

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Nonparametric data segmentation in multivariate time series
via joint characteristic functions**

By E. T. MCGONIGLE

School of Mathematics, University of Bristol, Bristol BS8 1UG, U.K.
School of Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, U.K.
e.t.mcgonigle@soton.ac.uk

AND H. CHO

School of Mathematics, University of Bristol, Bristol BS8 1UG, U.K.
haeran.cho@bristol.ac.uk

SUMMARY

Modern time series data often exhibit complex dependence and structural changes which are not easily characterised by shifts in the mean or model parameters. We propose a nonparametric data segmentation methodology for multivariate time series termed NP-MOJO. By considering joint characteristic functions between the time series and its lagged values, NP-MOJO is able to detect change points in the marginal distribution, but also those in possibly non-linear serial dependence, all without the need to pre-specify the type of changes. We show the theoretical consistency of NP-MOJO in estimating the total number and the locations of the change points, and demonstrate the good performance of NP-MOJO against a variety of change point scenarios. We further demonstrate its usefulness in applications to seismology and economic time series.

Some key words: change point detection, joint characteristic function, moving sum, multivariate time series, nonparametric

1. INTRODUCTION

Change point analysis has been an active area of research for decades, dating back to [Page \(1954\)](#). Literature on change point detection continues to expand rapidly due to its prominence in numerous applications, including biology ([Jewell et al., 2020](#)), financial analysis ([Lavielle & Teysiere, 2007](#)) and environmental sciences ([Carr et al., 2017](#)). Considerable efforts have been made for developing computationally and statistically efficient methods for data segmentation, a.k.a. multiple change point detection, in the mean of univariate data under independence ([Killick et al., 2012](#); [Frick et al., 2014](#); [Fryzlewicz, 2014](#)) and permitting serial dependence ([Tecuapetla-Gómez & Munk, 2017](#); [Dette et al., 2020](#); [Cho & Kirch, 2022](#); [Cho & Fryzlewicz, 2023](#)). There also exist methods for detecting changes in the covariance ([Aue et al., 2009](#); [Wang et al., 2021](#)), parameters under linear regression ([Bai & Perron, 1998](#); [Xu et al., 2024](#)) or other models ([Fryzlewicz & Subba Rao, 2014](#); [Safikhani & Shojaie, 2022](#)) in fixed and high dimensions. For an overview, see [Truong et al. \(2020\)](#) and [Cho & Kirch \(2023+\)](#).

Any departure from distributional assumptions such as independence and Gaussianity tends to result in poor performance of change point algorithms. Furthermore, it may not be realistic to assume any knowledge of the type of change point that occurs, or to make parametric assumptions on the data generating process, for time series that possess complex structures and are observed

over a long period. Searching for change points in one property of the data (e.g. mean), when the time series instead undergoes changes in another (e.g. variance), may lead to misleading conclusions and inference on such data. Therefore, it is desirable to develop flexible, nonparametric change point detection techniques that are applicable to detect general changes in the underlying distribution of serially dependent data.

There are several strategies for the nonparametric change point detection problem, such as those based on the empirical cumulative distribution and density functions (Carlstein, 1988; Zou et al., 2014; Haynes et al., 2017; Madrid Padilla et al., 2021; Vanegas et al., 2022; Madrid Padilla et al., 2022, 2023), kernel transforms of the data (Harchaoui et al., 2009; Celisse et al., 2018; Arlot et al., 2019; Li et al., 2019) or U -statistics measuring the ‘energy’-based distance between different distributions (Matteson & James, 2014; Chakraborty & Zhang, 2021; Boniece et al., 2023). There also exist graph-based methods applicable to non-Euclidean data (Chen & Zhang, 2015; Chu & Chen, 2019). All these methods can only detect changes in the marginal distribution of the data and apart from Madrid Padilla et al. (2023), assume serial independence. We also mention Cho & Fryzlewicz (2012), Preuß et al. (2015) and Korkas & Fryzlewicz (2017) where the problem of detecting changes in the second-order structure is addressed, but their methods do not have power against changes in non-linear dependence.

We propose NP-MOJO, a NonParametric MOving sum procedure for detecting changes in the JOint characteristic function, which detects multiple changes in serial, possibly non-linear dependence as well as marginal distributions of a multivariate time series $\{X_t\}_{t=1}^n$. We adopt a moving sum procedure to scan the data for multiple change points. The moving sum methodology has successfully been applied to a variety of change point testing (Chu et al., 1995; Huskova & Slaby, 2001) and data segmentation problems (Eichinger & Kirch, 2018). Here, we combine it with a detector statistic carefully designed to detect changes in complex dependence structure beyond those detectable from considering the marginal distribution only. Specifically, we utilise an energy-based distributional discrepancy that measures any change in the joint characteristic function of the time series at some lag $\ell \geq 0$, which allows for detecting changes in the joint distribution of $(X_t, X_{t+\ell})$ beyond the changes in their linear dependence. To the best of our knowledge, NP-MOJO is the first nonparametric methodology which is able to detect changes in non-linear serial dependence in multivariate time series.

We establish that NP-MOJO achieves consistency in estimating the number and locations of the change points for a given lag, providing convergence rates for the change point location estimators, and propose a methodology that extends this desirable property of single-lag NP-MOJO to multiple lags. Combined with a dependent multiplier bootstrapping procedure, NP-MOJO and its multi-lag extension perform well across a wide range of change point scenarios in simulations and real data applications. Accompanying R software implementing NP-MOJO is available as the R package CptNonPar (McGonigle & Cho, 2023) on CRAN.

2. MODEL AND MEASURE OF DISCREPANCY

We observe a multivariate time series $\{X_t\}_{t=1}^n$ of (finite) dimension p , where

$$X_t = \sum_{j=0}^q X_t^{(j)} \cdot \mathbb{I}\{\theta_j + 1 \leq t \leq \theta_{j+1}\} \quad (1)$$

with $X_t = (X_{t1}, \dots, X_{tp})^\top$ and $0 = \theta_0 < \theta_1 < \dots < \theta_q < \theta_{q+1} = n$. For each sequence $\{X_t^{(j)} : t \geq 1\}$, $j = 0, \dots, q$, there exists an $\mathbb{R}^p$ -valued measurable function $g^{(j)}(\cdot) = (g_1^{(j)}(\cdot), \dots, g_p^{(j)}(\cdot))^\top$ such that $X_t^{(j)} = g^{(j)}(\mathcal{F}_t)$ with $\mathcal{F}_t = \sigma(\varepsilon_s : s \leq t)$, and i.i.d. random ele-

ments ε_t . We assume that $g^{(j-1)} \neq g^{(j)}$ for all $j = 1, \dots, q$, such that under the model (1), the time series undergoes q change points at locations $\Theta = \{\theta_1, \dots, \theta_q\}$, with the notational convention that $\theta_0 = 0$ and $\theta_{q+1} = n$. That is, $\{X_t\}_{t=1}^n$ consists of $q+1$ stationary segments where the j -th segment is represented in terms of a segment-dependent ‘output’ $g^{(j)}(\mathcal{F}_t)$, with the common ‘input’ $\mathcal{F}_t$ shared across segments such that dependence across the segments is not ruled out. Each segment has a non-linear Wold representation as defined by Wu (2005); this representation includes commonly adopted time series models including ARMA and GARCH processes.

Denote the inner product of two vectors x and y by $\langle x, y \rangle = x^\top y$ and ι the imaginary unit with $\iota^2 = -1$. At some integer ℓ , define the joint characteristic function of $\{X_t^{(j)}\}_{t \in \mathbb{Z}}$ at lag ℓ , as

$$\phi_\ell^{(j)}(u, v) = \mathbb{E} \left\{ \exp \left(\iota \langle u, X_1^{(j)} \rangle + \iota \langle v, X_{1+\ell}^{(j)} \rangle \right) \right\}, \quad 0 \leq j \leq q.$$

We propose to measure the size of changes between adjacent segments under (1), using an ‘energy-based’ distributional discrepancy given by

$$d_\ell^{(j)} = \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \left| \phi_\ell^{(j)}(u, v) - \phi_\ell^{(j-1)}(u, v) \right|^2 w(u, v) du dv, \quad 1 \leq j \leq q, \quad (2)$$

where $w(u, v)$ is a positive weight function for which the above integral exists. For given lag $\ell \geq 0$, the quantity $d_\ell^{(j)}$ measures the weighted L_2 -norm of the distance between the lag ℓ joint characteristic functions of $\{X_t^{(j-1)}\}_{t \in \mathbb{Z}}$ and $\{X_t^{(j)}\}_{t \in \mathbb{Z}}$. A discrepancy measure of this form is a natural choice for nonparametric data segmentation, since:

LEMMA 1. For any $\ell \geq 0$, $d_\ell^{(j)} = 0$ if and only if $(X_1^{(j)}, X_{1+\ell}^{(j)}) \stackrel{d}{=} (X_1^{(j-1)}, X_{1+\ell}^{(j-1)})$.

