Statistica Sinica Preprint No: SS-2021-0048	
Title	Moving Sum Data Segmentation for Stochastic Processes Based on Invariance
Manuscript ID	SS-2021-0048
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0048
Complete List of Authors	Claudia Kirch and Philipp Klein
Corresponding Author	Claudia Kirch
E-mail	claudia.kirch@ovgu.de
Notice: Accepted version subject to English editing.	

Statistica Sinica

MOVING SUM DATA SEGMENTATION FOR STOCHASTIC PROCESSES BASED ON INVARIANCE

Claudia Kirch, Philipp Klein

Otto-von-Guericke-University Magdeburg

Abstract:

The segmentation of data into stationary stretches also known as multiple change point problem is important for many applications in time series analysis as well as signal processing. Based on strong invariance principles, we analyze data segmentation methodology using moving sum (MOSUM) 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 change points is sublinear. We study the asymptotic behavior of the corresponding change point estimators, show 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 – a result that 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 et al. (2016), Chan and Chen (2017)), finance (Aggarwal et al. (1999), Cho and Fryzlewicz (2012)), astrophysics (see Fisch et al. (2018)) or oceanographics (Killick et al. (2010)).

Early literature focused on testing for a single change point in the mean, moving on to changes in more complex data structure where currently a main interest lies in detecting changes in high-dimensional data; see e.g. Csörgö and Horvàth (1997); Horváth and Rice (2014); Cho and Kirch (2020).

During the last two decades interest shifted from testing to the multiple change problem aiming at segmenting the data into stationary stretches often focusing on changes in the mean of i.i.d. Gaussian data (Cho and Kirch (2020)). While moving sum (MOSUM) statistics were first considered for testing (Bauer and Hackl (1980), Hušková and Slabỳ (2001)), they are better suited as a basis for data segmentation (Yau and Zhao (2016), Eichinger and Kirch (2018), Meier et al. (2019), Cho and Kirch (2021)).

We adopt a MOSUM approach to localize multiple changes in multivariate renewal processes where the analysis of neuronal firing patterns, so called spike trains, is a very prominent example where data segmentation methods for renewal processes are useful. Indeed, many methods, e.g. Grün et al. (2002) or Schneider (2008) use local approaches applied on segments with approximately constant intensity to model the data. Furthermore, it is of great interest to study the joint behavior of spike trains, compare e.g. Perkel et al. (1967), Brown and Mitra (2004) and Grün and Rotter (2010). Chen et al. (2019) use non-parametric methods to detect change points in neuropixel data, which consists of a large amount of neuronal firing patterns, in order to make meaningful assertions about the whole or parts of the data. In particular, they study firing patterns in several different brain areas and make assertions on 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 consistency of the change point estimators and derive the corresponding localization rates where we allow for both linear as well as sublinear bandwidths. Without the latter, consistency cannot be achieved in the important situation where the distance between change points is sublinear. Additionally, we go beyond the univariate case including some multivariate point processes based on renewal processes in our analysis.

While our main interest lies in the detection of multiple changes in renewal processes, we adopt a more general framework for deriving our theoretical results that also includes detecting changes in partial sum as well as 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) where consistency for the number of change points has been shown. Those results are now extended to include MOSUM methodology for the estimation of a possibly unbounded number of change points in a multivariate setting, where we achieve a minimax optimal separation rate in addition to a minimax optimal localization rate (for the change point estimators) in case of a bounded number of change points as well as for Wiener processes with drift (see Remark 4.2 below).

Organization of the material

In Subsection 2.1, we introduce a general multiple change point model followed by a discussion of renewal processes as an import 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 in addition to 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 Materials S2.

2. Multiple change point problem

2.1 Model

While 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 $(\boldsymbol{\Sigma}_{j,T} \cdot t)$ fulfilling regularity assumptions specified in Assumption 2.1 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 only observe 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 $0 = c_0 < c_1 < \ldots < c_{q_T} < c_{q_T+1} = T$ 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 slow enough (see Assumption 3.1). For ease of notation we frequently drop the dependency on T for the above quantities in the following. The aim of data segmentation involves the consistent estimation of the number and location of the change points as well as the derivation of the corresponding localization rates.

We assume that the underlying processes $\{\mathbf{R}_{t,T}^{(j)}\}$, $j = 1, \ldots, P$, fulfill the following joint invariance principle towards a 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 2.1.

