Statistica Sinica **33** (2023), 873-892
doi:https://doi.org/10.5705/ss.202021.0048

# MOVING SUM DATA SEGMENTATION FOR STOCHASTIC PROCESSES BASED ON INVARIANCE

Claudia Kirch and Philipp Klein

*Otto-von-Guericke-University Magdeburg*

*Abstract:* The segmentation of data into stationary stretches, also known as the multiple change point problem, is important for many applications in time series analysis and signal processing. Based on strong invariance principles, we analyze a data segmentation methodology using moving sum statistics for a class of regime-switching multivariate processes, where each switch results in a change in the drift. In particular, this framework includes the data segmentation of multivariate partial sum, integrated diffusion, and renewal processes, even if the distance between the change points is sublinear. We study the asymptotic behavior of the corresponding change point estimators, show their consistency, and derive the corresponding localization rates, which are minimax optimal in a variety of situations, including an unbounded number of changes in Wiener processes with drift. Furthermore, we derive the limit distribution of the change point estimators for local changes. This result can, in principle, be used to derive confidence intervals for the change points.

*Key words and phrases:* Change point analysis, data segmentation, invariance principle, moving sum statistics, multivariate processes, regime-switching processes

## 1. Introduction

Change point analysis aims at detecting and localizing structural breaks in time series data, with applications in a variety of fields, such as neurophysiology (see Messer et al. (2014)), genomics (compare Olshen et al. (2004), Niu and Zhang (2012), Li, Munk and Sieling (2016), Chan and Chen (2017)), finance (Aggarwal, Inclan and Leal (1999), Cho and Fryzlewicz (2012)), astrophysics (see Fisch, Eckley and Fearnhead (2022)), and oceanographics (Killick et al. (2010)).

Early literature focused on testing for a single change point in the mean. Later studies examined changes in more complex data structures, and currently focus on detecting changes in high-dimensional data; see for example Csörgö and Horvàth (1997), Horváth and Rice (2014), and Cho and Kirch (2021).

During the last two decades, interest has shifted from testing to the multiple change problem, aiming at segmenting data into stationary stretches, often fo-

Corresponding author: Claudia Kirch, Institute for Mathematical Statistics, Department of Mathematics, Otto-von-Guericke University, 39106 Magdeburg, Germany. E-mail: claudia.kirch@ovgu.de.

cusing on changes in the mean of independent and identically distributed (i.i.d.) Gaussian data (Cho and Kirch (2021)). Although moving sum (MOSUM) statistics were first considered for testing (Bauer and Hackl (1980), Hušková and Slabý (2001)), they are better suited as a basis for data segmentation (Yau and Zhao (2016), Eichinger and Kirch (2018), Meier, Cho and Kirch (2021), Cho and Kirch (2022)).

We adopt a MOSUM approach to localize multiple changes in multivariate renewal processes. Here, the analysis of neuronal firing patterns, so called spike trains, is a prominent example, where data segmentation methods for renewal processes are useful. Indeed, many methods, including Grün, Diesmann and Aertsen (2002) and Schneider (2008), apply local approaches to segments with approximately constant intensity to model the data. Furthermore, it is of great interest to study the joint behavior of spike trains; see for example Perkel, Gernstein and Moore (1967), Brown, Kass and Mitra (2004), and Grün and Rotter (2010). Chen, Chen and Ding (2019) use nonparametric methods to detect change points in neuropixel data, which consist of a large number of neuronal firing patterns, in order to make meaningful assertions about the whole or part of the data. In particular, they study firing patterns in several different brain areas, and make assertions on the possible coordination between regions based on their change point patterns. Messer et al. (2014) propose a MOSUM multiscale procedure to detect changes in the firing intensity. assuming that the firing patterns follow renewal processes with piecewise constant intensity. Our work extends their results in several ways.

First, we prove the consistency of the change point estimators and derive the corresponding localization rates, where we allow for both linear and sublinear bandwidths. Without the latter, consistency cannot be achieved in the important situation where the distance between the change points is sublinear. In addition, we go beyond the univariate case, including multivariate point processes based on renewal processes in our analysis.

Although our main interest lies in detecting multiple changes in renewal processes, we adopt a more general framework for deriving our theoretical results that also includes detecting changes in partial sum and diffusion processes. A univariate version of that model with at most one change point has been considered by Horváth and Steinebach (2000) and Kühn and Steinebach (2002). A univariate version for finitely many change points has been considered by Kühn (2001), who shows the consistency for the number of change points. We extend those results to include a MOSUM methodology for estimating a possibly unbounded number of change points in a multivariate setting. Here we achieve a minimax optimal separation rate, in addition to a minimax optimal localization rate (for the change point estimators) for a bounded number of change points and for Wiener processes with drift (see Remark 3 below).

The remainder of the paper is organized as follows. In Subsection 2.1, we introduce a general multiple change point model, followed by a discussion of renewal processes as an example for the model in Subsection 2.2. In Section 3, we describe how to estimate change points based on MOSUM statistics. First, we introduce the MOSUM statistics in 3.1, before presenting the estimators for the structural breaks in 3.2. In 3.3, we derive some asymptotic results for the MOSUM statistics that are required for threshold selection, and can also be used in a testing context. In Section 4, we show that the corresponding data segmentation procedure is consistent. Finally, we derive the localization rates and the corresponding asymptotic distribution of the change point estimators for local changes. In Section 5, we present some results from a small simulation study. The proofs can be found in the Supplementary Material S2.

## 2. Multiple Change Point Problem

### 2.1. Model

Although our main interests lies in detecting changes in renewal processes, we prove the results for the following more general model that additionally includes partial sum and certain diffusion processes.

Consider $P < \infty$ stochastic processes $\{\mathbf{R}_{t,T}^{(j)} : 0 \leq t \leq T\}$ in continuous time of dimension $p$ with (unknown) drift $(\boldsymbol{\mu}_T^{(j)} \cdot t)$ and (unknown) covariance $(\mathbf{\Sigma}_{j,T} \cdot t)$ fulfilling the regularity assumptions specified in Assumption 1 below. These $P$ processes can be thought of as background processes, with only one of them being active at each time, in the sense of driving the increments of our observation process. Consequently, at each time point, we observe only the active process, and do not know the exact structure of any of these processes. To elaborate, for $c_\ell < t \leq c_{\ell+1}$, we observe

$$\mathbf{Z}_{t,T} = \left(\mathbf{R}_{t,T}^{(c_{\ell+1})} - \mathbf{R}_{c_\ell,T}^{(c_{\ell+1})}\right) + \sum_{j=1}^{\ell} \left(\mathbf{R}_{c_j,T}^{(c_j)} - \mathbf{R}_{c_{j-1},T}^{(c_j)}\right), \tag{2.1}$$

where $0 = c_0 < c_1 < \cdots < c_{q_T} < c_{q_T+1} = T$ are the unknown change points and the number of change points $q_T$ can be bounded or unbounded.