Lemma 1 extends the observation made in Matteson & James (2014) about the correspondence between the characteristic function and marginal distribution. It shows that by considering the joint characteristic functions $\phi_\ell^{(j)}(u, v)$ at multiple lags $\ell \geq 0$, the discrepancy $d_\ell^{(j)}$ is able to capture changes in the serial dependence as well as those in the marginal distribution of $\{X_t\}_{t=1}^n$.

The following lemma lists some choices of the weight function $w(u, v)$ and the associated representation of $d_\ell^{(j)}$ as the kernel-based discrepancy between $Y_t^{(j)} = (X_t^{(j)}, X_{t+\ell}^{(j)})$ and $Y_t^{(j-1)}$, extending the observation made in Matteson & James (2014) for the setting where a sequence of independent observations are undergoing changes in the marginal distribution. Let $\|x\|$ denote the Euclidean norm of a vector x , and define $\tilde{Y}_t^{(j)} = (\tilde{X}_t^{(j)}, \tilde{X}_{t+\ell}^{(j)})$ where $\tilde{X}_t^{(j)} = g^{(j)}(\tilde{\mathcal{F}}_t)$ with $\tilde{\mathcal{F}}_t = \sigma(\tilde{\varepsilon}_s : s \leq t)$ and $\tilde{\varepsilon}_t$ is an independent copy of ε_t .

LEMMA 2. (i) For any $\beta > 0$, suppose that $d_\ell^{(j)}$ in (2) is obtained with respect to the following weight function:

$$w_1(u, v) = C_1(\beta, p)^{-2} \exp \left\{ -\frac{1}{2\beta^2} (\|u\|^2 + \|v\|^2) \right\} \quad \text{with } C_1(\beta, p) = (2\pi)^{p/2} \beta^p.$$

Then, the function $h_1 : \mathbb{R}^{2p} \times \mathbb{R}^{2p} \rightarrow [0, 1]$ defined as $h_1(x, y) = \exp(-\beta^2 \|x - y\|^2 / 2)$ for $x, y \in \mathbb{R}^{2p}$, satisfies

$$d_\ell^{(j)} = \mathbb{E} \left\{ h_1 \left(Y_1^{(j)}, \tilde{Y}_1^{(j)} \right) \right\} + \mathbb{E} \left\{ h_1 \left(Y_1^{(j-1)}, \tilde{Y}_1^{(j-1)} \right) \right\} - 2\mathbb{E} \left\{ h_1 \left(\tilde{Y}_1^{(j)}, Y_1^{(j-1)} \right) \right\}.$$

4

E. T. MCGONIGLE AND H. CHO

(ii) For any $\delta > 0$, suppose that $d_\ell^{(j)}$ is obtained with

$$w_2(u, v) = C_2(\delta, p)^{-2} \prod_{s=1}^p u_s^2 v_s^2 \exp \left\{ -\delta (u_s^2 + v_s^2) \right\} \quad \text{with} \quad C_2(\delta, p) = \frac{\pi^{p/2}}{2^p \delta^{3p/2}}.$$

Then, the function $h_2 : \mathbb{R}^{2p} \times \mathbb{R}^{2p} \rightarrow [-2e^{-2/3}, 1]$ defined as

$$h_2(x, y) = \prod_{r=1}^{2p} \frac{\{2\delta - (x_r - y_r)^2\} \exp \left\{ -\frac{1}{4\delta} (x_r - y_r)^2 \right\}}{2\delta}$$

for $x = (x_1, \dots, x_{2p})^\top$ and $y = (y_1, \dots, y_{2p})^\top$, satisfies

$$d_\ell^{(j)} = \mathbb{E} \left\{ h_2 \left(Y_1^{(j)}, \tilde{Y}_1^{(j)} \right) \right\} + \mathbb{E} \left\{ h_2 \left(Y_1^{(j-1)}, \tilde{Y}_1^{(j-1)} \right) \right\} - 2\mathbb{E} \left\{ h_2 \left(\tilde{Y}_1^{(j)}, Y_1^{(j-1)} \right) \right\}.$$

Lemma 2 is a special case of Bochner's Theorem applied to the chosen weight functions, see for example Section 5.3 of [Sejdinovic et al. \(2013\)](#). The weight function w_1 is commonly referred to as the Gaussian weight function. Both w_1 and w_2 are unit integrable and separable in their arguments, such that $d_\ell^{(j)}$ is well-defined due to the boundedness of the characteristic function. We provide an alternative weight function in Appendix A.2 and also refer to [Fan et al. \(2017\)](#) for other suitable choices.

Remark 1. From Lemma 2, $d_\ell^{(j)}$ can be viewed as the squared maximum mean discrepancy (MMD) on a suitably defined reproducing kernel Hilbert space with the associated kernel function; see Lemma 6 of [Gretton et al. \(2012\)](#) and Section 2.6 of [Celisse et al. \(2018\)](#). We also note the literature on the (auto)distance correlation for measuring and testing dependence in multivariate ([Székely et al., 2007](#)) and time series ([Zhou, 2012](#); [Fokianos & Pitsillou, 2017](#); [Davis et al., 2018](#)) settings.

Remark 2. In Model (1) (and in our theoretical results), the dimension p of the time series is assumed fixed. We would expect practical performance to deteriorate with increasing dimension since we use an energy-based method. For example, when the time series undergoes a change in both mean and variance, the pre- and post-change segments of the time series can be separated into an "inner layer" and "outer layer" based on their pairwise Euclidean distances. However, as [Chen & Friedman \(2017\)](#) note, "data points in the outer layer find themselves to be closer to points in the inner layer than other points in the outer layer", due to the curse of dimensionality. See for example [Ramdas et al. \(2015\)](#) or Section 2.2 of [Chu & Chen \(2019\)](#) for further discussion.

3. METHODOLOGY

3.1. The NP-MOJO procedure

In this section we describe our proposed nonparametric moving sum procedure for detecting changes in the joint characteristic function, henceforth referred to as NP-MOJO. The identities given in Lemma 2 allow for the efficient computation of the statistics approximating $d_\ell^{(j)}$ and their weighted sums, which forms the basis for the NP-MOJO procedure for detecting multiple change points from a multivariate time series $\{X_t\}_{t=1}^n$ under the model (1). Throughout, we present the procedure with a generic kernel h associated with some weight function w . We first introduce NP-MOJO for the problem of detecting changes in the joint distribution of $Y_t = (X_t, X_{t+\ell})$ at a given lag $\ell \geq 0$, and extend it to the multi-lag problem in Section 3.3.

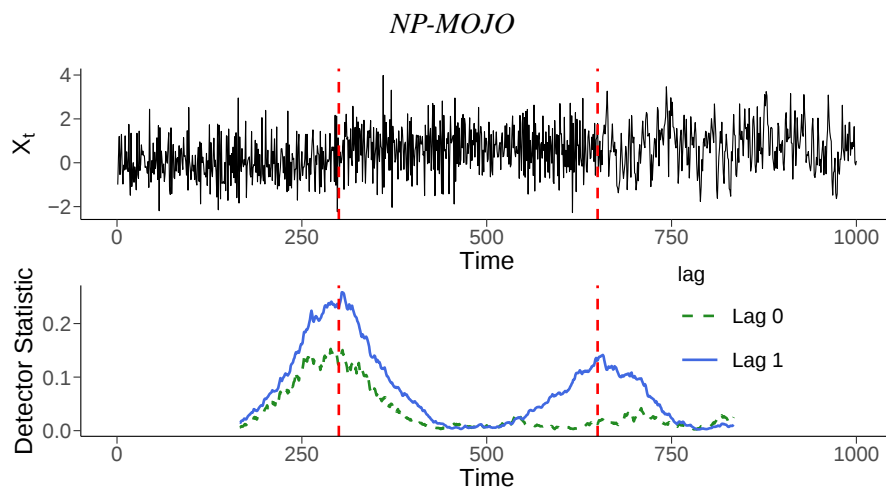


Fig. 1: Top: time series of length $n = 1000$ with change points $\theta_1 = 300$ and $\theta_2 = 650$ (vertical dashed lines), see Example 1. Bottom: corresponding detector statistics $T_\ell(G, k)$ computed at lags $\ell = 0$ (dashed) and $\ell = 1$ (solid).