Denote the joint process by $\mathbf{R}_{t,T} = \left(\mathbf{R}_{t,T}^{(1)'}, \ldots, \mathbf{R}_{t,T}^{(P)'}\right)'$ *as well the joint drift by* $\boldsymbol{\mu}_T = \left(\boldsymbol{\mu}_T^{(1)'}, \ldots, \boldsymbol{\mu}_T^{(P)'}\right)'$, *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 $\nu_T \to 0$

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

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 below limit results. 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 due 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 will not depend on T, in which case this assumption is automatically fulfilled under positive definiteness. The convergence rate ν_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 ν_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), Gut and Steinebach (2002; 2009). Kühn and Steinebach (2002) make use of the Schwarz information criterion for the estimation of 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 if compared to 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 $\widetilde{R}_{t,j}^{(i)}$, $j = 1, \ldots, \tilde{p}$, from which we derive a p-dimensional point process $\mathbf{R}_t^{(i)} = \mathbf{B}^{(i)} \left(\widetilde{R}_{t,1}^{(i)}, \ldots, \widetilde{R}_{t,\tilde{p}}^{(i)} \right)'$, where $\mathbf{B}^{(i)}$ is a $(p \times \tilde{p})$ -matrix with non-negative integer-valued entries. By Lemma 4.2 in Steinebach and

Eastwood (1996) Assumption 2.1 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 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 2.1 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 – a setting that we adopt in the simulation study. If the intensities differ, then by Steinebach and Eastwood (1996) an invariance principle towards a Gaussian process can still be obtained, where each component still is 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) as well as Messer and Schneider (2017). However, they have not derived any consistency results for their change point estimators and only considered 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 moving sum (MOSUM) statistic that is based on the moving difference of increments with bandwidth $h = h_T$

$$\begin{aligned}\mathbf{M}_t = \mathbf{M}_{t,T,h_T}(\mathbf{Z}) &= \frac{1}{\sqrt{2h}}\left[(\mathbf{Z}_{t+h} - \mathbf{Z}_t) - (\mathbf{Z}_t - \mathbf{Z}_{t-h})\right] \\ &= \frac{1}{\sqrt{2h}}\left(\mathbf{Z}_{t+h} - 2\mathbf{Z}_t + \mathbf{Z}_{t-h}\right). \end{aligned} \tag{3.3}$$

If there is no change, then this difference will fluctuate around 0, while it will be different from 0 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 the order h). On the other hand, the increments should not be contaminated by a second change as 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 3.1. *For ν_T as in Assumption (2.1) 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

$$h \leq \frac{1}{2}\,\Delta_i, \qquad \textit{where } \Delta_i = \min(c_{i+1} - c_i, c_i - c_{i-1}). \tag{3.4}$$

Additionally, the signal needs to be large enough to be detectable by this bandwidth, i.e.

$$\frac{\|\mathbf{d}_i\|^2\, h}{\log\left(\frac{T}{h}\right)} \to \infty. \tag{3.5}$$

Combining (3.4) and (3.5) shows that – with an appropriate bandwidth h – changes are detectable as soon as

$$\frac{\|\mathbf{d}_i\|^2\, \Delta_i}{\log\left(\frac{T}{\Delta_i}\right)} \to \infty. \tag{3.6}$$

In 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 that cannot be improved (see Proposition 1 of Arias-Castro et al. (2011)).

The assumption on the distance of the first and last change point to the boundary of the process in (3.4) can be relaxed as no boundary effects can occur there.

3.2 Change point estimators

The MOSUM statistic $\mathbf{M}_t = \mathbf{m}_t + \boldsymbol{\Lambda}_t$ as in (3.3) decomposes into a piecewise linear signal term $\mathbf{m}_t = \mathbf{m}_{t,h,T}$ and a centered noise term $\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}_{t,h,T}$ with

$$\sqrt{2h}\,\mathbf{m}_t = \begin{cases} (h - t + c_i)\,\mathbf{d}_i, & \text{for } c_i < t \le c_i + h, \\ 0, & \text{for } c_i + h < t \le c_{i+1} - h, \\ (h + t - c_{i+1})\,\mathbf{d}_{i+1}, & \text{for } c_{i+1} - h < t \le c_{i+1}, \end{cases} \tag{3.7}$$

$$\sqrt{2h}\,\boldsymbol{\Lambda}_t = \sqrt{2h}\,\boldsymbol{\Lambda}_t(\widetilde{\mathbf{R}}) \tag{3.8}$$

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

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 0 outside h-intervals around the change points. Additionally, the noise term is asymptotically negligible compared to the signal term (see Theorem 3.1 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

