamma}(U(1))]$ is given by (3.2). Since $\tilde{u}_{\text{BH}}$ is random, similarly to the definition of $\text{FDR}_{\text{BH},\gamma}(b)$, we define power in the BH procedure as

$$\text{Power}_{\text{BH},\gamma}(b) = \text{E}\left[\frac{1}{J_L}\left(\sum_{j \in I^+} \mathbb{1}\left(\tilde{T}_\gamma^+(\tilde{u}_{\text{BH}}) \cap (v_j - b, v_j + b) \neq \emptyset\right)\right.\right.$$

$$\left.\left. + \sum_{j \in I^-} \mathbb{1}\left(\tilde{T}_\gamma^-(\tilde{u}_{\text{BH}}) \cap (v_j - b, v_j + b) \neq \emptyset\right)\right)\right].$$

Theorem 3.3. *Let conditions (C1) and (C2) hold.*
(i) Suppose Algorithm 2.1 is applied with a fixed threshold $u > 0$. Then

$$\text{Power}_\gamma(u;b) = 1 - O(a^{-2}).$$

(ii) Suppose Algorithm 2.1 is applied with the random threshold $\tilde{u}_{\text{BH}}$ (3.1). Then

$$\text{Power}_{\text{BH},\gamma}(b) = 1 - O(a^{-2} + L^{-1/2}).$$

Proof. The desired results follow from similar arguments for showing Theorem 5 in [5]. □

Example 3.1 (Gaussian autocorrelation model). *Let the noise $z(t)$ in (2.1) be constructed as*

$$z(t) = \sigma \int_{\mathbb{R}} \frac{1}{\nu}\phi\left(\frac{t-s}{\nu}\right) dB(s), \qquad \sigma, \nu > 0, \tag{3.5}$$

where ϕ is the standard Gaussian density, $dB(s)$ is Gaussian white noise and $\nu > 0$ ($z(t)$ is regarded by convention as Gaussian white noise when $\nu = 0$). Convolving with a Gaussian kernel $w_\gamma(t) = (1/\gamma)\phi(t/\gamma)$ with $\gamma > 0$ as in (2.4) produces a zero-mean infinitely differentiable stationary ergodic Gaussian field $z_\gamma(t)$ such that

$$z'_\gamma(t) = w'_\gamma(t) * z(t) = \sigma \int_{\mathbb{R}} \frac{-(t-s)}{\xi^3}\phi\left(\frac{t-s}{\xi}\right) dB(s), \quad \xi = \sqrt{\gamma^2 + \nu^2},$$

with $\sigma_\gamma'^2 = \sigma^2/(4\sqrt{\pi}\xi^3)$, $\lambda_{4,\gamma} = 3\sigma^2/(8\sqrt{\pi}\xi^5)$ and $\lambda_{6,\gamma} = 15\sigma^2/(16\sqrt{\pi}\xi^7)$. We have

$$\text{SNR}_{j,\gamma} = \frac{a_j w_\gamma(0)}{\sigma_\gamma'} = \frac{\sqrt{2}|a_j|}{\sigma\pi^{1/4}} \left[\frac{(\gamma^2 + \nu^2)^{3/4}}{\gamma} \right]. \tag{3.6}$$

As a function of γ, the SNR has a local minimum at $\gamma^* = \sqrt{2}\nu$ and is strictly increasing for large γ. In particular, when $\nu = 0$, it is strictly increasing in γ. Thus we generally expect the detection power to increase with γ for $\gamma > \sqrt{2}\nu$. This will be confirmed in the simulations below. Note that for $\nu > 0$, the SNR is unbounded as $\gamma \to 0$, however in practice γ cannot be too small: if the support of w_γ becomes smaller than the sampling interval, then the derivative μ_γ' cannot be estimated.

4. Simulation studies

4.1. Performance of the dSTEM algorithm

Simulations were used to evaluate the performance and limitations of the dSTEM algorithm for signals $\mu(t) = a\lfloor t/d \rfloor$, where $t = 1, \ldots, L$, $L = 12000$, and signal strength $a \in \{1, 1.5, 2\}$. Under this setting, the true change points are $v_j = jd$ for $j = 1, \ldots, L/d - 1$, and the distance between neighboring change points is $d = 100$. The noise is generated as the Gaussian process constructed in (3.5) with $\sigma = 1$ and varying ν. Notice that the random error is white noise when $\nu = 0$, and is autocorelated when $\nu > 0$. The smoothing kernels are $w_\gamma(t) = (1/\gamma)\phi(t/\gamma)\mathbb{1}(t \in [-4\gamma, 4\gamma])$ for varying γ. The BH procedure was applied at FDR level $\alpha = 0.1$ and the tolerance $b = 5$. Results were averaged over 1,000 replications to simulate the expectations.