For fixed bandwidth $G \in \mathbb{N}$, NP-MOJO scans the data using a detector statistic computed on neighbouring moving windows of length G , which approximates the discrepancy between the local joint characteristic functions of the corresponding windows measured analogously as in (2). Specifically, the detector statistic at location k is given by the following two-sample V -statistic:

$$T_\ell(G, k) = \frac{1}{(G - \ell)^2} \left\{ \sum_{s, t=k-G+1}^{k-\ell} h(Y_s, Y_t) + \sum_{s, t=k+1}^{k+G-\ell} h(Y_s, Y_t) - 2 \sum_{s=k-G+1}^{k-\ell} \sum_{t=k+1}^{k+G-\ell} h(Y_s, Y_t) \right\}$$

for $k = G, \dots, n - G$, as an estimator of the local discrepancy measure

$$\mathcal{D}_\ell(G, k) = \sum_{j=0}^q \left(\frac{G - \ell - |k - \theta_j|}{G - \ell} \right)^2 d_\ell^{(j)} \cdot \mathbb{I}\{|k - \theta_j| \leq G - \ell\}. \quad (3)$$

At given k , the statistic $T_\ell(G, k)$ measures the difference in the distribution of Y_t over the disjoint intervals of length $G - \ell$ around k , and satisfies

$$\mathbb{E}\{T_\ell(G, k)\} = \mathcal{D}_\ell(G, k) + O(G^{-1/2}). \quad (4)$$

We have $\mathcal{D}_\ell(G, k) = 0$ when the section of the data $\{X_t, |t - k| \leq G - \ell\}$ does not undergo a change and accordingly, $T_\ell(G, k)$ is expected to be close to zero. On the other hand, if $|k - \theta_j| < G - \ell$, then $\mathcal{D}_\ell(G, k)$ increases and then decreases around θ_j with a local maximum at $k = \theta_j$. The statistic $T_\ell(G, k)$ is expected to behave similarly: in particular, at any change point location θ_j , we have that $\mathbb{E}\{T_\ell(G, \theta_j)\} = d_\ell^{(j)} + O(G^{-1/2})$ (see Lemma D.4 in the supplementary material for further information). We illustrate this using the following example.

Example 1. A univariate time series $\{X_t\}_{t=1}^n$ of length $n = 1000$ is generated as $X_t = \mu_t + \varepsilon_t$, where $\mu_t = 0.7 \cdot \mathbb{I}\{t > \theta_1\}$ and $\varepsilon_t = \varepsilon_t^{(1)} \cdot \mathbb{I}\{t < \theta_2\} + \varepsilon_t^{(2)} \cdot \mathbb{I}\{t \geq \theta_2\}$, with $\theta_1 = 300$ and $\theta_2 = 650$. Each $\varepsilon_t^{(j)}$ is an autoregressive (AR) process of order 1: $\varepsilon_t^{(1)} = 0.5\varepsilon_{t-1}^{(1)} + W_t$ and $\varepsilon_t^{(2)} = -0.5\varepsilon_{t-1}^{(2)} + W_t$, where $\{W_t\}_{t \in \mathbb{Z}}$ is a white noise process with $\text{var}(W_t) = \sqrt{1 - 0.5^2}$. This choice leads to $\text{var}(X_t) = 1$ for all t , see the top panel of Figure 1 for a realisation. Then, the mean shift at θ_1 is detectable at all lags while the autocorrelation change at θ_2 is detectable at odd lags only, i.e. $d_\ell^{(2)} = 0$ for even $\ell \geq 0$. The bottom panel of Figure 1 plots $T_\ell(G, k)$, $G \leq k \leq n - G$,

6 E. T. MCGONIGLE AND H. CHO

computed using kernel h_2 in Lemma 2 (ii) with $G = 166$. At lag $\ell = 0$, the detector statistic forms a prominent peak around θ_1 but it is flat around θ_2 ; at lag $\ell = 1$, the statistic $T_1(G, k)$ forms local maxima around both θ_j , $j = 1, 2$.

Based on these observations, it is reasonable to detect and locate the change points in the joint distribution of $(X_t, X_{t+\ell})$ as significant local maximisers of $T_\ell(G, k)$. We adopt the selection criterion, first considered by Eichinger & Kirch (2018) in the context of detecting mean shifts from univariate time series, for simultaneous estimation of multiple change points. For some fixed constant $\eta \in (0, 1)$ and a threshold $\zeta_\ell(n, G) > 0$, we identify any local maximiser of $T_\ell(G, k)$, say $\hat{\theta}$, which satisfies

$$T_\ell(G, \hat{\theta}) > \zeta_\ell(n, G) \quad \text{and} \quad \hat{\theta} = \arg \max_{k: |k - \hat{\theta}| \leq \eta G} T_\ell(G, k). \quad (5)$$

That is, $\hat{\theta}$ is declared a change point if it is a local maximiser of $T_\ell(G, k)$ over a sufficiently large interval of size ηG , at which the threshold $\zeta_\ell(n, G)$ is exceeded. We denote the set of such estimators fulfilling (5) by $\hat{\Theta}_\ell$ with $\hat{q}_\ell = |\hat{\Theta}_\ell|$. The choice of $\zeta_\ell(n, G)$ is discussed in Section 3.4.

3.2. Theoretical properties

For some finite integer $\ell \geq 0$, we define the index set of the change points detectable at lag ℓ as $\mathcal{I}_\ell = \{1 \leq j \leq q : d_\ell^{(j)} \neq 0\}$, and denote its cardinality by $q_\ell = |\mathcal{I}_\ell| \leq q$. Not all change points are detectable at all lags, see Example 1 where we have $\mathcal{I}_0 = \{1\}$ and $\mathcal{I}_1 = \{1, 2\}$. In this section, we show that the single-lag NP-MOJO described in Section 3.1 consistently estimates the total number q_ℓ and the locations $\{\theta_j, j \in \mathcal{I}_\ell\}$ of the change points detectable at lag ℓ , by $\hat{\Theta}_\ell$.

Writing $g_{ti}(\cdot) = \sum_{j=0}^q g_i^{(j)}(\cdot) \mathbb{I}\{\theta_j + 1 \leq t \leq \theta_{j+1}\}$, define $X_{ti, \{t-s\}} = g_{ti}(\mathcal{F}_{t, \{t-s\}})$, where $F_{t, \{t-s\}} = \sigma(\dots, \varepsilon_{t-s-1}, \tilde{\varepsilon}_{t-s}, \varepsilon_{t-s+1}, \dots, \varepsilon_t)$ is a coupled version of $\mathcal{F}_t$ with ε_{t-s} replaced by its independent copy $\tilde{\varepsilon}_{t-s}$. For a random variable Z and $\nu > 0$, let $\|Z\|_\nu = \{\mathbb{E}(|Z|^\nu)\}^{1/\nu}$. Analogously as in Xu et al. (2024), we define the element-wise functional dependence measure and its cumulative version as

$$\delta_{s, \nu, i} = \sup_{t \in \mathbb{Z}} \|X_{ti} - X_{ti, \{t-s\}}\|_\nu \quad \text{and} \quad \Delta_{m, \nu} = \max_{1 \leq i \leq p} \sum_{s=m}^{\infty} \delta_{s, \nu, i}, \quad m \in \mathbb{Z}. \quad (6)$$

Then, we make the following assumptions on the degree of serial dependence in $\{X_t\}_{t=1}^n$.

Assumption 1. There exist some constants $C_F, C_X \in (0, \infty)$ and $\gamma_1 \in (0, 2)$ such that

$$\sup_{m \geq 0} \exp(C_F m^{\gamma_1}) \Delta_{m, 2} \leq C_X.$$

Assumption 2. The time series $\{X_t\}_{t=1}^n$ is continuous and β -mixing with $\beta(m) \leq C_\beta m^{-\gamma_2}$ for some constants $C_\beta \in (0, \infty)$ and $\gamma_2 \geq 1$, where

$$\beta(m) = \sup_{t \in \mathbb{Z}} \left(\sup \frac{1}{2} \sum_{r=1}^R \sum_{s=1}^S |\text{pr}(A_r \cap B_s) - \text{pr}(A_r)\text{pr}(B_s)| \right).$$

Here, the inner supremum is taken over all pairs of finite partitions $\{A_1, \dots, A_R\}$ of $\mathcal{F}_t = \sigma(\varepsilon_u, u \leq t)$ and $\{B_1, \dots, B_S\}$ of $\sigma(\varepsilon_u, u \geq t+m)$.

Assumptions 1 and 2 require the serial dependence in $\{X_t\}_{t=1}^n$, measured by $\Delta_{m, 2}$ and $\beta(m)$, to decay exponentially, and both are met by a range of linear and non-linear processes (Wu, 2005; Mokkadem, 1988). Under Assumption 1, we have $\|X_{it}\|_2 < \infty$ for all i and t . Assumption 1 is required for bounding $T_\ell(G, k) - \mathbb{E}\{T_\ell(G, k)\}$ uniformly over k , while Assumption 2 is used

for controlling the bias $\mathbb{E}\{T_\ell(G, k)\} - \mathcal{D}_\ell(G, k)$ which is attributed to serial dependence. A condition similar to Assumption 2 is often found in the time series literature making use of distance correlations, see e.g. Davis et al. (2018) and Yousuf & Feng (2022). Under the stronger assumption that $\{X_t^{(j)}\}$ and $\{X_t^{(j+1)}\}$ are independent, we can derive the analogous results as those presented in Theorems 1 and 3, under Assumption 2 only.

Assumption 3. The kernel function h is symmetric and bounded, and can be written as $h(x, y) = h_0(x - y)$ for some function $h_0 : \mathbb{R}^{2p} \rightarrow \mathbb{R}$ that is Lipschitz continuous with respect to $\|\cdot\|$ with Lipschitz constant $C_h \in (0, \infty)$.

Assumption 3 on the kernel function h is met by h_1 and h_2 introduced in Lemma 2, with constants C_h bounded by $\beta e^{-1/2}$ and $2\sqrt{2}p^{3/2}\delta^{-1/2}$, respectively.