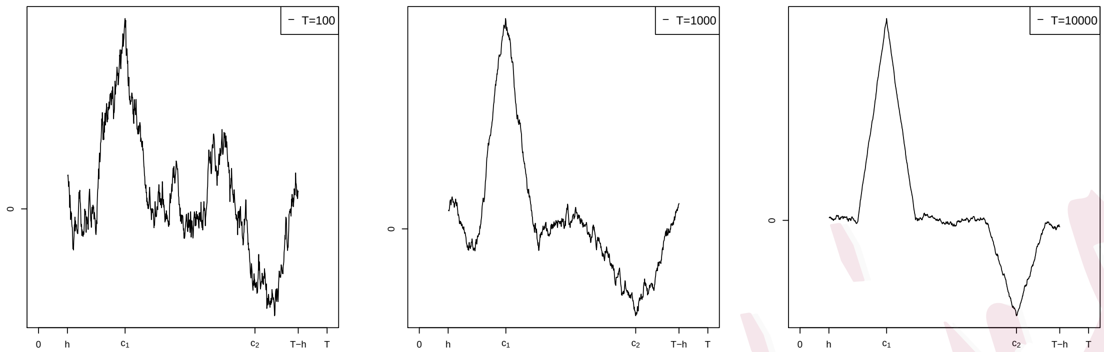


Figure 1: Univariate MOSUM statistic with $T = 100,\, 1\,000,\, 10\,000$ (from left to right), where the noise term (fluctuating around the signal) becomes smaller and smaller relative to the signal term.

a suitable threshold $\beta = \beta_{h,T}$ (see Section 3.3 for a detailed discussion) we define *significant* time points, where a point t^* is *significant* if

$$\mathbf{M}'_{t^*}\,\widehat{\mathbf{A}}_{t^*}^{-1}\,\mathbf{M}_{t^*} \geq \beta. \tag{3.9}$$

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

Assumption 3.2.

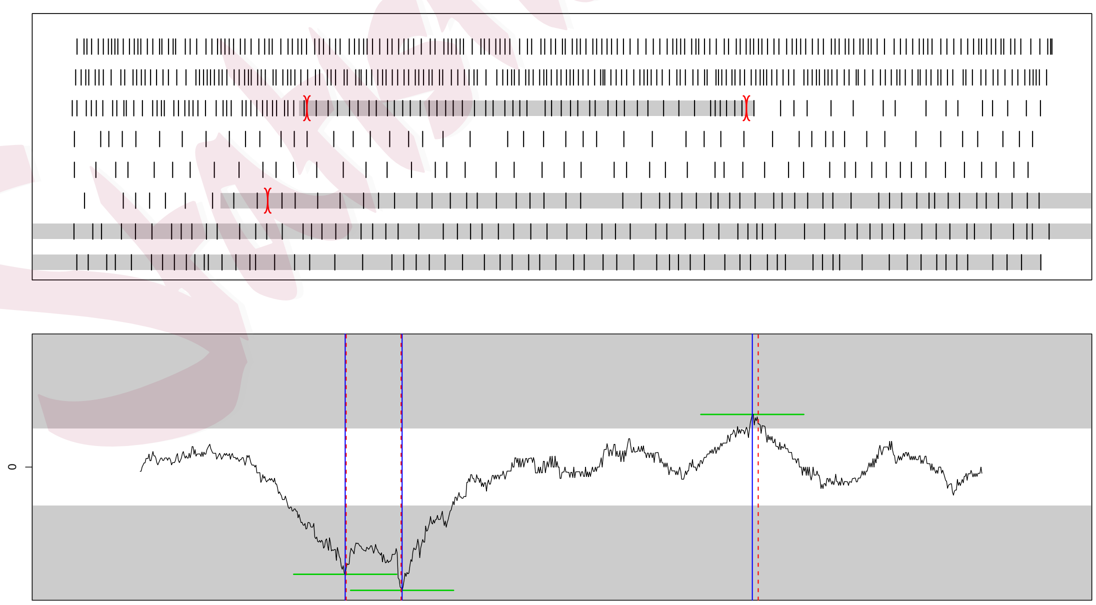
$$\sup_{h\leq t\leq T-h}\left\|\widehat{\mathbf{A}}_{t,T}^{-1}\right\| = O_P\left(1\right), \qquad \sup_{i=1,\ldots,q_T}\ \sup_{|t-c_i|\leq h}\left\|\widehat{\mathbf{A}}_{t,T}\right\| = O_P(1).$$

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

$$\mathbf{\Sigma}_t = \mathbf{\Sigma}_{t,T} = \mathbf{\Sigma}_T^{(c_i)} \tag{3.10}$$

Figure 2: In the upper panel, the observed event times of a univariate renewal process with 3 change points (i.e. 4 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 3.1) 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 ηh-environments around the estimators.