The results of FDR and power are shown in Figure 3. We see that for fixed γ and ν, as the strength of the signal a increases, FDR will decrease while the power will increase; moreover, FDR is eventually controlled below the nominal level and the power tends to 1, which is consistent with Theorems 3.1 and 3.3. For each fixed a, the power is seen to first decrease quickly and then increase again as γ increases. This phenomenon coincides with the behavior of the SNR (3.6) derived in Example 3.1, predicting the power to decrease for $\gamma \leq \sqrt{2}\nu$ and increase for $\gamma > \sqrt{2}\nu$. Meanwhile, if a is moderate or large, the FDR is seen to first increase and then decrease as γ increases, with the maximum of FDR still controlled below the nominal level.

In the simulations in Figure 3 above, the distance of neighboring change points $d = 100$ is large enough so that the kernel smoothing effectively affects only one change point at a time. However, if d is small, then the kernel smoothing with large γ may produce interference between neighboring change points, causing the power decrease. To illustrate this, we take the case where the signal strength is $a = 1.5$ and perform simulations for FDR and power with $d = 40, 30$ and 20. As shown in Figure 4, too large a γ makes FDR increase and power decrease, due to the overlap between the kernel smoothing at neighboring change
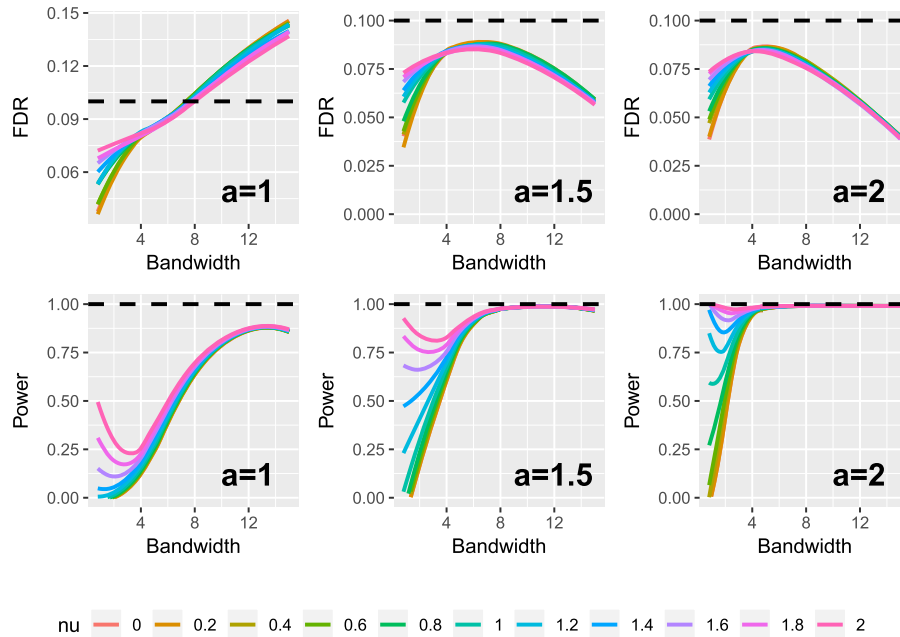
FIG 3. *The FDR (top) and power (bottom) vs. different combinations of the smoothness parameter ν (ranging from 0 to 2), the signal strength a (taking values 1, 1.5 and 2) and the bandwidth γ (ranging from 0.2 to 15). Here, the significance level $\alpha = 0.1$, tolerance $b = 5$ and $d = 100$.*

points. This phenomenon becomes more evident when d is small ($d = 20$). Theoretically, the neighboring interference would happen when d is less than the support of the smoothing kernel, which is about 8γ in our Gaussian case here. In particular, we see that the turning point of the power, attaining almost its maximum, appears at around $\gamma = 8$ for cases $d = 40$ and $d = 30$, while at around $\gamma = 5$ for the case $d = 20$. This suggests that $\gamma = d/4$ is a good choice of bandwidth when d is not large.

4.2. Choice of the bandwidth γ

Figure 3 shows that the bandwidth γ will greatly affect the performance of dSTEM. We see that larger γ tends to attain a smaller FDR and larger power. However, as shown in Figure 4, if γ is too large, it will produce interference between neighboring change points and contamination error, thereby decreasing the power. Thus, the choice of γ is critical for the performance of dSTEM.