The upper index $^{(c_j)}$ at the process $\mathbf{R}_{\cdot,T}$ indicates (with a slight abuse of notation) the active process between the $(j-1)$th and the $j$th change point. We define the change in drift between two neighboring regimes by

$$\mathbf{d}_{i,T} := \boldsymbol{\mu}_T^{(c_{i+1})} - \boldsymbol{\mu}_T^{(c_i)} \neq 0 \quad \text{for all } i = 1, \ldots, q_T, \tag{2.2}$$

where $\mathbf{d}_{i,T}$ is bounded, but we allow for $\mathbf{d}_{i,T} \to 0$, as long as the convergence is sufficiently slow (see Assumption 2). For ease of notation, we frequently drop the dependency on $T$ for the above quantities. The aim of data segmentation is the consistent estimation of the number and location of the change points, as well as deriving of the corresponding localization rates.

We assume that the underlying processes $\{\mathbf{R}_{t,T}^{(j)}\}$, for $j = 1, \ldots, P$, fulfill the following joint invariance principle toward Wiener processes. If the underlying processes are independent, then this simplifies to the validity of an invariance principle for each of these $P$ processes.

**Assumption 1.** *Denote the joint process by* $\mathbf{R}_{t,T} = (\mathbf{R}_{t,T}^{(1)'}, \ldots, \mathbf{R}_{t,T}^{(P)'})'$ *and the joint drift by* $\boldsymbol{\mu}_T = (\boldsymbol{\mu}_T^{(1)'}, \ldots, \boldsymbol{\mu}_T^{(P)'})'$, *where* $'$ *indicates the matrix transpose. For every* $T > 0$ *there exist* $(p \cdot P)$*-dimensional Wiener processes* $\mathbf{W}_{t,T}$ *with covariance matrix* $\boldsymbol{\Sigma}_T$ *and*

$$\boldsymbol{\Sigma}_T^{(i)} = (\boldsymbol{\Sigma}_T(l,k))_{l,k=p\,(i-1)+1,\ldots,p\,i},$$

*with*

$$\|\boldsymbol{\Sigma}_T^{(i)}\| = O(1), \qquad \|\boldsymbol{\Sigma}_T^{(i)^{-1}}\| = O(1),$$

*such that, possibly after a change of probability space, it holds that for some sequence* $\nu_T \to 0$,

$$\sup_{0 \le t \le T} \|\widetilde{\mathbf{R}}_{t,T} - \mathbf{W}_{t,T}\| = \sup_{0 \le t \le T} \|(\mathbf{R}_{t,T} - \boldsymbol{\mu}_T\, t) - \mathbf{W}_{t,T}\| = O_P\left(T^{1/2}\, \nu_T\right),$$

*where* $\widetilde{\mathbf{R}}_{t,T} = \mathbf{R}_{t,T} - \boldsymbol{\mu}_T\, t$ *denotes the centered process.*

The covariance matrix $\boldsymbol{\Sigma}_T^{(i)}$ relates to the $i$th underlying process $\{\mathbf{R}_{t,T}^{(i)}\}$, and plays an important role in the limit results below. On the other hand, the cross-dependence between different driving processes does not influence these limit results, because at each time, only one process actively influences the observed process and the increments of the joint process are asymptotically independent, owing to the joint invariance principle.

The assumption on the norm of the covariance matrices is equivalent to the smallest eigenvalue of $\boldsymbol{\Sigma}_T^{(i)}$ being bounded, in addition to being bounded away from zero (both uniformly in $T$). In many situations, the covariance matrices do not depend on $T$, in which case this assumption is automatically fulfilled under positive definiteness. The convergence rate $\nu_T$ in the invariance principle typically depends on the number of moments that exist. Roughly speaking, the more moments the original process has, the faster $\nu_T$ converges.

The corresponding univariate model with at most one change was first considered by Horváth and Steinebach (2000), and further used in a single-change setting by Steinebach (2000), Kirch and Steinebach (2006), and Gut and Steinebach (2002; 2009). Kühn and Steinebach (2002) use the Schwarz information criterion to estimate the number of change points in a related univariate framework with a bounded number of change points. Using information criteria is computationally much more expensive, with quadratic computational complexity than MOSUM procedures with linear computational complexity, as proposed in this paper.

### 2.2. Renewal and some related point processes

In this section, we explain the connection of our model to renewal processes, which are also considered in the simulation study. Further examples, such as partial sum and diffusion processes, can be found in the Supplementary Material (see Section S1.1). We consider $P$ independent sequences of $p$-dimensional point processes that are related to renewal processes in the following way. For each $i = 1, \ldots, P$, we start with $\tilde{p} \geq p$ independent renewal processes $\widetilde{R}_{t,j}^{(i)}$, for $j = 1, \ldots, \tilde{p}$, from which we derive a $p$-dimensional point process $\mathbf{R}_t^{(i)} = \mathbf{B}^{(i)}(\widetilde{R}_{t,1}^{(i)}, \ldots, \widetilde{R}_{t,\tilde{p}}^{(i)})'$, where $\mathbf{B}^{(i)}$ is a $(p \times \tilde{p})$ matrix with nonnegative integer-valued entries. By Lemma 4.2 in Steinebach and Eastwood (1996), Assumption 1 is fulfilled for a block-diagonal $\mathbf{\Sigma}_T$ with

$$\mathbf{\Sigma}_T^{(i)} = \mathbf{B}^{(i)}\mathbf{D}\left(\frac{\boldsymbol{\sigma}^2(i)}{\boldsymbol{\mu}^3(i)}\right)\mathbf{B}^{(i)\prime},$$

$$\text{with } \mathbf{D}\left(\frac{\boldsymbol{\sigma}^2(i)}{\boldsymbol{\mu}^3(i)}\right) = \operatorname{diag}\left(\frac{\sigma_1^2(i)}{\mu_1^3(i)}, \ldots, \frac{\sigma_{\tilde{p}}^2(i)}{\mu_{\tilde{p}}^3(i)}\right),$$

where $\mu_j(i)$ and $\sigma_j^2(i)$ are the mean and variance, respectively of the corresponding inter-event times. Steinebach and Eastwood (1996) and Csenki (1979) consider $\tilde{p} = p$, but use inter-event times that are dependent for $j = 1, \ldots, p$. In such a situation, the invariance principle in Assumption 1 still holds if the intensities are the same across components with $\Sigma_T^{(i)} = \Sigma_{\text{IET}}^{(i)}/\mu_1^3(i)$, where $\Sigma_{\text{IET}}^{(i)}$ is the covariance of the vector of inter-event times. We adopt this setting in the simulation study. If the intensities differ, then by Steinebach and Eastwood (1996), an invariance principle toward a Gaussian process can still be obtained. Here each component is still a Wiener process, but the increments from one component may depend on the lagged behavior of the other components, where the lag increases with time. MOSUM procedures for related univariate renewal processes have been considered in Messer et al. (2014), Messer et al. (2017) nd Messer and Schneider (2017). However, they do not derive any consistency results for their change point estimators and only consider linear bandwidths.