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

$$\sup_{i=1,\ldots,q_T} \sup_{|t-c_i|>h} \left\| \widehat{\boldsymbol{\Sigma}}_{t,T}^{-1/2} - \boldsymbol{\Sigma}_t^{-1/2} \right\| = o_P\left(\left(\log \frac{T}{h} \right)^{-1} \right) \tag{3.11}$$

in addition to the above boundedness assumptions. In particular, this permits local estimators that are consistent only away from change points but contaminated by the change in a local environment thereof. The latter is typically the case for covariance estimators, think e.g. of 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 (due to the continuity of the signal) such that only local extrema of such intervals actually indicate a change point. To define what a local extremum is, 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 ηh-environment, i.e. if

$$t^* = \min\left\{ \underset{t^*-\eta h \leq t \leq t^*+\eta h}{\operatorname{argmax}} \|\mathbf{M}_t\|. \right\} \tag{3.12}$$

The threshold β distinguishes between *significant* and *spurious* local extrema that are purely associated 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 the 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 point. Additionally, the figure shows that (3.4) is required for theoretic considerations only but can be weakened in practice in combination with a suitable η with ηh defining the minimal distance that two MOSUM estimators can have. This is one of the major advantages of the η-criterion based on *significant* local maxima as described here (in comparison to the ϵ-criterion originally investigated by Eichinger and Kirch (2018) in the context of mean changes, see also the discussion in Meier et al. (2019)). Results for the ϵ-criterion can be obtained along the lines of our proofs below. In practical applications, if η is chosen too large, some pairs of change points may be indistinguishable by our procedure. On the other hand, too small a choice of η can lead to an increase in the number of spurious and duplicate estimators, as can be seen in S1 in the Supplementary Material. The latter is not a problem if a post-processing step as in e. g. Cho and Kirch (2021) is applied.

3.3 Threshold selection

The procedure clearly depends on the choice of a threshold $\beta = \beta_{h,T}$ (see (3.9)) that can distinguish between significant and spurious local extrema. The following theorem gives the magnitudes of the signal as well as the noise terms:

Theorem 3.1. *Let the Assumptions 2.1, 3.1 and 3.2 hold.*

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

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

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

(b) *For the noise term it holds for $q_T = 0$, i.e. in the no-change situation,*

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

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

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 3.1 that*

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

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

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

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

The assertions remain true if an estimator for the covariance is used fulfilling (3.11) *uniformly over all* $h \le t \le T - h$.

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

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

To obtain consistency of the estimators, the threshold needs to be small enough to be asymptotically negligible compared to the squared signal term as in Theorem 3.1 (a) to guarantee that every change is detected with asymptotic probability 1. At the same time, the threshold needs to grow faster than the squared noise term in Theorem 3.1 (c) so that false positives occur with asymptotic probability 0. Hence, both conditions are fulfilled under the following assumption:

Assumption 3.3. *The threshold fulfills:*

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

In particular, larger bandwidths h_T lead to a better detectability of the change point, where due to (3.4) an upper bound related to the distance to the neighboring change points applies. This is also confirmed by the simulation results in Table 4 in the Supplementary Material.

The following remark introduces a threshold that has a nice interpretation in connection with change point testing:

Remark 3.1. The threshold is often obtained as the asymptotic α_T-quantile based on the limit result in Theorem 3.1 (b) for some sequence $\alpha_T \to 0$. In this case a choice of

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

similar to Eichinger and Kirch (2018) can replace the slightly stronger lower bound of Assumption 3.3 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 α_T asymptotically related to testing each time point for a possible change. In fact, Theorem 3.1 shows that such a threshold with a constant sequence α yields an asymptotic test at level α which has asymptotic power one by Theorem 4.1. Tests designed for the at-most-one-change as in Hušková and Steinebach (2000), Hušková and Steinebach (2002) often have a better power, but are not as good at localizing change points (see Figure 1 in Cho and Kirch (2020) for an illustration).

4. Consistency of the segmentation procedure

In this section, we show consistency of the above segmentation procedure for both the estimators of the number and 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 (3.12) are consistent for the number and locations of the change points.