The optimal bandwidth γ is the value that maximizes the power while controlling the FDR under the significance level. It is difficult to obtain the optimal γ theoretically. However, in our model, the signal is a piecewise constant function, and it is possible to obtain the optimal γ in practice by making more
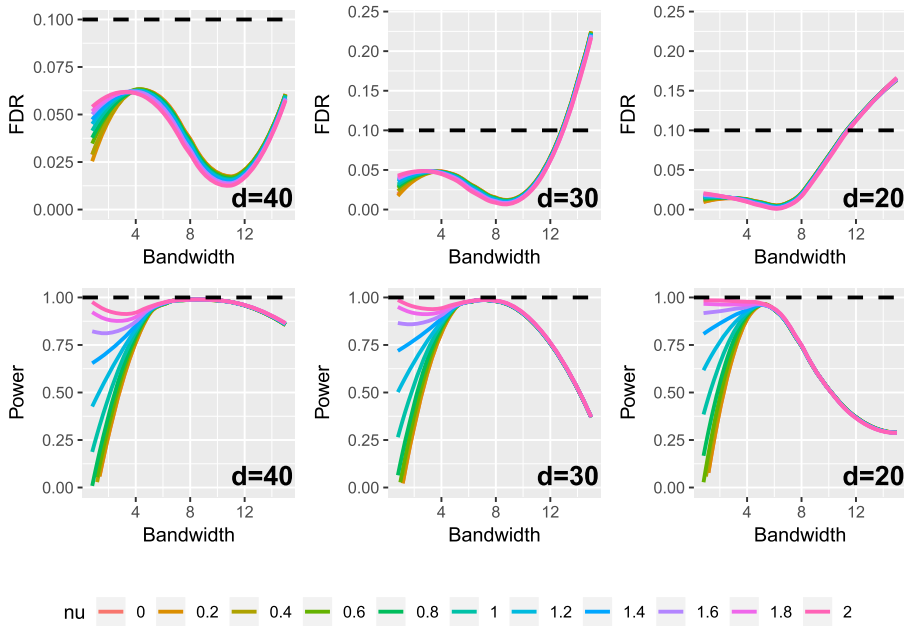
FIG 4. *The FDR (top) and power (bottom) vs. different combinations of the smoothness parameter ν (ranging from 0 to 2), the width between neighboring change points d (taking values 40, 30 and 20) and the bandwidth γ (ranging from 0.2 to 15). Here, the significance level $\alpha = 0.1$, tolerance $b = 5$ and signal strength $a = 1.5$.*

assumptions on the noise, such as in Example 3.1 using the Gaussian autocorrelation model. Figure 3 suggests that γ should be chosen to be about 8 for weak signals, while it can be as small as 4 for strong signals, almost regardless of the noise autocorrelation. To avoid producing interference between neighboring change points, the minimal distance between change points $d = \inf_j |v_j - v_{j-1}|$ defined in (2.2) should be large. On the other hand, if d is not large, then we can choose the bandwidth γ to be about d/c where the effective support of the smoothing kernel is $\pm c\gamma$. This has been shown in simulations in Figure 4 where $c = 4$ and $\gamma = d/4$.

4.3. Comparison with algorithm FDRSeg

As mentioned in Section 1, FDRSeg is the newest method which can control FDR for change point detection. In this subsection, we compare the performance of our method dSTEM with the algorithm FDRSeg. First, it is worth mentioning that our method is mainly designed for autocorrelated random noise, while FDRSeg requires independent and identically distributed random error, which is just a special case of our method ($\nu = 0$). Note that FDRSeg contains only one parameter α_F, which controls the theoretical upper bound of FDR at $2\alpha_F/(1 -$

TABLE 1

Performance comparison of dSTEM and FDRSeg under independent noise. The parameter
α_F is the only parameter in FDRSeg, controlling the theoretical upper bound of FDR to
$2\alpha_F/(1 - \alpha_F)$.

		dSTEM			**FDRSeg**	
	γ	FDR	Power	α_F	FDR	Power
	9	0.113	0.723	0.05	0.119	0.840
$a = 1$	10	0.117	0.781			
	11	0.124	0.820	0.1	0.155	0.865
	12	0.131	0.848			
	γ	FDR	Power	α_F	FDR	Power
	6	0.091	0.896	0.05	0.033	0.943
$a = 1.5$	7	0.089	0.943			
	8	0.088	0.965	0.1	0.090	0.957
	9	0.083	0.974			
	γ	FDR	Power	α_F	FDR	Power
	4	0.085	0.932	0.05	0.008	0.971
$a = 2$	5	0.088	0.978			
	6	0.082	0.987	0.1	0.049	0.983
	7	0.085	0.989			

TABLE 2

Performance comparison of dSTEM and FDRSeg under autocorrelated noise ($\nu = 1$).