Assumption 4. (i) $G^{-1} \log(n) \rightarrow 0$ as $n \rightarrow \infty$ while $\min_{0 \leq j \leq q} (\theta_{j+1} - \theta_j) \geq 2G$.
(ii) $\sqrt{G/\log(n)} \min_{j \in \mathcal{I}_\ell} d_\ell^{(j)} \rightarrow \infty$.

Recall that $\mathcal{I}_\ell$ denotes the index set of detectable change points at lag ℓ , i.e. $d_\ell^{(j)} > 0$ iff $j \in \mathcal{I}_\ell$. However, this definition of detectability is too weak to ensure that all θ_j , $j \in \mathcal{I}_\ell$, are detected by NP-MOJO with high probability at lag ℓ , since we do not rule out the case of local changes where $d_\ell^{(j)} \rightarrow 0$. Consider Example 1: the change in the autocorrelations results in $d_\ell^{(2)} > 0$ for all odd ℓ but the size of change is expected to decay exponentially fast as ℓ increases. Assumption 4 allows for local changes provided that $\sqrt{G/\log(n)} d_\ell^{(j)}$ diverges sufficiently fast. Assumption 4 (i) on the minimum spacing of change points, is commonly imposed in the literature on change point detection using moving window-based procedures. Assumption 4 does not rule out $G/n \rightarrow 0$ and permits the number of change points q to increase in n . We discuss the selection of bandwidth in Section 4.

THEOREM 1. Let Assumptions 1, 2, 3 and 4 hold and $\ell \geq 0$ be a finite integer, and set the threshold as $\zeta_\ell(n, G) = c_\zeta \sqrt{\log(n)}/G$ for some constant $c_\zeta > 0$. Then, there exists $c_0 > 0$, depending only on C_F , C_X , γ_1 , C_β , γ_2 and p , such that as $n \rightarrow \infty$,

$$\text{pr} \left(\hat{q}_\ell = q_\ell, \max_{j \in \mathcal{I}_\ell} \min_{\hat{\theta} \in \hat{\Theta}_\ell} d_\ell^{(j)} |\hat{\theta} - \theta_j| \leq c_0 \sqrt{G \log(n)} \right) \rightarrow 1.$$

Theorem 1 establishes that, for given ℓ , NP-MOJO correctly estimates the total number and the locations of the change points detectable at lag ℓ (including the no change case where $q_\ell = 0$). In particular, by Assumption 4, the change point estimators satisfy

$$\min_{\hat{\theta} \in \hat{\Theta}_\ell} |\hat{\theta} - \theta_j| = O_P \left\{ (d_\ell^{(j)})^{-1} \sqrt{G \log(n)} \right\} = o_P \{ \min(\theta_j - \theta_{j-1}, \theta_{j+1} - \theta_j) \} \text{ for all } j \in \mathcal{I}_\ell,$$

i.e. the change point estimators converge to the true change point locations in the rescaled time. Further, the rate of estimation is inversely proportional to the size of change $d_\ell^{(j)}$, such that the change points associated with larger $d_\ell^{(j)}$ are estimated with better accuracy. Also making use of the energy-based distributional discrepancy, Matteson & James (2014) establish the consistency of their proposed E-Divisive method for detecting changes in (marginal) distribution under independence. In addition to detection consistency, we further derive the rate of estimation for NP-MOJO which is applicable to detect changes in complex time series dependence besides those in marginal distribution, in broader situations permitting serial dependence.

Compared to the optimal rate of estimation known for some parametric change point problems, the rate reported in Theorem 1 is sub-optimal due to the bias of order $O(G^{-1/2})$ (see (4)) in U -

and V -statistics in the presence of serial dependence. In the next theorem, we relax Assumptions 1 and 2 to serial independence, and derive a faster rate of estimation for detecting change points in the marginal distribution (namely θ_j , $j \in \mathcal{I}_0 = \{1, \dots, q_0\}$) using NP-MOJO with lag $\ell = 0$.

THEOREM 2. *Let Assumptions 3 and 4 hold, the latter with $\ell = 0$, and assume that $\{X_t\}_{t=1}^n$ are independent over time, so that $q_0 = q$. Set the threshold as $\zeta(n, G) = c_\zeta \sqrt{\log(n)/G}$ for some constant $c_\zeta > 0$. Then, there exists $c_0 > 0$ depending on p , such that as $n \rightarrow \infty$,*

$$\Pr \left(\hat{q} = q, \max_{1 \leq j \leq q} \min_{\hat{\theta} \in \hat{\Theta}_0} (d_0^{(j)})^2 |\hat{\theta} - \theta_j| \leq c_0 \log(n) \right) \rightarrow 1.$$

3.3. Multi-lag extension of NP-MOJO

In this section, we address the problem of combining the results of the NP-MOJO procedure when it is applied with multiple lags. Let $\mathcal{L} \subset \mathbb{N}_0 = \{0, 1, \dots\}$ denote a (finite) set of non-negative integers. Recall that given $\ell \in \mathcal{L}$, NP-MOJO returns a set of change points estimators $\hat{\Theta}_\ell$. Denote the union of change point estimators over all lags $\mathcal{L}$ by $\tilde{\Theta} = \bigcup_{\ell \in \mathcal{L}} \hat{\Theta}_\ell = \{\hat{\theta}_j, 1 \leq j \leq Q : \hat{\theta}_1 < \dots < \hat{\theta}_Q\}$, and denote by $\mathbb{T}(\tilde{\theta}) = \max_{\ell \in \mathcal{L}} T_\ell(G, \tilde{\theta})$ the maximum detector statistic at $\tilde{\theta}$ across all $\ell \in \mathcal{L}$. We propose to find a set of the final change point estimators $\hat{\Theta} \subset \tilde{\Theta}$ by taking the following steps; we refer to this procedure as multi-lag NP-MOJO.

Step 0. Set $\hat{\Theta} = \emptyset$ and select a constant $c \in (0, 2]$.

Step 1. Set $\tilde{\Theta}_1 = \tilde{\Theta}$ and $m = 1$. Iterate Steps 2–4 for $m = 1, 2, \dots$, while $\tilde{\Theta}_m \neq \emptyset$.

Step 2. Let $\tilde{\theta}_m = \min \tilde{\Theta}_m$ and identify $C_m = \{\tilde{\theta} \in \tilde{\Theta}_m : \tilde{\theta} - \tilde{\theta}_m < cG\}$.

Step 3. Identify $\hat{\theta}_m = \arg \max_{\tilde{\theta} \in C_m} \mathbb{T}(\tilde{\theta})$; if there is a tie, we arbitrarily break it.

Step 4. Add $\hat{\theta}_m$ to $\hat{\Theta}$ and update $m \leftarrow m + 1$ and $\tilde{\Theta}_m = \tilde{\Theta}_{m-1} \setminus C_{m-1}$.

At iteration m of the multi-lag NP-MOJO, Step 2 identifies the minimal element from the current set of candidate change point estimators $\tilde{\Theta}_m$, and a cluster of estimators C_m whose elements are expected to detect the identical change points from multiple lags. Then, Step 3 finds an estimator $\hat{\theta} \in C_m$, which is associated with the largest detector statistic at some lag, and it is added to the set of final estimators. This choice is motivated by Theorem 1, which shows each θ_j is estimated with better accuracy at the lag associated with the largest change in the lagged dependence (measured by $d_\ell^{(j)}$). Iterating these steps until all the elements of $\tilde{\Theta}$ are either added to $\hat{\Theta}$ or discarded, we obtain the set of final change point estimators.

We define a subset of $\mathcal{L}$ containing the lags at which the j -th change point is detectable, as $\mathcal{L}^{(j)} = \{\ell \in \mathcal{L} : d_\ell^{(j)} \neq 0\}$. Re-visiting Example 1, when we set $\mathcal{L} = \{0, 1\}$, it follows that $\mathcal{L}^{(1)} = \{0, 1\}$ and $\mathcal{L}^{(2)} = \{1\}$. To establish the consistency of the multi-lag NP-MOJO, we formally assume that all changes points are detectable at some lag $\ell \in \mathcal{L}$.

Assumption 5. For $\mathcal{L} \subset \mathbb{N}_0$ with $L = |\mathcal{L}| < \infty$, we have $\bigcup_{\ell \in \mathcal{L}} \mathcal{I}_\ell = \{1, \dots, q\}$. Equivalently, $\mathcal{L}^{(j)} \neq \emptyset$ for all $j = 1, \dots, q$.