Theorem 4.1. *Let Assumptions 2.1, 3.1 – 3.3 hold. Let* $0 < \hat{c}_1 < \ldots < \hat{c}_{\hat{q}_T}$ *be the change point estimators* (3.12)*. Then for any* $\tau > 0$ *it holds*

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

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 further improved:

Assumption 4.1. *(a) For any of the centered processes* $\widetilde{\mathbf{R}}^{(j)}$ *as in* (3.8) *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)*

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

(b) *Let now the upper index θ_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 $D_T > 0$*

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

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 (2021), Proposition 2.1 (c)(ii)).

Remark 4.1. (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 2.1, 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 (2021) for some tight bounds in case of partial sum processes).

(c) Often, there exist forward and backwards invariance principles from some arbitrary starting value θ_i for each regime. This is the case for partial sum processes and for (backward and forward) Markov processes due to the Markov property. For renewal processes, this can be shown along the lines of the original proof for the invariance principle (Csörgö et al. (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 Materials) 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 4.2.

Let Assumptions 2.1, 3.1 – 3.3 in addition to 4.1 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 4.2 (Minimax optimality)**.** We have already mentioned beneath (3.6) that the separation rate given there is minimax optimal (see Proposition 1 of Arias-Castro et al. (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 minimax optimal localization rate for a single change point and in extension 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 et al. (2020)). Consequently, our procedure achieves minimax optimality for a bounded number of change points under weak assumptions (as pointed out in Remark 4.1 (c)). Secondly, 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 (2021)). 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 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 can be done along the same lines. We need the following assumption:

Assumption 4.2. *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*

$$\mathbf{Y}_s^{(1)} = \widetilde{\mathbf{R}}^{(c_i)}_{c_i - h + \frac{s-D}{\|d_i\|^2}} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i - h - \frac{D}{\|d_i\|^2}},$$

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

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

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 3.1 it holds $h\|\mathbf{d}_i\|^2 \to \infty$, such that the distance $h - \frac{2D}{\|d_i\|^2}$ 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 T large enough. Additionally, under weak assumptions such as mixing conditions this independence still holds asymptotically in the sense that $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$, $\mathbf{W}^{(3)}$ are independent.

Functional central limit theorems for these processes follow from invariance principles as in Assumption 2.1 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 4.1 (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 point.

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

Theorem 4.3.

Let Assumptions 2.1, 3.1 – 3.3, 4.1 (a) with $\omega_T = 1$ and 4.2 hold. For $\hat{q}_T < q_T$ define $\hat{c}_i = T$ for $i = \hat{q}_T + 1, \ldots, q_T$. Let

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

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 4.2 holds jointly for all q change points, then the result also holds jointly.

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

Remark 4.3. (a) If $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$, $\mathbf{W}^{(3)}$ are independent which is typically the case (see discussion beneath Assumption 4.2), 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 $\sigma_{(j)}^2 = \mathbf{u}_i' \boldsymbol{\Xi}^{(j)} \mathbf{u}_i$. Usually (see discussion beneath Assumption 4.2) $\sigma_{(21)} = \sigma_{(1)}$ and $\sigma_{(22)} = \sigma_{(3)}$ further simplifying the expression. For some examples such as partial sum processes it holds $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ 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 $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ is typically not fulfilled for renewal processes because the covariance depends on the changing intensity of the process.

(b) If $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$, $\mathbf{W}^{(3)}$ are independent and $\mathbf{M}_t$ in (3.12) is replaced by $\boldsymbol{\Sigma}_t^{-1/2}\mathbf{M}_t$, 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 $\hat{c}_i - c_i$ does only depend on the magnitude of the change $\mathbf{d}_i$ but not on its direction $\mathbf{u}_i$.

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

5. Simulations and Discussions

5.1 Summary of Simulation results

For univariate partial sum processes extensive simulations as well as data examples for MOSUM statistics have been conducted in Eichinger and Kirch (2018); Meier et al. (2019), while results for renewal processes have been obtained by Messer et al. (2014, 2017). In the Supplementary Material S1 we complement these findings by simulations for multivariate renewal processes with multiple changes, where both dependent and independent components as well as several choices for the matrix $\widehat{\mathbf{A}}_t$ as in (3.9) are considered.

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, while 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 a 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 amount 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 (2021), Definition 2.1, or Cho and Kirch (2020), Definition 2.1). In some applications with *multiscale* signals, where frequent large changes as well as 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 (2021) for an information criterion based approach for partial sum processes as well as 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 et al. (2019) may be necessary.