		dSTEM			**FDRseg**	
	γ	FDR	Power	α_F	FDR	Power
	9	0.112	0.733	0.05	0.808	1.000
$a = 1$	10	0.118	0.792			
	11	0.127	0.827	0.1	0.815	1.000
	12	0.134	0.851			
	γ	FDR	Power	α_F	FDR	Power
	6	0.088	0.908	0.05	0.796	1.000
$a = 1.5$	7	0.086	0.949			
	8	0.086	0.968	0.1	0.802	1.000
	9	0.084	0.976			
	γ	FDR	Power	α_F	FDR	Power
	4	0.084	0.952	0.05	0.785	1.000
$a = 2$	5	0.083	0.980			
	6	0.083	0.988	0.1	0.795	1.000
	7	0.081	0.990			

α_F). However, they suggest that in practice their method should give FDR $\leq \alpha_F$. Thus, we let α_F be 0.05 and 0.1.

Table 1 shows the realized FDR and detection power under independent noise situation. We see that for small signal $a = 1$, dSTEM can almost control FDR and its power is 84.8% when $\gamma = 12$; while FDRSeg can attain a little larger power, but it is hard to control FDR when $\alpha = 0.1$. For larger signal a, both two methods can control FDR and attain a similar large power. Table 2 shows the results under the situation of autocorrelated noise ($\nu = 1$). In this case, the performance of dSTEM is nearly the same as that in independent scenario, while FDRSeg tends to estimate a large number of change points, leading to a large FDR, which means it can hardly control FDR.

5. Data example

5.1. Magnetometer sensor readings

In the field of mobile security, two-factor authentication/verification is of great importance, which is an extra layer of security of your mobile device, such as smartphones, wearable, and smart home devices, designed to ensure that you are the only person who can unlock your device, even if someone knows your password. In recent years, gesture based key establishment is a popular topic in communication security and computer science [29, 28, 16].

Modern mobile devices embedded with various motion sensors including accelerometer, gyroscope and magnetometer are used to measure and record the gesture performing process. Obtaining accurate readings of magnetometer is the foundation of magnetometer baesd research. However, during data collection, there are always noises which might be caused by hardware imperfection, manipulation error or sensitivity of the sensor. Particularly, finding the start point and associated end point for each gesture is a big challenge. The goal of this analysis is to find such change points.

In this paper, the data was collected from an experiment where several simple gestures, for example, shaking the smartphone in different directions and at different speeds, were designed. In particular, the smartphone defines a coordinate system of the embedded magnetometer, which is shown in Figure 5. The magnetometer can record the speed of smartphone movement as the readings along X, Y and Z axes. In our case, we will only show the results of the readings on X-axis, since readings along other two axes could be processed similarly.

In this data analysis, the sample size is $n = 6510$. The bandwidth was chosen not too large in order to avoid interference between neighboring change points. Choosing $\gamma = 18$ implies that the effective support of the Gaussian kernel up to 5 standard deviations on each side, for a total of 180 points, is still well below the typical distance between change points, which is about 300 sample points. Figure 6 shows results of the detected change points. Due to the measurement error and magnetic-field interference, the real underlying data will be interfered by slight fluctuations, leading to lots of (349) local maxima and minima, as shown in Figure 6 (top right). However, despite that, our method can still find the true change points, whose number is actually not large, as shown in the bottom left panel. In the bottom right panel, it is obvious that the start points and associated end points are very well detected.

Figure 7 shows the results of FDRSeg, it is obvious that FDRSeg estimates too many (1609) change points, which is nonsense and consistent with the simulation results under autocorrelated noise.

5.2. Array-CGH data

Array-based comparative genomic hybridization (array-CGH) is a high-throughput high-resolution technique used to evaluate changes in the number of copies
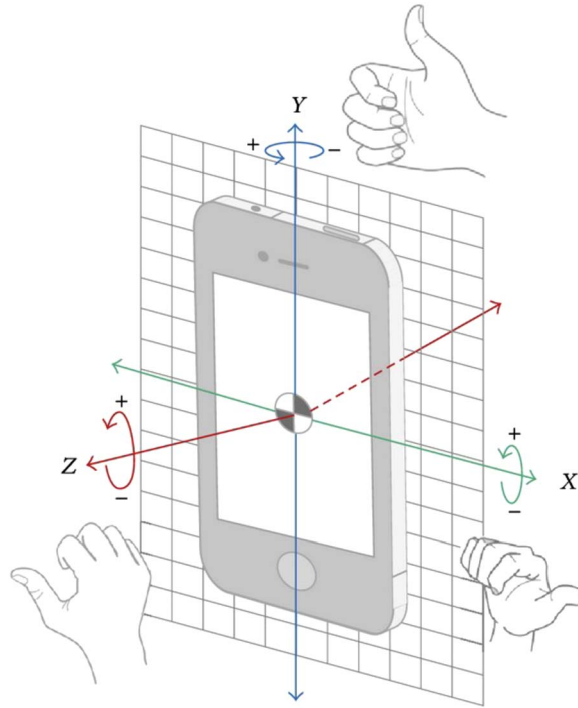
Fig 5. *The coordinate system of the magnetometer embedded in a smartphone.*

of alleles at thousands of genomic loci simultaneously. The output is often called Copy Number Variation (CNV) data. Changes in copy number are represented by segments whose mean is displaced with respect to the background. To detect these changes, it is costumary to search for change points along the genome.