Under Assumptions 1–5, the consistency of the multi-lag NP-MOJO procedure is largely a consequence of Theorem 1. Assumption 4 (ii) requires that at any lag $\ell \in \mathcal{L}$ and a given change point θ_j , we have either $j \in \mathcal{I}_\ell$ with $d_\ell^{(j)}$ large enough (in the sense that $\sqrt{G/\log(n)} d_\ell^{(j)} \rightarrow \infty$), or $j \notin \mathcal{I}_\ell$ such that $d_\ell^{(j)} = 0$. Such a dyadic classification of the change points rules out the possibility that for some j , we have $d_\ell^{(j)} > 0$ but $d_\ell^{(j)} = O\{\sqrt{\log(n)/G}\}$, in which case θ_j may escape detection by NP-MOJO at lag ℓ . We therefore consider the following alternative:

Assumption 6. (i) $G^{-1} \log(n) \rightarrow 0$ as $n \rightarrow \infty$ while $\min_{0 \leq j \leq q} (\theta_{j+1} - \theta_j) \geq 4G$.
(ii) $\sqrt{G/\log(n)} \min_{1 \leq j \leq q} \max_{\ell \in \mathcal{L}^{(j)}} d_\ell^{(j)} \rightarrow \infty$.

Compared to Assumption 4, Assumption 6 requires that the change points are further apart from one another relative to G by the multiplicative factor of two. At the same time, the latter only requires that for each $j = 1, \dots, q$, there exists *at least one* lag $\ell \in \mathcal{L}$ at which $d_\ell^{(j)}$ is large enough to guarantee the detection of θ_j by NP-MOJO with large probability. Theorem 3 establishes the consistency of multi-lag NP-MOJO under either Assumption 4 or 6.

THEOREM 3. Suppose that Assumptions 1, 2, 3 and 5 hold and at each $\ell \in \mathcal{L}$, we set $\zeta_\ell(n, G) = c_{\zeta, \ell} \sqrt{\log(n)/G}$ with some constants $c_{\zeta, \ell} > 0$. Let $\hat{\Theta} = \{\hat{\theta}_j, 1 \leq j \leq \hat{q} : \hat{\theta}_1 < \dots < \hat{\theta}_{\hat{q}}\}$ denote the set of estimators returned by multi-lag NP-MOJO with tuning parameter c .

(i) If Assumption 4 holds for all $\ell \in \mathcal{L}$ and $c = 2\eta$ with $\eta \in (0, 1/2]$, then with c_0 as in Theorem 1, depending only on $C_F, C_X, \gamma_1, C_\beta, \gamma_2$ and p ,

$$\Pr\left(\hat{q} = q, \max_{1 \leq j \leq q} \max_{\ell \in \mathcal{L}^{(j)}} d_\ell^{(j)} |\hat{\theta}_j - \theta_j| \leq c_0 \sqrt{G \log(n)}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

(ii) If Assumption 6 holds and $c = 2$, then the conclusion of (i) holds.

Under Assumption 6 (ii), which is weaker than Assumption 4 (ii), we may encounter a situation where $\sqrt{G/\log(n)} d_\ell^{(j)} = O(1)$ while $d_\ell^{(j)} > 0$ at some lag $\ell \in \mathcal{L}$. Then, we cannot guarantee that such θ_j is detected by NP-MOJO at lag ℓ and, even so, we can only show that its estimator $\tilde{\theta} \in \tilde{\Theta}_\ell$ satisfies $|\tilde{\theta} - \theta_j| = O(G)$. This requires setting the tuning parameter c maximally for the clustering in Step 2 of multi-lag NP-MOJO, see Theorem 3 (ii). At the same time, there exists a lag well-suited for the localisation of each change point and Step 3 identifies an estimator detected at such lag, and the final estimator inherits the rate of estimation attained at the favourable lag.

3.4. Threshold selection via dependent wild bootstrap

Theorem 1 gives the choice of the threshold $\zeta_\ell(n, G) = c_\zeta \sqrt{\log(n)/G}$ which guarantees the consistency of NP-MOJO in multiple change point estimation. The choice of c_ζ influences the finite sample performance of NP-MOJO but it depends on many unknown quantities involved in specifying the degree of serial dependence in $\{X_t\}_{t=1}^n$ (see Assumptions 1 and 2), which makes the theoretical choice of little practical use. Resampling is popularly adopted for the calibration of change point detection methods including threshold selection. However, due to the presence of serial dependence, permutation-based approaches such as that adopted in Matteson & James (2014) or sample splitting adopted in Madrid Padilla et al. (2021) are inappropriate.

We propose to adopt the dependent wild bootstrap procedure proposed in Leucht & Neumann (2013), in order to approximate the quantiles of $\max_{G \leq k \leq n-G} T_\ell(G, k)$ in the absence of any change point, from which we select $\zeta_\ell(n, G)$. Let $\{W_t^{[r]}\}_{t=1}^{n-G}$ denote a bootstrap sequence generated as a Gaussian AR(1) process with $\text{var}(W_t^{[r]}) = 1$ and the AR coefficient $\exp(-1/b_n)$, where the sequence $\{b_n\}$ is chosen such that $b_n = o(n)$ and $\lim_{n \rightarrow \infty} b_n = \infty$. We construct bootstrap

10

E. T. MCGONIGLE AND H. CHO

replicates using $\{W_t^{[r]}\}_{t=1}^{n-G}$ as $T_\ell^{[r]} = \max_{G \leq k \leq n-G} T_\ell^{[r]}(G, k)$, where

$$T_\ell^{[r]}(G, k) = \frac{1}{(G - \ell)^2} \left\{ \sum_{s,t=k-G+1}^{k-\ell} \bar{W}_{s,k}^{[r]} \bar{W}_{t,k}^{[r]} h(Y_s, Y_t) + \sum_{s,t=k+1}^{k+G-\ell} \bar{W}_{s-G,k}^{[r]} \bar{W}_{t-G,k}^{[r]} h(Y_s, Y_t) \right. \\ \left. - 2 \sum_{s=k-G+1}^{k-\ell} \sum_{t=k+1}^{k+G-\ell} \bar{W}_{s,k}^{[r]} \bar{W}_{t-G,k}^{[r]} h(Y_s, Y_t) \right\},$$

with $\bar{W}_{t,k}^{[r]} = W_t^{[r]} - (G - \ell)^{-1} \sum_{u=k-G+1}^{k-\ell} W_u^{[r]}$. Independently generating $\{W_t^{[r]}\}_{t=1}^{n-G}$ for $r = 1, \dots, R$ (R denoting the number of bootstrap replications), we store $T_\ell^{[r]}$ and select the threshold as $\zeta_\ell(n, G) = q_{1-\alpha}(\{T_\ell^{[r]}\}_{r=1}^R)$, the $(1 - \alpha)$ -quantile of $\{T_\ell^{[r]}\}_{r=1}^R$ for the chosen level $\alpha \in (0, 1]$. Additionally, we can compute the importance score for each $\hat{\theta} \in \hat{\Theta}_\ell$ as

$$s(\hat{\theta}) = \frac{\left| \left\{ 1 \leq r \leq R : T_\ell(G, \hat{\theta}) \geq T_{\ell,r}^{[r]} \right\} \right|}{R}. \quad (7)$$

Taking a value between 0 and 1, the larger $s(\hat{\theta})$ is, the more likely that there exists a change point close to $\hat{\theta}$ empirically. The bootstrap procedure generalises to the multi-lag NP-MOJO straightforwardly. In practice, we observe that setting $\hat{\theta}_j = \arg \max_{\hat{\theta} \in \hat{\Theta}_j} s(\hat{\theta})$ (with some misuse of the notation, $s(\cdot)$ is computed at the relevant lag for each $\hat{\theta}$) works well in Step 3 of multi-lag NP-MOJO. This is attributed to the fact that this score inherently takes into account the varying scale of the detector statistics at multiple lags and ‘standardises’ the importance of each estimator. In all numerical experiments, our implementation of multi-lag NP-MOJO is based on this choice of $\hat{\theta}_j$. We provide the algorithmic descriptions of NP-MOJO and its multi-lag extension in Algorithms 1 and 2 in Appendix A.5.

4. IMPLEMENTATION OF NP-MOJO

In this section, we discuss the computational aspects of NP-MOJO and provide recommendations for the choice of tuning parameters based on extensive numerical results. Numerical studies analysing NP-MOJO’s sensitivity to these tuning parameters can be found in Appendix B.

Computational complexity: owing to the moving sum-based approach, the cost of sequentially computing $T_\ell(G, k)$ from $T_\ell(G, k - 1)$ is $O(G)$, giving the overall cost of computing $T_\ell(G, k)$, $G \leq k \leq n - G$, as $O(nG)$. Exact details of the sequential update are given in Appendix A.1. The bootstrap procedure described in Section 3.4 is performed once per lag for simultaneously detecting multiple change points, in contrast with E-Divisive (Matteson & James, 2014) that requires the permutation-based testing to be performed for detecting each change point. With R bootstrap replications, the total computational cost is $O(|\mathcal{L}|RnG)$ for multi-lag NP-MOJO using the set of lags $\mathcal{L}$ and bootstrapping, as opposed to $O(Rqn^2)$ for E-Divisive. Furthermore, the bootstrap procedure can be parallelised in a straightforward manner, which we include as an option in the implementation of the method.