## 3. Data Segmentation Procedure

### 3.1. Moving sum statistics

By assumption, the drifts of the two active processes to the left and right of a change point differ, see (2.2). On the other hand, in a stationary stretch away from any change point, the drift is the same. Because the difference in drift can be estimated by a difference of increments, we propose the following MOSUM-statistic that is based on the moving difference of increments with bandwidth $h = h_T$:

$$\mathbf{M}_t = \mathbf{M}_{t,T,h_T}(\mathbf{Z}) = \frac{1}{\sqrt{2h}}\left[(\mathbf{Z}_{t+h} - \mathbf{Z}_t) - (\mathbf{Z}_t - \mathbf{Z}_{t-h})\right]$$
$$= \frac{1}{\sqrt{2h}}\left(\mathbf{Z}_{t+h} - 2\mathbf{Z}_t + \mathbf{Z}_{t-h}\right). \tag{3.1}$$

If there is no change, then this difference will fluctuate around zero, but will be different from zero close to a change point. On the one hand, the bandwidth should be chosen to be as large as possible (to get a better estimate obtained from a larger "effective sample size" of order $h$). On the other hand, the increments should not be contaminated by a second change, because this can lead to situations where the change point can no longer be reliably localized by the signal. This observation is reflected in the following assumptions on the bandwidth.

**Assumption 2.** *For $\nu_T$ as in Assumption 1, the bandwidth $h < T/2$ fulfills*

$$\frac{\nu_T^2 T \log T}{h} \to 0.$$

*Furthermore, it isolates the $i$th change point in the sense of*

$$h \leq \frac{1}{2}\Delta_i, \qquad \text{where } \Delta_i = \min(c_{i+1} - c_i, c_i - c_{i-1}). \tag{3.2}$$

*In addition, the signal needs to be large enough to be detectable by this bandwidth, that is,*

$$\frac{\|\mathbf{d}_i\|^2 h}{\log(T/h)} \to \infty. \tag{3.3}$$

Combining (3.2) and (3.3) shows that, with an appropriate bandwidth $h$, changes are detectable as soon as

$$\frac{\|\mathbf{d}_i\|^2\,\Delta_i}{\log\left(T/\Delta_i\right)} \to \infty. \tag{3.4}$$

In the case of the classical mean change model, as in Subsection S1.1 of the Supplementary Material, this is known to be the minimax-optimal separation rate, which cannot be improved (see Proposition 1 of Arias-Castro, Candes and Durand (2011)).

The assumption on the distance of the first and last change point to the boundary of the process in (3.2) can be relaxed, because no boundary effects can occur there.

### 3.2. Change point estimators

The MOSUM statistic $\mathbf{M}_t = \mathbf{m}_t + \mathbf{\Lambda}_t$, as in (3.1), decomposes into a piecewise linear signal term $\mathbf{m}_t = \mathbf{m}_{t,h,T}$ and a centered noise term $\mathbf{\Lambda}_t = \mathbf{\Lambda}_{t,h,T}$, with

$$\sqrt{2h}\mathbf{m}_t = \begin{cases} (h - t + c_i)\,\mathbf{d}_i, & \text{for } c_i < t \le c_i + h, \\ 0, & \text{for } c_i + h < t \le c_{i+1} - h, \\ (h + t - c_{i+1})\,\mathbf{d}_{i+1}, & \text{for } c_{i+1} - h < t \le c_{i+1}, \end{cases} \tag{3.5}$$

$$\sqrt{2h}\,\mathbf{\Lambda}_t = \sqrt{2h}\,\mathbf{\Lambda}_t(\widetilde{\mathbf{R}}) \tag{3.6}$$

$$= \begin{cases} \widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_t^{(c_{i+1})} + \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_i)} + \widetilde{\mathbf{R}}_{t-h}^{(c_i)}, & \text{for } c_i < t \le c_i + h, \\ \widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_t^{(c_{i+1})} + \widetilde{\mathbf{R}}_{t-h}^{(c_{i+1})}, & \text{for } c_i + h < t \le c_{i+1} - h, \\ \widetilde{\mathbf{R}}_{t+h}^{(c_{i+2})} - \widetilde{\mathbf{R}}_{c_{i+1}}^{(c_{i+2})} + \widetilde{\mathbf{R}}_{c_{i+1}}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_t^{(c_{i+1})} + \widetilde{\mathbf{R}}_{t-h}^{(c_{i+1})}, & \text{for } c_{i+1} - h < t \le c_{i+1}, \end{cases}$$

where $\widetilde{\mathbf{R}}_t := \mathbf{R}_t - t\boldsymbol{\mu}$ for $i = 0, \ldots, q_T$, and the upper index $c_j$ denotes the active regime between the $(j-1)$th and $j$th change point (with a slight abuse of notation).

The signal term is a piecewise linear function that takes its extrema at the change points, and is zero outside $h$-intervals around the change points. Furthermore, the noise term is asymptotically negligible compared to the signal term (see Theorem 1 for the corresponding theoretical statement and Figure 1 for an illustrative example).

This motivates the following data segmentation procedure that considers local extrema that are big enough (in absolute value) as change point estimators: For a suitable threshold $\beta = \beta_{h,T}$ (see Section 3.3 for a detailed discussion), we

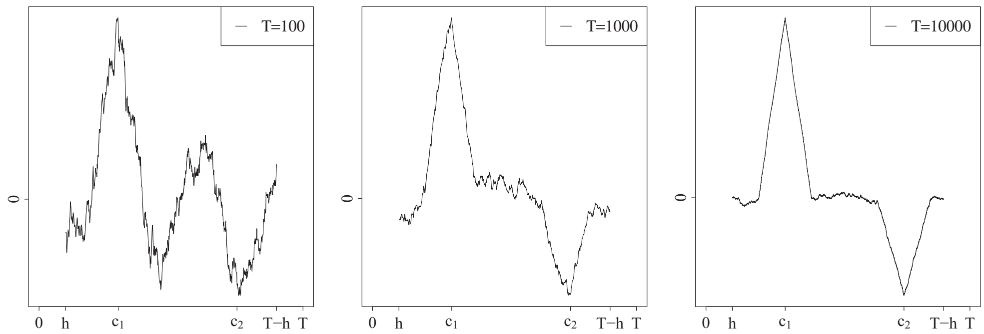


Figure 1. Univariate MOSUM statistic with $T = 100, 1000, 10000$ (from left to right), where the noise term (fluctuating around the signal) becomes smaller and smaller relative to the signal term.