In this paper, we apply our method to chromosome 1 of tumor sample #18 from the dataset of [12] and [21]. This sample is one of 37 formalin-fixed breast cancer tumors in that dataset and it was chosen for its visual appeal in the illustration of our method. The data in chromosome 1 of tumor sample #18 consists of 968 average copy number reads mapped onto 968 unequally spaced locations along the chromosome. For simplicity, the data was analyzed ignoring the gaps in the genomic locations. Figure 8 (top left) shows the data with spacings between reads artificially set to 1. Note that ignoring the spacings does not affect the presence or absence of change points.

To analyze the data, the dSTEM algorithm was applied with a truncated Gaussian smoothing kernel $w_\gamma(t) = (1/\gamma)\phi(t/\gamma)\mathbb{1}(t \in [-4\gamma, 4\gamma])$ with $\gamma = 10$. Again, a bandwidth of $\gamma = 10$ was chosen not too large in order to avoid interference between neighboring change points. Figure 8 (top right) shows the estimated first derivative (2.6). Figure 8 (bottom left) marks 19 local maxima (green) and 19 local minima (red).
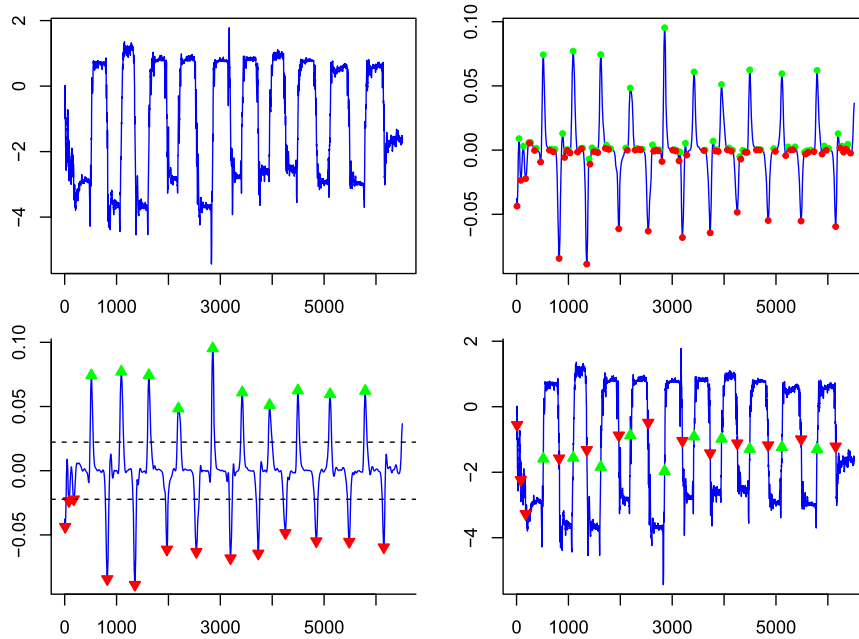
Fig 6. *Magnetometer example. Top left: Observed data. Top right: Estimated first derivative of the smoothed data, and local maxima (green), local minima (red), and significant height threshold (black dashed line). Bottom left: The first derivative, and the detected positive (green upward triangle) and negative (red downward triangle) change points. Bottom right: The observed data and its change points.*
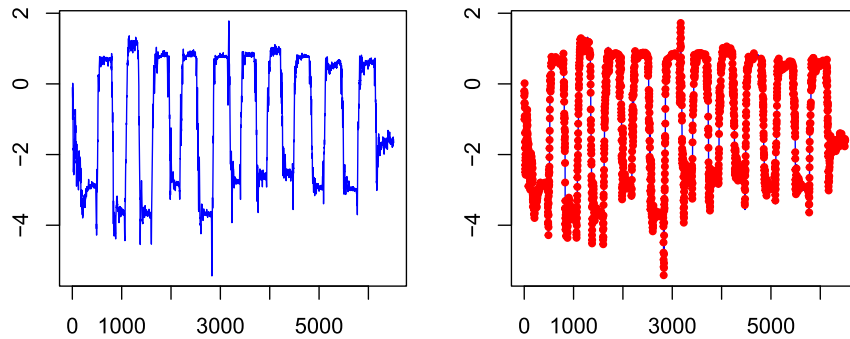