5.2 Discussion and Outlook

In this paper, a data segmentation for multivariate processes with changes in the drift is introduced and corresponding consistency results are obtained 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 – a framework that includes renewal, partial sum and also some diffusion processes with change points.

One drawback of the procedure is the use of a single bandwidth. In practice, the identification of the optimal bandwidth turns out to be rather difficult as pointed out e. g. by Cho and Kirch (2021) and Messer et al. (2014): On the one hand, one wants to choose a large bandwidth in order to have maximum power, while on the other hand, choosing a too large bandwidth may lead to 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 (2021)) no single bandwidth can detect all change points. Therefore, one future topic of interest is the extension of the proposed procedure to a true multiscale setup as in Cho and Kirch (2021).

Acknowledgments

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

References

Aggarwal, R., C. Inclan, and R. Leal (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis 34(1)*, 33–55.

REFERENCES

Antoch, J. and M. Hušková (1999). *Asymptotics, Nonparametrics, and Time Series*, Chapter Estimators of changes, pp. 533–578. Marcel Dekker.

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

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

Brown, E. N. an Kass, R. E. and P. P. Mitra (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience 7(5)*, 456–461.

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

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

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

Cho, H. and C. Kirch (2020). Data segmentation algorithms: Univariate mean change and beyond. *arXiv:2012.12814*.

Cho, H. and C. Kirch (2021+). Two-stage data segmentation permitting multiscale changepoints, heavy tails and dependence. *Annals of the Institute of Statistical Mathematics, to appear*.

Csenki, A. (1979). An invariance principle in k-dimensional extended renewal theory. *J. Appl. Prob. 16*, 567–574.

Csörgö, M. and L. Horvàth (1997). *Limit theorems in change-point analysis* (vol. 18 ed.). John Wiley & Sons Inc.

Csörgö, M., L. Horváth, and J. Steinebach (1987). Invariance principles for renewal processes. *Ann. Probab. 14*, 1441–1460.

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

Fisch, A. T. M., I. A. Eckley, and P. Fearnhead (2018). A linear time method for the detection of point and collective anomalies. *arXiv:1806.01947.*.

Grün, S., M. Diesmann, and A. Aertsen (2002). Unitary events' in multiple single-neuron activity. ii. non-stationary data. *Neural Computation 14(1)*, 81–119.

Grün, S. and S. Rotter (2010). *Analysis of parallel spike trains.* Springer.

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

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

Horváth, L. and G. Rice (2014). Extensions of some classical methods in change point analysis. *Test 23*(2), 219–255.

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

Hušková, M. and A. Slabỳ (2001). Permutation tests for multiple changes. *Kybernetika 37(5)*, 605–622.

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

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

Killick, R., I. A. Eckley, K. Ewans, and P. Jonathan (2010). Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Engineering 37*, 1120–1126.

Kirch, C. and J. Steinebach (2006). Permutation principles for the change analysis of stochastic processes under strong invariance. *Journal of Computational and Applied Mathematics 184*, 64–88.

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

Meier, A., H. Cho, and C. Kirch (2019). mosum: A package for moving sums in change point analysis. *Journal of Statistical Software 97(8)*, 1–42.

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

Messer, M., M. Kirchner, J. Schiemann, J. Roeper, R. Neininger, and G. Schneider (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *The Annals of Applied Statistics 8(4)*, 2027–2067.

Messer, M. and G. Schneider (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 H. Zhang (2012). The screening and ranking algorithm to detect dna copy number variations. *The Annals of Applied Statistics 6*, 1306–1326.

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

Perkel, D. H., G. L. Gernstein, and M. G. P. (1967). Neuronal spike trains and stochastic point processes. ii. simultaneous spike trains. *Biophysical Journal 7(4)*, 419–440.

Schneider, G. (2008). Messages of oscillatory correlograms - a spike-train model. *Neural Computation 20(5)*, 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 V. R. Eastwood (1996). Extreme value asymptotics for multivariate renewal processes. *Journal of multivariate analysis 56(2)*, 284–302.

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

Wang, D., Y. Yu, and A. Rinaldo (2020). Univariate mean change point detection: Penalization, cusum and optimality. *Electronic Journal of Statistics 14*(1), 1917–1961.

Yau, C. Y. and Z. Zhao (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*(4), 895–916.

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

E-mail: claudia.kirch@ovgu.de

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

E-mail: philipp.klein@ovgu.de