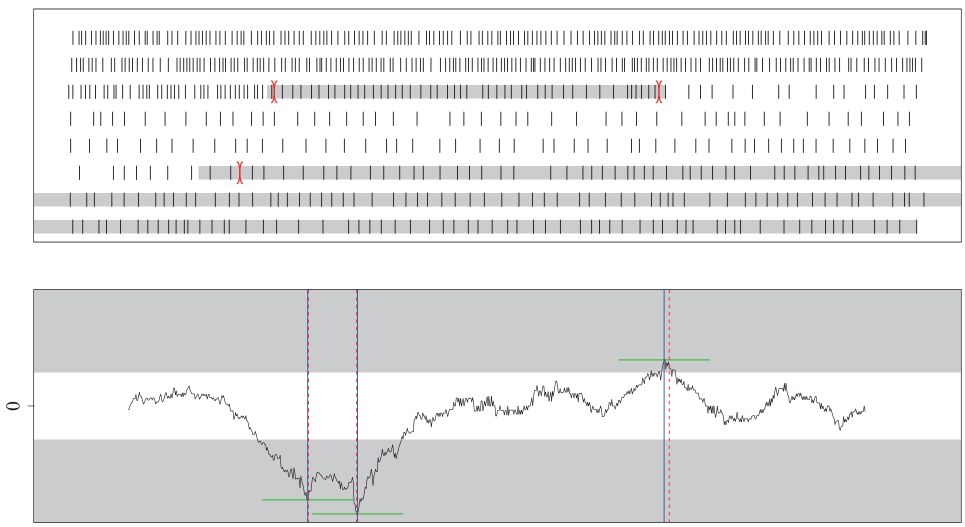


Figure 2. In the upper panel, the observed event times of a univariate renewal process with three change points (i.e., four stationary segments) are displayed (where the plot needs to be read like a text: It starts in the upper row on the left, then continues in the first row, and jumps to the second row and, so on). The gray and white regions mark the estimated segmentation of the data, while the red intervals mark the true segmentation. In the lower panel, the corresponding MOSUM statistic with (relative) bandwidth $h/T = 0.07$ is displayed. The gray areas are the regions where the threshold ($\alpha = 0.05$, as in Remark 1) is exceeded (in absolute value). The blue solid lines indicate the change point estimates obtained as local extrema that fall within the gray area (making them *significant*). The true change points are indicated by the red dashed lines. The green horizontal lines denote $\eta h$-environments around the estimators.

define *significant* time points, where a point $t^*$ is *significant* if

$$\mathbf{M}'_{t^*}\,\widehat{\mathbf{A}}_{t^*}^{-1}\,\mathbf{M}_{t^*} \geq \beta. \tag{3.7}$$

Here, $\widehat{\mathbf{A}}_{t^*}$ is a symmetric positive-definite matrix that may depend on the data and fulfills the following assumption.

**Assumption 3.**

$$\sup_{h\leq t\leq T-h}\left\|\widehat{\mathbf{A}}_{t,T}^{-1}\right\| = O_P\left(1\right), \qquad \sup_{i=1,\ldots,q_T}\ \sup_{|t-c_i|\leq h}\left\|\widehat{\mathbf{A}}_{t,T}\right\| = O_P(1).$$

A good (non-data-driven) choice fulfilling this assumption is given by

$$\mathbf{\Sigma}_t = \mathbf{\Sigma}_{t,T} = \mathbf{\Sigma}_T^{(c_i)}, \tag{3.8}$$

for $c_{i-1} < t \leq c_i$, which guarantees scale-invariance of the procedure and allows for nicely interpretable thresholds (see Section 3.3). The latter remains true for estimators as long as they fulfill

$$\sup_{i=1,\ldots,q_T}\ \sup_{|t-c_i|>h}\left\|\widehat{\mathbf{\Sigma}}_{t,T}^{-1/2} - \mathbf{\Sigma}_t^{-1/2}\right\| = o_P\left(\left(\log\frac{T}{h}\right)^{-1}\right), \tag{3.9}$$

in addition to the above boundedness assumptions. In particular, this permits local estimators that are consistent only away from change points, but are contaminated by the change in the latter's local environment. This is typically the case for covariance estimators, such as the sample variance contaminated by a change point. To improve detectability, it is beneficial if the estimator is additionally consistent directly at the change point (see, e.g., Eichinger and Kirch (2018)).

Typically, there are intervals of significant points (owing to the continuity of the signal) such that only local extrema of such intervals actually indicate a change point. To define a local extremum, we require a tuning parameter $0 < \eta < 1$. This parameter defines the locality requirement on the extremum, where a point $t^*$ is a local extremum if it maximizes the absolute MOSUM statistic within its $\eta h$-environment, that is, if

$$t^* = \min\left\{\underset{t^*-\eta h\leq t\leq t^*+\eta h}{\operatorname{argmax}}\|\mathbf{M}_t\|\right\}. \tag{3.10}$$

The threshold $\beta$ distinguishes between *significant* and *spurious* local extrema that are associated purely with the noise term. The set of all significant local extrema is the set of change point estimators with its cardinality an estimator for the number of change points.

Figure 2 shows an example illustrating these ideas. Away from the change points, the MOSUM statistic fluctuates around 0 (within the white area that is beneath the threshold in absolute value), while it falls within the gray area close to the change points, making corresponding local extrema significant. Furthermore, the statistic does not need to return to the white area in order to have all changes estimated, as can be seen between the first and second change points. In addition, the figure shows that (3.2) is required for theoretic considerations only, but can be weakened in practice in combination with a suitable $\eta$, with $\eta h$ defining the minimal distance that two MOSUM estimators can have. This is one of the major advantages of the $\eta$-criterion based on *significant* local maxima, as described here (in comparison to the $\epsilon$-criterion originally investigated by Eichinger and Kirch (2018) in the context of mean changes; see also the discussion in Meier, Cho and Kirch (2021)). Results for the $\epsilon$-criterion can be obtained along the lines of our proofs below. In practical applications, if $\eta$ is chosen too large, some pairs of change points may be indistinguishable by our procedure. On the other hand, too small a choice of $\eta$ can increase the number of spurious and duplicate estimators, as shown in S1 in the Supplementary Material. However, the latter is not a problem if we apply a post-processing step, as in Cho and Kirch (2022).

### 3.3. Threshold selection

The procedure clearly depends on the choice of a threshold $\beta = \beta_{h,T}$ (see (3.7)) that can distinguish between significant and spurious local extrema. The following theorem gives the magnitudes of the signal and the noise terms:

**Theorem 1.** *Let Assumptions* 1, 2 *and* 3 *hold.*

*(a) For the signal* $\mathbf{m}_t$ *with* $c_i - h < t < c_i + h$, *it holds that*

$$\mathbf{m}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{m}_t \geq \frac{1}{2\|\widehat{\mathbf{A}}_t\|} \frac{(h - |t - c_i|)^2}{h} \|\mathbf{d}_i\|^2.$$

*At other time points, the noise term is equal to zero.*

*(b) For the noise term the following hold for* $q_T = 0$, *that is, in the no-change situation:*

*(i) for a linear bandwidth* $h = \gamma T$ *with* $0 < \gamma < 1/2$,

$$\sup_{\gamma T \le t \le T - \gamma t} \mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t \xrightarrow{\mathcal{D}} \sup_{\gamma \le s \le 1-\gamma} \frac{1}{2\gamma} \left(\mathbf{B}_{s+\gamma} - 2\mathbf{B}_s + \mathbf{B}_{s-\gamma}\right)' \left(\mathbf{B}_{s+\gamma} - 2\mathbf{B}_s + \mathbf{B}_{s-\gamma}\right),$$

*where* $\mathbf{B}$ *denotes a multivariate standard Wiener process. In particular, the squared noise term is of order* $O_P(1)$.