We ran simulations to compare the computational speed of the competing nonparametric methods. We simulate realisations under the change in mean model (B.1), with increasing values of sample size n and the number of equispaced change points q ($(n, q) \in \{(100, 1), (500, 2), (1000, 3), (2000, 5), (5000, 10), (10000, 20)\}$). We use the same settings for each method as in the main simulation study, using the parallelised version of multi-lag NP-MOJO when $n \geq 2000$, and compute the average run time over 100 realisations. The results are displayed

in Figure 2. The fastest method by far is `cpt.np`, followed by KCPA and NP-MOJO. E-Divisive and NWBS are noticeably slower than the other methods. In particular, when $n = 10000$, the average running time of `cpt.np` is 0.17 seconds, KCPA is 46.26 seconds, NP-MOJO is 2.31 minutes, NWBS is 30.06 minutes, and E-Divisive is 70.37 minutes. Also, we observe that KCPA's running time is increasing at a faster rate than NP-MOJO's, and may exceed the running time of NP-MOJO for larger values of n .

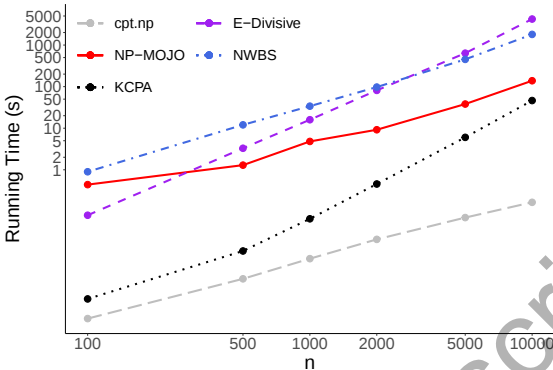


Fig. 2: Running time comparisons between the competing nonparametric methods. Both axes are in the log scale.

Kernel function: as with any kernel-based approach, NP-MOJO's performance will vary with the choice of kernel, and a kernel that works well for one type of change point may not be the best for another type of change point. Based on empirical performance and versatility to a wide range of change point scenarios (see Appendix B.3.1), we recommend the use of the kernel function h_2 in Lemma 2 (ii). The parameter δ is set using the 'median trick', a common heuristic used in kernel-based methods (Li et al., 2019). Specifically, we set δ to be a half the median of all $\|Y_s - Y_t\|^2$ involved in the calculation of $T_\ell(G, k)$. For p -variate i.i.d. Gaussian data with common variance σ^2 , this corresponds to $\delta \approx \sigma p$ as the dimension p increases (Ramdas et al., 2015). As with the kernel h_2 , the median trick can also be used when setting β if the kernel h_1 is used.

Bandwidth: due to the nonparametric nature of NP-MOJO, it is advised to use a larger bandwidth than that shown to work well for the moving sum procedure for univariate mean change detection (Eichinger & Kirch, 2018). In our simulation studies and data applications, we set $G = \lfloor n/6 \rfloor$. It is often found that using multiple bandwidths and merging the results improves the adaptivity of moving window-based procedures, such as the 'bottom-up' merging proposed by Messer et al. (2014) or the localised pruning of (Cho & Kirch, 2022). We empirically explore the multiscale extension of the multi-lag NP-MOJO with bottom-up merging, see Appendix A.3 for details of its implementation and Appendix B.5 for a proof of concept numerical study involving multiscale change point scenarios. We leave a theoretical investigation into the multiscale extension of NP-MOJO for future research.

Parameters for change point estimation: we set $\eta = 0.4$ in (5) following the recommendation in Meier et al. (2021). For multi-lag NP-MOJO, we set $c = 1$ for clustering the estimators from multiple lags, a choice that lies between those recommended in Theorem 3 (i) and (ii), since we do not know whether Assumptions 4 or 6 hold in practice. Appendices B.3.3 and B.3.4 demonstrate that within a reasonable range, NP-MOJO is insensitive to the choices of η and c . To further guard against spurious estimators, we only accept those $\hat{\theta}$ that lie in intervals of length greater than $\lfloor 0.02G \rfloor$ where the corresponding $T_\ell(G, k)$ exceeds $\zeta_\ell(n, G)$.

Parameters for the bootstrap procedure: the choice of b_n sets the level of dependence in the multiplier bootstrap sequences. [Leucht & Neumann \(2013\)](#) show that a necessary condition is that $\lim_{n \rightarrow \infty} (b_n^{-1} + b_n n^{-1}) = 0$, giving a large freedom for choice of b_n . We recommend $b_n = 1.5n^{1/3}$, which works well in practice. Appendix B.3.2 demonstrates that within a reasonable range, NP-MOJO is insensitive to the choice of b_n . As for α , its choice amounts to setting the level of significance in statistical testing. This provides a more systematic alternative to the problem of model selection in multiple change point detection compared to others, such as those requiring the selection of a threshold that is known up to a rate (or a range of rates, see e.g. [Madrid Padilla et al., 2023](#)), or constants involved in the penalty of a penalised cost function ([Arlot et al., 2019](#)). In all numerical experiments, we use $\alpha = 0.1$ with $R = 499$ bootstrap replications.

Set of lags $\mathcal{L}$: the flexibility of NP-MOJO in its ability to detect changes in dependence, comes at the price of having to select the set of lags $\mathcal{L}$. The choice of $\mathcal{L}$ depends on the practitioner's interest and domain knowledge, a problem commonly faced by general-purpose change point detection methods, such as the choice of the quantile level in [Vanegas et al. \(2022\)](#), the parameter of interest in [Zhao et al. \(2022\)](#) and the estimating equation in [Kirch & Reckruehm \(2024\)](#). For example, for monthly data, using $\mathcal{L} = \{0, 3, 12\}$ allows for detecting changes in the quarterly and yearly seasonality. Even when the interest lies in detecting changes in the marginal distribution only, it helps to jointly consider multiple lags, since any marginal distributional change is likely to result in changes in the joint distribution of $(X_t, X_{t+\ell})$. As we consider time series that exhibit short range dependence, we would expect that NP-MOJO will not have detection power at large lags. In simulations, we use $\mathcal{L} = \{0, 1, 2\}$ which works well not only for detecting changes in the mean and the second-order structure, but also for detecting changes in (non-linear) serial dependence and higher-order characteristics. For a practical approach to lag selection, see Appendix A.4 in the supplementary material, where we propose a semi-automatic method for choosing the set of lags $\mathcal{L}$ given some initial set $\tilde{\mathcal{L}}$.

5. SIMULATION STUDY

We conduct extensive simulation studies with varying change point scenarios (30 scenarios where $q \geq 1$, 7 with $q = 0$), sample sizes ($n \in \{500, 1000, 2000, 10000\}$) and dimensions $p \in \{1, 2, 5, 10\}$, and consider both evenly-spaced and multiscale change point settings. We provide complete descriptions of the simulation studies in Appendix B where, for comparison, we consider not only nonparametric but also parametric data segmentation procedures well-suited to detect the types of changes in consideration, which include changes in the mean, second-order and higher-order moments and non-linear serial dependence. Due to space constraint, here we focus on a selection of the results in the evenly-spaced setting with $n = 1000$ comparing both single-lag and multi-lag NP-MOJO (denoted by NP-MOJO- ℓ and NP-MOJO- $\mathcal{L}$ respectively), with the nonparametric competitors: E-Divisive ([Matteson & James, 2014](#)), NWBS ([Madrid Padilla et al., 2021](#)), KCPA ([Celisse et al., 2018](#); [Arlot et al., 2019](#)) and cpt.np ([Haynes et al., 2017](#)). E-Divisive and KCPA are applicable to multivariate data segmentation whilst NWBS and cpt.np are not. The scenarios are:

- (B5) $X_t = \sum_{j=0}^3 \Sigma_j^{1/2} \mathbb{I}\{\theta_j + 1 \leq t \leq \theta_{j+1}\} \cdot \varepsilon_t$, where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})^\top$ with $\varepsilon_{it} \sim_{\text{i.i.d.}} t_5$, $(\theta_1, \theta_2, \theta_3) = (250, 500, 750)$, $\Sigma_0 = \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\Sigma_1 = \Sigma_3 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$.
- (C1) $X_t = X_t^{(j)} = a_j X_{t-1}^{(j)} + \varepsilon_t$ for $\theta_j + 1 \leq t \leq \theta_{j+1}$, where $q = 2$, $(\theta_1, \theta_2) = (333, 667)$ and $(a_0, a_1, a_2) = (-0.8, 0.8, -0.8)$.
- (C3) $X_t = X_t^{(j)} = \sigma_t^{(j)} \varepsilon_t$ with $(\sigma_t^{(j)})^2 = \omega_j + \alpha_j (X_{t-1}^{(j)})^2 + \beta_j (\sigma_{t-1}^{(j)})^2$ for $\theta_j + 1 \leq t \leq \theta_{j+1}$, where $q = 1$, $\theta_1 = 500$, $(\omega_0, \alpha_0, \beta_0) = (0.01, 0.7, 0.2)$ and $(\omega_1, \alpha_1, \beta_1) = (0.01, 0.2, 0.7)$.