Fig 7. *Magnetometer example. Left: Observed data. Right: Detected change points (red) by FDRSeg.*

P-values corresponding to local maxima and minima were computed according to (2.9) using the distribution (2.11). The required parameters $\sigma_\gamma'^2 = \mathrm{Var}(z_\gamma'(t))$, $\lambda_{4,\gamma} = \mathrm{Var}(z_\gamma''(t))$, $\lambda_{6,\gamma} = \mathrm{Var}(z_\gamma'''(t))$ were estimated empirically from the estimated first, second and third derivatives over the observed data

sequence. However, the empirical variances were computed using truncated averages instead of regular averages in order to avoid bias from the extreme derivatives at the change points without assuming their presence or location in advance. The BH algorithm was applied to the 38 p-values FDR level 0.2, yielding a p-value significance threshold of 4.42×10^{-4}. The corresponding absolute height threshold of 0.089 is marked as dashed lines in Figure 8 (bottom left). The significant peaks are plotted on the original data in Figure 8 (bottom right) with a location tolerance of $b = 2$ for visual reference.
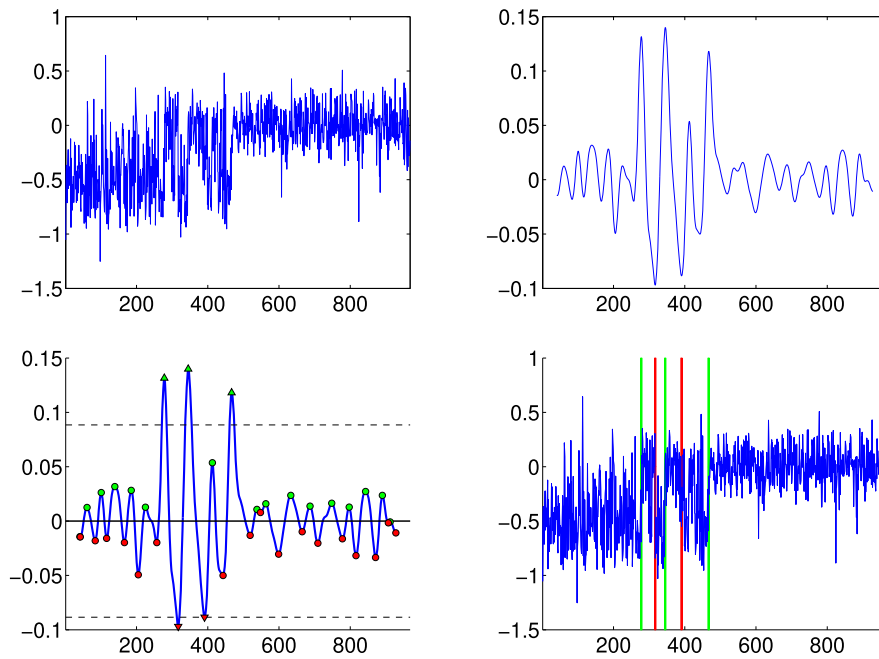

FIG 8. *Data example. Top left: Observed data. Top right: Estimated first derivative. Bottom left: Local maxima (upward triangles), local minima (downward triangles) of the estimated first derivative, and significance height threshold (black dashed line). Bottom right: The detected positive (green) and negative (red) change points.*

6. Discussion

6.1. Increasing and decreasing change points

In this paper, we combined both local maxima and minima of the derivative as candidate peaks, and then applied a multiple testing procedure to find a uniform threshold (in absolute value) for detecting all change points. This approach is sensible when the distributions (number and height) of true increasing and decreasing change points are about the same. Alternatively, different thresholds

for detecting increasing and decreasing change points could be found by applying separate multiple testing procedures to the sets of candidate local maxima and local minima. While we applied the BH algorithm to control FDR, in principle other multiple testing procedures may be used to control other error rates.

6.2. The smoothing bandwidth

A natural and important question is how to choose the smoothing bandwidth γ. We can see that either a small γ (if the noise is highly autocorrelated) or a relatively large γ (if the noise is less autocorrelated) is preferred in order to increase power, but only to the extent that the smoothed signal supports $h'_{j,\gamma}(t)$ have little overlap and that detected change points are not displaced by more than the desired tolerance b (recall that the value of b is not used in the dSTEM algorithm itself, but it may be determined by the needs of the specific scientific application). Considering the Gaussian kernel to have an effective support of $\pm c\gamma$, a good value of γ may be about $\min(b, d/(2c))$, where d is the separation between change points. For example, if we consider the Gaussian kernel to have an effective support of $\pm 4\gamma$ and the separation between change points is $d = 100$, we may choose γ to be no larger than $\gamma = 10$. Since the location of the change points is unknown, a more precise optimization of γ may require an iterative procedure. Moreover, if some change points are close together and others are far apart, an adaptive bandwidth may be preferable. We leave these as problems for future research.