*(ii) for a sublinear bandwidth* $h/T \to 0$, *it holds under Assumption 2 that*

$$a\left(\frac{T}{h}\right) \sup_{h \le t \le T-h} \sqrt{\mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t} - b\left(\frac{T}{h}\right) \xrightarrow{\mathcal{D}} E,$$

*where* $E$ *follows a Gumbel distribution with* $\mathrm{P}(E \le x) = e^{-2e^{-x}}$ *and*

$$\begin{aligned} a(x) &= \sqrt{2\log x} \\ b(x) &= 2\log x + \frac{p}{2}\log\log x + \log\frac{3}{2} - \log\Gamma\left(\frac{p}{2}\right). \end{aligned}$$

*In particular, the squared noise term is of order* $O_P\left(\log(T/h)\right)$.

*The assertions remain true if an estimator for the covariance is used fulfilling* (3.9) *uniformly over all* $h \le t \le T - h$.

*(c) In the situation of multiple change points, it holds that*

$$\sup_{h \le t \le T-h} \|\mathbf{\Lambda}_t\| = O_P\left(\sqrt{\log\left(\frac{T}{h}\right)}\right).$$

To obtain the consistency of the estimators, the threshold needs to be small enough to be asymptotically negligible compared to the squared signal term, as in Theorem 1 (a), to guarantee that every change is detected with asymptotic probability one. At the same time, the threshold needs to grow faster than the squared noise term in Theorem 1 (c), so that false positives occur with asymptotic probability zero. Both conditions are fulfilled under the following assumption.

**Assumption 4.** *The threshold fulfills*

$$\frac{\beta_{h,T}}{h_T \min_{i=1,\ldots,q_T} \|\mathbf{d}_i\|^2} \to 0, \qquad \frac{\log(T/h_T)}{\beta_{h,T}} \to 0 \qquad (T \to \infty).$$

In particular, larger bandwidths $h_T$ lead to better detectability of the change point, where (3.2) means that an upper bound related to the distance to the neighboring change points applies. This is confirmed by the simulation results in Table 4 in the Supplementary Material.

The following remark introduces a threshold that has a nice interpretation in terms of change point testing.

**Remark 1.** The threshold is often obtained as the asymptotic $1-\alpha_T$-quantile based on the limit result in Theorem 1 (b), for some sequence $\alpha_T \to 0$. In this case, a choice of

$$\frac{\left(-\log\log(1/\sqrt{1-\alpha_T})\right)^2}{\log(T/h_T)} = O(1),$$

similar to Eichinger and Kirch (2018), can replace the slightly stronger lower bound of Assumption 4 on the threshold without compromising our theoretical results. In the simulation study in Section 5, we use this threshold with $\alpha_T = 0.05$. This controls the family-wise error rate at level $\alpha_T$ asymptotically related to testing each time point for a possible change. In fact, Theorem 1 shows that such a threshold with a constant sequence $\alpha$ yields an asymptotic test at level $\alpha$, which has asymptotic power one by Theorem 2. Tests designed for at most one change, as in Hušková and Steinebach (2000) and Hušková and Steinebach (2002), often have better power, but are not as good at localizing change points (see Figure 1 in Cho and Kirch (2021)).

## 4. Consistency of the Segmentation Procedure

In this section, we show the consistency of the above segmentation procedure for the estimators of both the number and the locations of the change points. Furthermore, we derive localization rates for the estimators of the locations of the change points for some special cases, showing that they cannot be improved, in general. This is complemented by the observation that these localization rates are indeed minimax-optimal if the number of change points is bounded in addition to observing Wiener processes with drift. The following theorem shows that the change point estimators defined in (3.10) are consistent for the number and locations of the change points.

**Theorem 2.** *Let Assumptions* 1–4 *hold. Let* $0 < \hat{c}_1 < \cdots < \hat{c}_{\hat{q}_T}$ *be the change point estimators* (3.10)*. Then, for any* $\tau > 0$*, it holds that*

$$\lim_{T\to\infty} \mathrm{P}\left(\max_{i=1,\ldots,\min(\hat{q}_T,q_T)} |\hat{c}_i - c_i| \le \tau h, \hat{q}_T = q_T\right) = 1.$$

The theorem shows in particular that the number of change points is estimated consistently. For the linear bandwidth, we additionally get consistency of the change point locations in rescaled time, while for the sublinear bandwidths, we even get a convergence rate of $h/T$ for the rescaled change points. Under the following stronger assumptions, the localization rates can be improved further.

**Assumption 5.**

(a) *For any of the centered processes* $\widetilde{\mathbf{R}}^{(j)}$ *as in* (3.6) *and any value* $\theta_i = \theta_{i,T}$ *(which will be* $c_i$ *or* $c_i \pm h$ *when the assumption is applied), it holds for any sequence* $D_T \geq 1$ *(bounded or unbounded) that*

$$\sup_{D_T/\|\mathbf{d}_i\|^2 \leq s \leq h} \frac{\sqrt{D_T}\left\|\widetilde{\mathbf{R}}^{(j)}_{\theta_i} - \widetilde{\mathbf{R}}^{(j)}_{\theta_i \pm s}\right\|}{s\,\|\mathbf{d}_i\|} = O_P(\omega_T),$$

*where neither* $\omega_T$ *nor the constant depends on* $D_T$*.*

(b) *Let now the upper index* $\theta_i$ *denote the active stretch in the stationary segment* $(\theta_i, \theta_i + s)$*, respectively* $(\theta_i - s, \theta_i)$*. Then, it holds for any sequence* $D_T > 0$ *that*

$$\max_{i=1,\ldots,q_T} \sup_{D_T/\|\mathbf{d}_i\|^2 \leq s \leq h} \frac{\sqrt{D_T}\left\|\widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i} - \widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i \pm s}\right\|}{s\,\|\mathbf{d}_i\|} = O_P(\tilde{\omega}_T),$$

*where neither* $\tilde{\omega}_T$ *nor the constant depends on* $D_T$*.*

The localization rates of the MOSUM procedure are determined by the rates $\omega_n, \tilde{\omega}_n$, which need to be derived for each example separately (at least for the tight ones). For partial sum processes, the suprema in (a) are stochastically bounded by the Hájék–Rényi inequality, while the assertion in (b) is fulfilled with a polynomial rate in $q_T$ (see Cho and Kirch (2022), Proposition 2.1 (c)(ii)).