(D3) $X_t = 0.4X_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim_{\text{i.i.d.}} \mathcal{N}(0, 0.5^2)$ for $t \leq \theta_1$ and $t \geq \theta_2 + 1$, and $\varepsilon_t \sim_{\text{i.i.d.}}$ Exponential(0.5) $- 0.5$ for $\theta_1 + 1 \leq t \leq \theta_2$, with $q = 2$ and $(\theta_1, \theta_2) = (333, 667)$.

Additional simulations for differing sample sizes can be found in Appendix B.4, and simulations with uneven spacing between neighbouring segments examining the performance of the multiscale version of multi-lag NP-MOJO are given in Appendix B.5. The above scenarios consider changes in the covariance of bivariate, non-Gaussian random vectors in (B5), changes in the autocorrelation (while the variance stays unchanged) in (C1), a change in the parameters of an ARCH(1, 1) process in (C3), and changes in higher moments of serially dependent observations in (D3). Table 1 reports the distribution of the estimated number of change points and the average covering metric (CM) and V-measure (VM) over 1000 realisations. Taking values between [0, 1], CM and VM close to 1 indicates better accuracy in change point location estimation, see Appendix B.2 for their definitions and complete discussions of change point scenarios.

In the case of (C1), we have $q_\ell = 0$, $\ell \neq 1$, while $q_1 = 2$, and thus we report $\hat{q}_\ell - q_\ell$ for the respective single-lag NP-MOJO- ℓ . Across all scenarios, NP-MOJO- $\mathcal{L}$ shows good detection and estimation accuracy and demonstrates the efficacy of considering multiple lags, see (C3) and (D3) in particular. As the competitors are calibrated for the independent setting, they tend to either over- or under-detect the number of change points in the presence of serial dependence in (C1), (C3) and (D3). In Appendix B.2, we compare NP-MOJO against change point methods proposed for time series data where it performs comparably to methods specifically calibrated for the change point scenarios considered.

Table 1: Distribution of the estimated number of change points and the average CM and VM over 1000 realisations. The modal value of $\hat{q} - q$ in each row is given in bold. Also, the best performance for each metric is underlined for each scenario.

Model	Method	$\hat{q} - q / \hat{q}_\ell - q_\ell$					CM	VM
		≤ -2	-1	0	1	≥ 2		
(B5)	NP-MOJO-0	0.000	0.001	0.997	0.002	0.000	<u>0.974</u>	<u>0.959</u>
	NP-MOJO-1	0.005	0.121	0.867	0.007	0.000	0.931	0.927
	NP-MOJO-2	0.006	0.103	0.884	0.007	0.000	0.935	0.929
	NP-MOJO- $\mathcal{L}$	0.000	0.001	0.999	0.000	0.000	<u>0.973</u>	<u>0.958</u>
	E-Divisive	0.670	0.189	0.101	0.032	0.008	0.431	0.335
	KCPA	0.322	0.000	0.662	0.015	0.001	<u>0.775</u>	<u>0.725</u>
(C1)	NP-MOJO-0	–	–	0.851	0.140	0.009	–	–
	NP-MOJO-1	0.000	0.002	0.956	0.042	0.000	0.978	0.961
	NP-MOJO-2	–	–	0.836	0.149	0.015	–	–
	NP-MOJO- $\mathcal{L}$	0.000	0.002	0.986	0.012	0.000	<u>0.980</u>	<u>0.963</u>
	E-Divisive	0.001	0.001	0.012	0.035	0.951	0.685	0.686
	KCPA	0.792	0.002	0.065	0.025	0.116	0.399	0.132
(C3)	NWBS	0.013	0.001	0.007	0.015	0.964	0.398	0.558
	cpt.np	0.000	0.000	0.002	0.003	0.995	<u>0.593</u>	<u>0.647</u>
	NP-MOJO-0	–	0.409	0.533	0.056	0.002	0.744	0.484
	NP-MOJO-1	–	0.236	0.682	0.081	0.001	0.819	0.633
	NP-MOJO-2	–	0.299	0.626	0.073	0.002	0.787	0.571
	NP-MOJO- $\mathcal{L}$	–	0.210	0.727	0.062	0.001	<u>0.823</u>	<u>0.645</u>
	E-Divisive	–	0.032	0.327	0.211	0.430	0.742	0.602
	KCPA	–	0.418	0.262	0.171	0.149	0.667	0.370
	NWBS	–	0.895	0.048	0.020	0.037	0.525	0.069

	cpt.np	–	0.000	0.013	0.047	0.940	0.634	0.554
(D3)	NP-MOJO-0	0.003	0.139	0.809	0.049	0.000	0.899	0.872
	NP-MOJO-1	0.006	0.155	0.792	0.047	0.000	0.892	0.864
	NP-MOJO-2	0.021	0.248	0.685	0.045	0.001	0.848	0.819
	NP-MOJO- $\mathcal{L}$	0.002	0.082	0.914	0.002	0.000	0.917	0.884
	E-Divisive	0.005	0.002	0.072	0.118	0.803	0.681	0.707
	KCPA	0.441	0.012	0.481	0.052	0.014	0.667	0.500
	NWBS	0.047	0.015	0.139	0.124	0.675	0.680	0.676
	cpt.np	0.000	0.000	0.045	0.055	0.900	0.726	0.756

6. DATA APPLICATIONS

6.1. California seismology measurements data set

We analyse a data set from the High Resolution Seismic Network, operated by the Berkeley Seismological Laboratory. Ground motion sensor measurements were recorded in three mutually perpendicular directions at 13 stations near Parkfield, California, USA for 740 seconds from 2am on December 23rd 2004. The data has previously been analysed in [Xie et al. \(2019\)](#) and [Chen et al. \(2022\)](#). [Chen et al. \(2022\)](#) pre-process the data by removing a linear trend and down-sampling, and the processed data is available in the ocd R package ([Chen et al., 2020](#)). According to the Northern California Earthquake Catalog, an earthquake of magnitude 1.47 Md hit near Atascadero, California (50 km away from Parkfield) at 02:09:54.01.

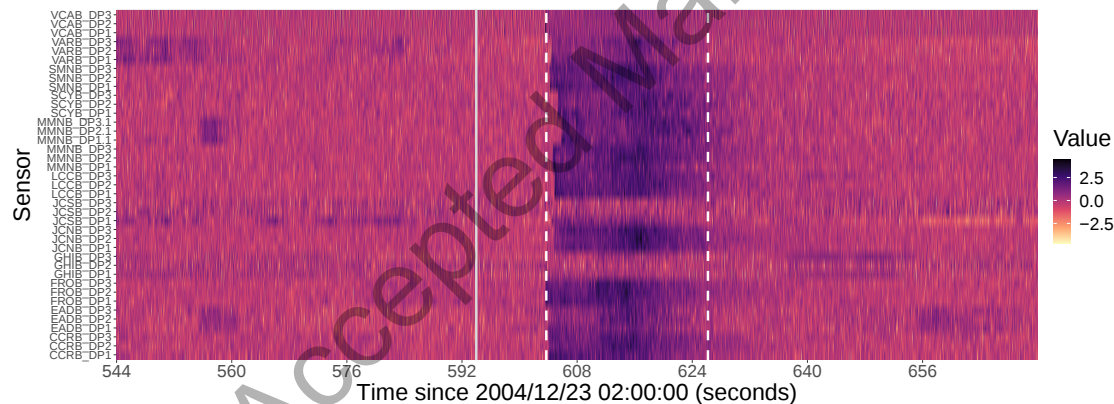


Fig. 3: Heat map of standardised sensor data. Change points detected by multi-lag NP-MOJO are shown in vertical dashed lines, and the time of the earthquake is given by solid vertical line.

We analyse time series of dimension $p = 39$ and length $n = 2000$ by taking a portion of the data set between 544 and 672 seconds after 2am, which covers the time at which the earthquake occurred (594 seconds after). We apply the multi-lag NP-MOJO with tuning parameters selected as in Section 4, using $G = 333$ and set of lags $\mathcal{L} = \{0, \dots, 4\}$. We detect two changes at all lags; the first occurs at between 603.712 and 603.968 seconds after 2am and may be attributed to the earthquake. As noted in [Chen et al. \(2022\)](#), P waves, which are the primary preliminary wave and arrive first after an earthquake, travel at up to 6km/s in the Earth's crust. This is consistent with the delay of approximately 9 seconds between the occurrence of the earthquake and the first change point detected by multi-lag NP-MOJO. We also note that performing online change point analysis, [Xie et al. \(2019\)](#) and [Chen et al. \(2022\)](#) report a change at 603.584 and 603.84 seconds after the