6.3. Extensions to more general models

The problem of change point detection has been extensively studied and there are many different models due to various applications. As a novel approach, we believe that our method can be extended to solve the detection problem in other useful models. The following are two possible extensions for future study.

6.3.1. Structure breaks detection in piecewise linear models

We have studied in this paper the case when the signal is a piecewise constant function. However, in practice, it is common that the signal is a linear trend with multiple structure breaks, which can be treated as change points. Detecting the structure breaks has important applications in econometrics and time series; see for example [2, 25]. Specifically, adopting similar notations in this paper, the target piecewise linear model can be reduced to the following signal-plus-noise model

$$y(t) = \mu(t) + z(t),$$

where $z(t)$ is the Gaussian noise and the signal $\mu(t)$ is a piecewise linear function of the form

$$\mu(t) = c_j + k_j t, \quad t \in [v_{j-1}, v_j),$$

with $c_j, k_j \in \mathbb{R}$, $j = 1, 2, 3, \ldots$ and $0 = v_0 < v_1 < v_2 < \cdots$. Assume the structures of $\mu(t)$ are different at neighboring v_j, $j \geq 1$. That is, the neighboring coefficient vectors (c_j, k_j) and (c_{j+1}, k_{j+1}) are not equal for $j \geq 1$. We are interested in finding the location and number of these structure breaks v_j.

Our dSTEM algorithm could be extended to solve this problem. However, there are several challenges. One is that the structure of differentiated smoothed signal $\mu'_\gamma(t)$ would be more complicated, especially depending on the values of coefficients c_j and k_j at neighboring v_j. For example, the local extrema of $\mu'_\gamma(t)$ depend on both the change of slopes and the jump at v_j, and their location may not be exactly at v_j. On the other hand, a change in slope would produce different values of the background signal on both sides of the structure breaks, creating a change point in the derivative. A possible solution to this problem could involve estimating and testing second derivatives by filtering with a second order differential kernel.

6.3.2. Change point detection for models with nonstationary Gaussian noise

We have assumed stationary Gaussian noise in our model for simplicity. The stationarity assumption allowed us to use an explicit formula for the height distribution of local extrema [4, 6]. However, in many applications, nonstationary noise is more realistic. To our knowledge, there are no existing methods for change point detection with nonstationary noise. Our work here provides a promising approach to solving this problem. To compute p-values, the height distribution of local extrema for smooth nonstationary Gaussian processes can be computed explicitly as long as the covariance function of the process is known [4, 6]. Otherwise, p-values could be approximated using the approximate overshoot distribution. However, challenges include how to estimate the covariance function and to prove FDR control and power consistency in this setting.

Acknowledgements

Supplementary Material

An R package named "dSTEM", for performing the dSTEM algorithm 2.1 for change point detection, is available in R Cran.

References

[1] AKRITAS, M. G. AND POLITIS, D. N. (2003). Inference for nonsmooth regression curves and surfaces using kernel-based methods. *Recent advances and trends in nonparametric statistics*, 183. MR2498239

[2] BAI, J. AND PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 1, 47–78. MR1616121

[3] BARRY, D. AND HARTIGAN, J. A. (1993). A bayesian analysis for change point problems. *Journal of the American Statistical Association* **88**, 421, 309–319. MR1212493

[4] CHENG, D. AND SCHWARTZMAN, A. (2015). Distribution of the height of local maxima of gaussian random fields. *Extremes 18*, 213–240. MR3351815

[5] CHENG, D. AND SCHWARTZMAN, A. (2017). Multiple testing of local maxima for detection of peaks in random fields. *The Annals of Statistics* **45**, 2, 529–556. MR3650392

[6] CHENG, D. AND SCHWARTZMAN, A. (2018). Expected number and height distribution of critical points of smooth isotropic gaussian random fields. *Bernoulli* **24**, 4B, 3422–3446. MR3788177

[7] CRAMÉR, H. AND LEADBETTER, M. R. (2013). *Stationary and related stochastic processes: Sample function properties and their applications.* Courier Corporation. MR0217860

[8] EFRON, B. AND ZHANG, N. R. (2011). False discovery rates and copy number variation. *Biometrika* **98**, 2, 251–271. MR2806427

[9] EILERS, P. H. AND DE MENEZES, R. X. (2004). Quantile smoothing of array cgh data. *Bioinformatics* **21**, 7, 1146–1153.