**Remark 2.**

(a) For Wiener processes with drift, we obtain $\omega_T = 1$ and $\tilde{\omega}_T = \sqrt{\log(q_T)}$ (see Proposition S2.1 in the Supplementary Material).

(b) By the invariance principle in Assumption 1, all rates are clearly dominated by $T^{1/2}\nu_T$. However, this is often too liberal a bound (see Proposition 2.1 in Cho and Kirch (2022) for some tight bounds in the case of partial sum processes).

(c) Often, there exist forward and backward invariance principles from some arbitrary starting value $\theta_i$ for each regime. This is the case for partial sum processes and for (backward and forward) Markov processes, owing to the Markov property. For renewal processes, this can be shown along the lines of the original proof for the invariance principle (Csörgö, Horváth and Steinebach (1987)), because the time to the next (previous) event is asymptotically negligible; see also Example 1.2 in Kühn and Steinebach (2002)). In this case, the Hájék–Rényi results for Wiener processes carry over (see Proposition S2.1 in the Supplementary Material) to the different processes underlying each regime, resulting in $\omega_T = 1$. For the situation with a bounded number of change points, this carries over to $\tilde{\omega}_T$.

**Theorem 3.** *Let Assumptions* 1–5 *hold. For* $\hat{q}_T < q_T$*, define* $\hat{c}_i = T$*, for* $i = \hat{q}_T + 1, \ldots, q_T$*.*

*(a) For a single change point estimator, the following localization rate holds:*

$$\|\mathbf{d}_i\|^2 |\hat{c}_i - c_i| = O_P\left(\omega_T^2\right).$$

*(b) The following uniform rate holds true:*

$$\max_{i=1,\ldots,q_T} \|\mathbf{d}_i\|^2 |\hat{c}_i - c_i| = O_P\left(\tilde{\omega}_T^2\right).$$

**Remark 3** (Minimax optimality)**.** As noted below (3.4), the separation rate given there is minimax optimal (see Proposition 1 of Arias-Castro, Candes and Durand (2011)). Minimax optimal localization rates (derived in the context of changes in the mean of univariate time series, which is covered by the partial sum processes in our framework) are known for a few special cases. First, the rate for a single change point, and thus also for a bounded number of change points, is given by $\omega_T = 1$ in the above notation (see, e.g., Lemma 2 in Wang, Yu and Rinaldo (2020)). Consequently, our procedure achieves minimax optimality for a bounded number of change points under weak assumptions (as pointed out in Remark 2 (c)). Second, the optimal localization rate for unbounded change points under sub-Gaussianity (attained for partial sum process of i.i.d. errors) is given by $\tilde{\omega}_T = \sqrt{\log T}$ (see Proposition 6 in Verzelen et al. (2020) and Proposition 2.3 in Cho and Kirch (2022)). Indeed, we match this rate for Wiener processes with drift.

The following theorem derives the limit distribution of the change point estimators for local changes, which shows in particular that the rates are tight. In principle, this result can be used to obtain asymptotically valid confidence intervals for the change point locations. In the case of fixed changes, the limit distribution depends on the underlying distribution of the original process (see Antoch and Hušková (1999) for the case of partial sum processes), where the proof follows the same lines. We need the following assumption.

**Assumption 6.** *Let* $\mathbf{d}_i = \mathbf{d}_{i,T} = \|\mathbf{d}_i\|\mathbf{u}_i + o(\|\mathbf{d}_i\|)$ *with* $\|\mathbf{u}_i\| = 1$ *and* $\|\mathbf{d}_{i,T}\| \to 0$. *Assume that* $\mathbf{Y}_s^{(j)} = \mathbf{Y}_s^{(j)}(c_i, D)$, *with*

$$\mathbf{Y}_s^{(1)} = \widetilde{\mathbf{R}}^{(c_i)}_{c_i-h+(s-D)/\|d_i\|^2} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i-h-D/\|d_i\|^2},$$
$$\mathbf{Y}_s^{(21)} = \widetilde{\mathbf{R}}^{(c_i)}_{c_i+(s-D)/\|d_i\|^2} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i-D/\|d_i\|^2}, \quad \mathbf{Y}_s^{(22)} = \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+(s-D)/\|d_i\|^2} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-D/\|d_i\|^2},$$
$$\mathbf{Y}_s^{(3)} = \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h+(s-D)/\|d_i\|^2} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h-D/\|d_i\|^2}$$

*fulfill the following multivariate functional central limit theorem for any constant* $D > 0$ *in an appropriate space equipped with the supremum norm*

$$\left\{\|\mathbf{d}_i\| \, (\mathbf{Y}_s^{(1)}, \mathbf{Y}_s^{(21)}, \mathbf{Y}_s^{(22)}, \mathbf{Y}_s^{(3)})' : 0 \le s \le 2D\right\} \xrightarrow{w} \left\{\widetilde{\mathbf{W}}_s : 0 \le s \le 2D\right\},$$

*where* $\widetilde{\mathbf{W}}$ *is a Wiener process with covariance matrix* $\boldsymbol{\Xi}$ *(not depending on* $D$*). For* $-D \le t \le D$, *denote* $\mathbf{W}_t = (\mathbf{W}_t^{(1)}, \mathbf{W}_t^{(21)}, \mathbf{W}_t^{(22)}, \mathbf{W}_t^{(3)})' = \widetilde{\mathbf{W}}_{D+t} - \widetilde{\mathbf{W}}_D$.

By Assumption 2, $h\|\mathbf{d}_i\|^2 \to \infty$ holds, such that the distance $h - 2D/\|d_i\|^2$ between $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2j)}$ (resp. between $\mathbf{Y}^{(2j)}$ and $\mathbf{Y}^{(3)}$) diverges to infinity. As such, for processes with independent increments, the processes $\mathbf{Y}^{(1)}$, $(\mathbf{Y}^{(21)}, \mathbf{Y}^{(22)})'$, $\mathbf{Y}^{(3)}$ are independent for sufficiently large $T$. Furthermore, under weak assumptions, such as mixing conditions, this independence still holds asymptotically in the sense that $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$ and $\mathbf{W}^{(3)}$ are independent.

Functional central limit theorems for these processes follow from invariance principles, as in Assumption 1 with $\boldsymbol{\Sigma}_T \to \boldsymbol{\Sigma}$, as long as such invariance principles still hold with an arbitrary (moving) starting value, which is typically the case (see also Remark 2 (c)). As such, it typically holds that $\boldsymbol{\Xi}^{(1)} = \boldsymbol{\Xi}^{(21)} = \boldsymbol{\Sigma}^{(c_i)}$ and $\boldsymbol{\Xi}^{(3)} = \boldsymbol{\Xi}^{(22)} = \boldsymbol{\Sigma}^{(c_{i+1})}$ where $\boldsymbol{\Xi}^j = \mathrm{Cov}(\mathbf{W}_1^{(j)})$ and $\boldsymbol{\Sigma}^{(c_i)}$ is the covariance matrix associated with the regime between the $(i-1)$th and $i$th change points.

The following theorem gives the asymptotic distribution for the change point estimators in the case of local change points.

**Theorem 4.** *Let Assumptions* 1–4 *and* 5(a) *hold with* $\omega_T = 1$, *and* 6 *hold. For* $\hat{q}_T < q_T$, *define* $\hat{c}_i = T$, *for* $i = \hat{q}_T + 1, \ldots, q_T$. *Let*

$$\Psi_t^{(i)} := -|t| + \begin{cases} \mathbf{u}_i'\mathbf{W}_t^{(1)} - 2\,\mathbf{u}_i'\mathbf{W}_t^{(21)} + \mathbf{u}_i'\mathbf{W}_t^{(3)}, & t < 0, \\ \mathbf{u}_i'\mathbf{W}_t^{(1)} - 2\,\mathbf{u}_i'\mathbf{W}_t^{(22)} + \mathbf{u}_i'\mathbf{W}_t^{(3)}, & t \ge 0. \end{cases}$$

*Then, for all $i = 1, \ldots, q_T$, it holds that for $T \to \infty$,*

$$\|\mathbf{d}_i\|^2 (\hat{c}_i - c_i) \xrightarrow{\mathcal{D}} \operatorname{argmax}\left\{ \Psi_t^{(i)} \middle| t \in \mathbb{R} \right\}.$$

*If there is a fixed number of changes $q_T = q$, with $q$, fixed and a functional central limit theorem as in Assumption 6 holds jointly for all $q$ change points, then the result also holds jointly.*

Owing to the Markov property of Wiener processes, $\{\Psi_t^{(i)} : t \geq 0\}$ is independent of $\{\Psi_t^{(i)} : t < 0\}$.

**Remark 4.**

(a) If $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$ and $\mathbf{W}^{(3)}$ are independent, which is typically the case (see the discussion beneath Assumption 6), then $\Psi_t^{(i)}$ simplifies to

$$\Psi_t^{(i)} := -|t| + \begin{cases} \sqrt{\sigma_{(1)}^2 + 4\,\sigma_{(21)}^2 + \sigma_{(3)}^2}\, B_t, & t < 0 \\ \sqrt{\sigma_{(1)}^2 + 4\,\sigma_{(22)}^2 + \sigma_{(3)}^2}\, B_t, & t \geq 0, \end{cases}$$

where $B$ is a (univariate) standard Wiener process and $\sigma_{(j)}^2 = \mathbf{u}_i' \mathbf{\Xi}^{(j)} \mathbf{u}_i$. Usually (see discussion beneath Assumption 6), $\sigma_{(21)} = \sigma_{(1)}$ and $\sigma_{(22)} = \sigma_{(3)}$, further simplifying the expression. For some examples, such as partial sum processes, $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ holds for all $t$, such that all $\sigma_{(j)}$ coincide. In this case, this further simplifies to

$$\Psi_t^{(i)} := -|t| + \sqrt{6}\,\sigma_{(1)}\, B_t.$$

For univariate partial sum processes, this result has already been obtained in Theorem 3.3 of Eichinger and Kirch (2018). However, the assumption of $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ is typically not fulfilled for renewal processes, because the covariance depends on the changing intensity of the process.

(b) If $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$ and $\mathbf{W}^{(3)}$ are independent and $\mathbf{M}_t$ in (3.10) is replaced by $\mathbf{\Sigma}_t^{-1/2}\mathbf{M}_t$, then the Wiener processes $\mathbf{W}^{(j)}$ are standard Wiener processes, such that $\Psi_t^{(i)}$ simplifies to

$$\Psi_t^{(i)} := -|t| + \sqrt{6}\, B_t.$$

This shows that in this case, the limit distribution of $\hat{c}_i - c_i$ depends only on the magnitude of the change $\mathbf{d}_i$, but not on its direction $\mathbf{u}_i$.

Statistically, however, this is difficult to achieve, because it requires a uniformly (in $t$) consistent estimator for the usually unknown covariance matrices $\mathbf{\Sigma}_t$.

## 5. Simulations and Discussion

### 5.1. Summary of simulation results

For univariate partial sum processes, extensive simulations and data examples for MOSUM statistics are conducted in Eichinger and Kirch (2018) and Meier, Cho and Kirch (2021), and results for renewal processes have been obtained by Messer et al. (2014, 2017). In the Supplementary Material S1, we complement these findings with simulations for multivariate renewal processes with multiple changes, where we consider both dependent and independent components as well as several choices for the matrix $\widehat{\mathbf{A}}_t$, as in (3.7).

It turns out that using the diagonal matrix with the asymptotic variances can lead to better or worse results than using the full asymptotic covariance matrix, depending on the nature of the change. From a statistical perspective, it is thus advantageous to use the diagonal matrix, because the local estimation of the inverse of a covariance matrix in moderately large or large dimensions is a very hard problem, leading to a loss in precision. In contrast the diagonal elements are far less difficult to estimate consistently. Using the diagonal matrix with the estimated variance instead of the true one leads to better detection power, with a substantial improvement for some changes. As such, local variance estimation can help boost the signal significantly, but comes at the cost of having a somewhat increased, while still reasonable number of spurious and duplicate change point estimators.

Using a single bandwidth only works well for *homogeneous* changes, in the sense that the smallest change in intensity is still large enough compared to the smallest distance to neighboring change points (for a detailed definition, we refer to Cho and Kirch (2022), Definition 2.1, or Cho and Kirch (2021), Definition 2.1). In some applications with *multiscale* signals, where frequent large changes and small isolated changes are present, this is no longer true. In such cases, several bandwidths need to be used, following by a pruning of the obtained candidates (see Cho and Kirch (2022) for an information criterion-based approach for partial sum processes and Messer et al. (2014) for a bottom-up approach for renewal processes). Similarly, if the distance to the neighboring change points is unbalanced, MOSUM procedures with asymmetric bandwidths, as suggested by Meier, Cho and Kirch (2021), may be necessary.

### 5.2. Discussion

In this paper, we introduce a data segmentation for multivariate processes with changes in the drift and obtain corresponding consistency results extending the work of Eichinger and Kirch (2018) and Messer et al. (2014). This is done in a general framework, where increments between change points are modeled by processes fulfilling a joint invariance principle. This framework includes renewal, partial sum, and some diffusion processes with change points.