[10] ERDMAN, C. AND EMERSON, J. W. (2008). A fast bayesian change point analysis for the segmentation of microarray data. *Bioinformatics* **24**, 19, 2143–2148.

[11] FRICK, K., MUNK, A., AND SIELING, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 3, 495–580. MR3210728

[12] HSU, L., SELF, S. G., GROVE, D., RANDOLPH, T., WANG, K., DELROW, J. J., LOO, L., AND PORTER, P. (2005). Denoising array-based comparative genomic hybridization data using wavelets. *Biostatistics* **6**, 2, 211–226.

[13] HUANG, T., WU, B., LIZARDI, P., AND ZHAO, H. (2005). Detection of dna copy number alterations using penalized least squares regression. *Bioinformatics* **21**, 20, 3811–3817.

[14] HUNG, Y., WANG, Y., ZARNITSYNA, V., ZHU, C., AND WU, C. J. (2013). Hidden markov models with applications in cell adhesion experiments. *Journal of the American Statistical Association* **108**, 504, 1469–1479. MR3174722

[15] JACKSON, B., SCARGLE, J. D., BARNES, D., ARABHI, S., ALT, A., GIOUMOUSIS, P., GWIN, E., SANGTRAKULCHAROEN, P., TAN, L., AND TSAI, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters* **12**, 2, 105–108.

[16] KILIBOZ, N. Ç. AND GÜDÜKBAY, U. (2015). A hand gesture recognition

technique for human–computer interaction. *Journal of Visual Communication and Image Representation* **28**, 97–104.

[17] KILLICK, R., ECKLEY, I. A., EWANS, K., AND JONATHAN, P. (2010). Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Engineering* **37**, 13, 1120–1126.

[18] LAI, W. R., JOHNSON, M. D., KUCHERLAPATI, R., AND PARK, P. J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array cgh data. *Bioinformatics* **21**, 19, 3763–3770.

[19] LANZANTE, J. R. (1996). Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology: A Journal of the Royal Meteorological Society* **16**, 11, 1197–1226.

[20] LI, H., MUNK, A., AND SIELING, H. (2016). Fdr-control in multiscale change-point segmentation. *Electronic Journal of Statistics* **10**, 1, 918–959. MR3486421

[21] LOO, L. W., GROVE, D. I., WILLIAMS, E. M., NEAL, C. L., COUSENS, L. A., SCHUBERT, E. L., HOLCOMB, I. N., MASSA, H. F., GLOGOVAC, J., LI, C. I., AND OTHERS. (2004). Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Research* **64**, 23, 8541–8549.

[22] MUGGEO, V. M. AND ADELFIO, G. (2010). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics* **27**, 2, 161–166.

[23] NAM, C. F., ASTON, J. A., AND JOHANSEN, A. M. (2012). Quantifying the uncertainty in change points. *Journal of Time Series Analysis* **33**, 5, 807–823. MR2969913

[24] OLSHEN, A. B., VENKATRAMAN, E., LUCITO, R., AND WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics* **5**, 4, 557–572.

[25] QIAN, J. AND SU, L. (2015). Shrinkage estimation of regression models with multiple structural changes. *Econometric Theory* **32**, 6, 1–58. MR3569279

[26] REEVES, J., CHEN, J., WANG, X. L., LUND, R., AND LU, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* **46**, 6, 900–915.

[27] SCHWARTZMAN, A., GAVRILOV, Y., AND ADLER, R. J. (2011). Multiple testing of local maxima for detection of peaks in 1d. *Annals of Statistics* **39**, 6, 3290. MR3012409

[28] SIMÃO, M., NETO, P., AND GIBARU, O. (2016). Natural control of an industrial robot using hand gesture recognition with neural networks. In *IECON 2016 – 42nd Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 5322–5327.

[29] TAN, W., WU, C., ZHAO, S., AND LI, J. (2010). Dynamic hand gesture recognition using motion trajectories and key frames. In *2010 2nd International Conference on Advanced Computer Control*. Vol. **3**. IEEE, 163–167.

[30] TIBSHIRANI, R. AND WANG, P. (2007). Spatial smoothing and hot spot

detection for cgh data using the fused lasso. *Biostatistics* **9**, 1, 18–29.

[31] Wang, P., Kim, Y., Pollack, J., Narasimhan, B., and Tibshirani, R. (2005). A method for calling gains and losses in array cgh data. *Biostatistics* **6**, 1, 45–58.

[32] Zeileis, A., Shah, A., and Patnaik, I. (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics & Data Analysis* **54**, 6, 1696–1706. MR2600868