One drawback of the procedure is the use of a single bandwidth. In practice, identifying the optimal bandwidth can be difficult, as pointed out e. g. by Cho and Kirch (2022) and Messer et al. (2014), among others. On the one hand, one wants to choose a large bandwidth in order to have maximum power, while on the other hand, choosing too large a bandwidth may lead to the misspecification or nonidentification of changes. Furthermore, as can be seen in the simulation study, in a multiscale change point situation (see Definition 2.1 of Cho and Kirch (2022)), no single bandwidth can detect all change points. Therefore, future work should extend the proposed procedure to a true multiscale setup, as in Cho and Kirch (2022).

## Supplementary Material

In Section S1, we present a small simulation study where we analyze the performance of our procedure using three-dimensional renewal processes. In Subsection S1.1, we present further types of processes that fulfill our model. Section S2 presents the proofs of Sections 3 and 4.

## Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 314838170, GRK 2297 MathCoRe.

## References

Aggarwal, R., Inclan, C. and Leal, R. (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis* **34**, 33–55.

Antoch, J. and Hušková, M. (1999). Estimators of changes. In *Asymptotics, Nonparametrics, and Time Series* (Edited by S. Ghosh), 533-578. Marcel Dekker, New York.

Arias-Castro, E., Candes, E. J. and Durand, A. (2011). Detection of an anomalous cluster in a network. *The Annals of Statistics* **39**, 278–304.

Bauer, P. and Hackl, P. (1980). An extension of the mosum technique for quality control. *Technometrics* **22**, 1–7.

Brown, E. N., Kass, R. E. and Mitra, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neuroscience* **7**, 456–461.

Chan, H. P. and Chen, H. (2017). Multi-sequence segmentation via score and higher-criticism tests. *arXiv:1706.07586.*

Chen, H., Chen, S. and Ding, X. (2019). A universal nonparametric event detection framework for neuropixels data. *bioRxiv:650671.*

Cho, H. and Fryzlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica* **22**, 207–229.

Cho, H. and Kirch, C. (2021). Data segmentation algorithms: Univariate mean change and beyond. *Econometrics and Statistics*, 2452–3062.

Cho, H. and Kirch, C. (2022). Two-stage data segmentation permitting multiscale changepoints, heavy tails and dependence. *Annals of the Institute of Statistical Mathematics* **74**, 653–684.

Csenki, A. (1979). An invariance principle in $k$-dimensional extended renewal theory. *Journal of Applied Probability* **16**, 567–574.

Csörgö, M. and Horvàth, L. (1997). *Limit Theorems in Change-Point Analysis*. 18th Edition. John Wiley & Sons, New York.

Csörgö, M., Horváth, L. and Steinebach, J. (1987). Invariance principles for renewal processes. *The Annals of Probability* **15**, 1441–1460.

Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24**, 526–564.

Fisch, A. T. M., Eckley, I. A. and Fearnhead, P. (2022). A linear time method for the detection of collective and point anomalies. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **15**, 494–508.

Grün, S., Diesmann, M. and Aertsen, A. (2002). Unitary events in multiple single-neuron spiking activity: II. nonstationary data. *Neural Computation* **14**, 81–119.

Grün, S. and Rotter, S. (2010). *Analysis of Parallel Spike Trains*. Springer, New York.

Gut, A. and Steinebach, J. (2002). Truncated sequential change-point detection based on renewal counting processes. *Scandinavian Journal of Statistics* **29**, 693–719.

Gut, A. and Steinebach, J. (2009). Truncated sequential change-point detection based on renewal counting processes II. *Journal of Statistical Planning and Inference* **139**, 1921–1936.

Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *Test* **23**, 219–255.

Horváth, L. and Steinebach, J. (2000). Testing for changes in the mean or variance of a stochastic process under weak invariance. *Journal of Statistical Planning and Inference* **91**, 365–376.

Hušková, M. and Slabỳ, A. (2001). Permutation tests for multiple changes. *Kybernetika* **37**, 605–622.

Hušková, M. and Steinebach, J. (2000). Limit theorems for a class of tests of gradual changes. *Journal of Statistical Planning and Inference* **89**, 57–77.

Hušková, M. and Steinebach, J. (2002). Asymptotic tests for gradual changes. *Statistics & Decisions* **20**, 137–151.

Kühn, C. (2001). An estimator of the number of change points based on a weak invariance principle. *Statistics & Probability Letters* **51**, 189–196.

Kühn, C. and Steinebach, J. (2002). On the estimation of change parameters based on weak invariance principles. *Limit Theorems in Probability and Statistics II* (Edited by I. Berkes, E. Csáki and M. Csörgö), 237–260. János Bolyai Math. Soc. Budapest.

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**, 1120–1126.

Kirch, C. and Steinebach, J. (2006). Permutation principles for the change analysis of stochastic processes under strong invariance. *Journal of Computational and Applied Mathematics* **186**, 64–88.

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

Meier, A., Cho, H. and Kirch, C. (2021). mosum: A package for moving sums in change point analysis. *Journal of Statistical Software* **97**, 1–42.

Messer, M., Costa, K. M., Roeper, J. and Schneider, G. (2017). Multi-scale detection of rate changes in spike trains with weak dependencies. *Journal of Computational Neuroscience* **42**, 187–201.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R. and Schneider, G. (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *The Annals of Applied Statistics* **8**, 2027–2067.

Messer, M. and Schneider, G. (2017). The shark fin function: Asymptotic behavior of the filtered derivative for point processes in case of change points. *Statistical Inference for Stochastic Processes* **20**, 253–272.

Niu, Y. S. and Zhang, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *The Annals of Applied Statistics* **6**, 1306–1326.

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**, 557–572.

Perkel, D. H., Gernstein, G. L. and Moore, G. P. (1967). Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. *Biophysical Journal* **7**, 419–440.

Schneider, G. (2008). Messages of oscillatory correlograms - a spike-train model. *Neural Computation* **20**, 1211–1238.

Steinebach, J. (2000). Some remarks on the testing of smooth changes in the linear drift of a stochastic process. *Theory of Probability and Mathematical Statistics* **61**, 173–185.

Steinebach, J. and Eastwood, V. R. (1996). Extreme value asymptotics for multivariate renewal processes. *Journal of Multivariate Analysis* **56**, 284–302.

Verzelen, N., Fromont, N., Lerasle, M. and Reynaud-Bouret, P. (2020). Optimal change point detection and localization. *arXiv:2010.11470*.

Wang, D., Yu, Y. and Rinaldo, A. (2020). Univariate mean change point detection: Penalization, cusum and optimality. *Electronic Journal of Statistics* **14**, 1917–1961.

Yau, C. Y. and Zhao, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**, 895–916.

Claudia Kirch

Institute for Mathematical Statistics, Department of Mathematics, Otto-von-Guericke University Magdeburg, Center for Behavioral Brain Sciences (CBBS), Universitätsplatz 2, 39106 Magdeburg, Germany.

E-mail: claudia.kirch@ovgu.de

Philipp Klein

Institute for Mathematical Statistics, Department of Mathematics, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany.

E-mail: philipp.klein@ovgu.de

(Received February 2021; accepted August 2021)