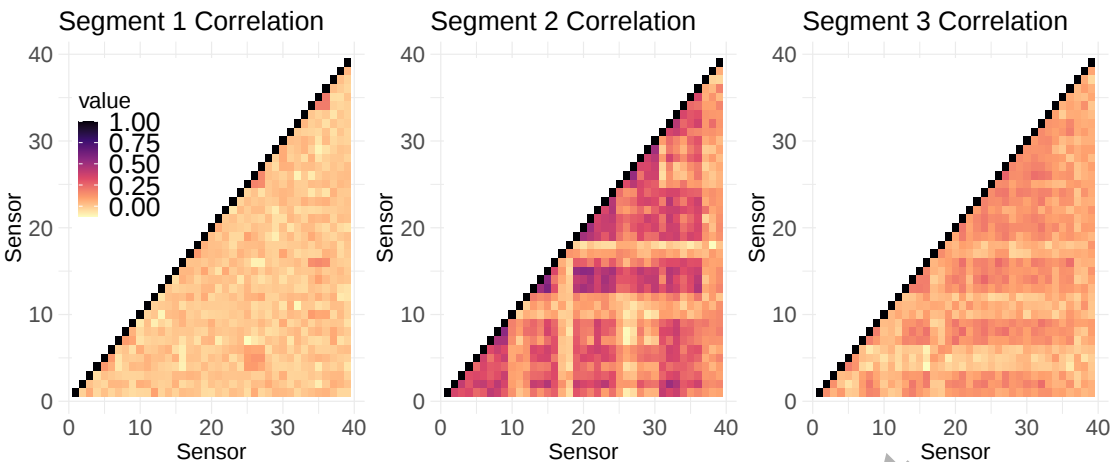


Fig. 4: Sample correlations from the three segments defined by the change point estimators.

earthquake, respectively. The second change is detected at between 626.176 and 626.496 seconds after 2am. It may correspond to the ending of the effect of the earthquake, as sensors return to ‘baseline’ behaviour. Figure 3 plots the heat map of the data with each series standardised for ease of visualisation, along with the onset of the earthquake and the two change points detected by the multi-lag NP-MOJO. It suggests, amongst other possible distributional changes, the time series undergoes mean shifts as found in Chen et al. (2022). We also examine the sample correlations computed on each of the three segments, see Figure 4 where the data exhibit a greater degree of correlation in segment 2 compared to the other two segments. Recalling that each station is equipped with three sensors, we notice that pairwise correlations from the sensors located at the same stations undergo greater changes in correlations. A similar observation is made about the sensors located at nearby stations.

6.2. US recession data

We analyse the US recession indicator data set. Recorded quarterly between 1855 and 2021 ($n = 667$), X_t is recorded as a 1 if any month in the quarter is in a recession (as identified by the Business Cycle Dating Committee of the National Bureau of Economic Research), and 0 otherwise. The data has previously been examined for change points under piecewise stationary autoregressive models for integer-valued time series in Hudecová (2013) and Diop & Kengne (2021). We apply the multi-lag NP-MOJO with $G = 111$ and $\mathcal{L} = \{0, \dots, 4\}$. All tuning parameters are set as recommended in Section 4 with one exception, δ for the kernel h_2 . We select $\delta = 1$ for lag 0 and 2 otherwise, since pairwise distances for binary data are either 0 or 1 when $\ell = 0$ such that the median heuristic would not work as desired.

At all lags, we detect a single change point located between 1933:Q1 and 1938:Q2. Multi-lag NP-MOJO estimates the change point at 1933:Q1, which is comparable to the previous analyses: Hudecová (2013) report a change at 1933:Q1 and Diop & Kengne (2021) at 1932:Q4. The change coincides with the ending of the Great Depression and beginning of World War II. The left panel of Figure 5 plots the detected change along with the sample average of X_t over the two segments (superimposed on $\{X_t\}_{t=1}^n$), showing that the frequency of recession is substantially lower after the change. The right panel plots the detector statistics $T_\ell(G, k)$ at lags $\ell \in \mathcal{L}$, divided by the respective threshold $\zeta_\ell(n, G)$ obtained from the bootstrap procedure. The thus-standardised $T_4(G, k)$, shown in solid line, displays the change point with the most clarity,

16

E. T. MCGONIGLE AND H. CHO

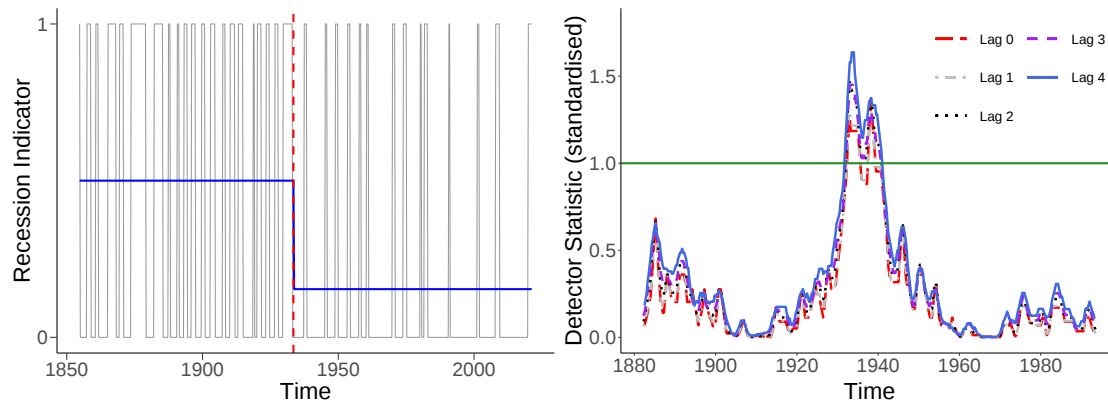


Fig. 5: Left: quarterly US recession indicator series. A change point detected by multi-lag NP-MOJO is shown in vertical dashed lines and the sample means over the two segments in solid line. Right: $T_\ell(G, k)$, $G \leq k \leq n - G$ for lags $\ell \in \mathcal{L}$, after standardisation by respective thresholds.

attaining the largest value over the widest interval above the threshold (standardised to be one). At lag 4, the detector statistic has the interpretation of measuring any discrepancy in the joint distribution of the recession indicator series and its yearly lagged values.

Supplementary material

The supplementary appendix contains additional discussion on the implementation of NP-MOJO and multi-lag NP-MOJO, the complete simulation results and the proofs of all theoretical results.

REFERENCES

- ARLOT, S., CELISSE, A. & HARCHAOU, Z. (2019). A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research* **20**, 1–56.
- AUE, A., HÖRMANN, S., HORVÁTH, L. & REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37**, 4046–4087.
- BAI, J. & PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47–78.
- BONIECE, B. C., HORVÁTH, L. & JACOBS, P. M. (2023). Change point detection in high dimensional data with u-statistics. *TEST*, 1–53.
- CARLSTEIN, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics* **16**, 188–197.
- CARR, J. R., BELL, H., KILLICK, R. & HOLT, T. (2017). Exceptional retreat of Novaya Zemlya’s marine-terminating outlet glaciers between 2000 and 2013. *The Cryosphere* **11**, 2149–2174.
- CELISSE, A., MAROT, G., PIERRE-JEAN, M. & RIGAILL, G. (2018). New efficient algorithms for multiple change-point detection with reproducing kernels. *Computational Statistics & Data Analysis* **128**, 200–220.
- CHAKRABORTY, S. & ZHANG, X. (2021). High-dimensional change-point detection using generalized homogeneity metrics. *arXiv preprint arXiv:2105.08976*.
- CHEN, H. & FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *Journal of the American statistical association* **112**, 397–409.
- CHEN, H. & ZHANG, N. (2015). Graph-based change-point detection. *The Annals of Statistics* **43**, 139 – 176.
- CHEN, Y., WANG, T. & SAMWORTH, R. J. (2020). *ocd: High-dimensional, multiscale online changepoint detection*. R package version 1.1.
- CHEN, Y., WANG, T. & SAMWORTH, R. J. (2022). High-dimensional, multiscale online changepoint detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **84**, 234–266.
- CHO, H. & FRYZLEWICZ, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Stat. Sinica* **22**, 207–229.
- CHO, H. & FRYZLEWICZ, P. (2023). Multiple change point detection under serial dependence: Wild contrast maximisation and gappy schwarz algorithm. *Journal of Time Series Analysis* **45**, 479–494.

CHO, H. & KIRCH, C. (2022). Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. *Annals of the Institute of Statistical Mathematics* **74**, 653–684.

CHO, H. & KIRCH, C. (2023+). Data segmentation algorithms: Univariate mean change and beyond. *Econometrics and Statistics (to appear)*.

CHU, C.-S. J., HORNIK, K. & KAUN, C.-M. (1995). MOSUM tests for parameter constancy. *Biometrika* **82**, 603–617.

CHU, L. & CHEN, H. (2019). Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data. *The Annals of Statistics* **47**, 382 – 414.

DAVIS, R. A., MATSUI, M., MIKOSCH, T. & WAN, P. (2018). Applications of distance correlation to time series. *Bernoulli* **24**, 3087–3116.

DETTE, H., ECKLE, T. & VETTER, M. (2020). Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics* **47**, 1243–1274.

DIOP, M. L. & KENGNE, W. (2021). Piecewise autoregression for general integer-valued time series. *Journal of Statistical Planning and Inference* **211**, 271–286.

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

FAN, Y., DE MICHEAUX, P. L., PENEV, S. & SALOPEK, D. (2017). Multivariate nonparametric test of independence. *Journal of Multivariate Analysis* **153**, 189–210.

FOKIANOS, K. & PITSILLOU, M. (2017). Consistent testing for pairwise dependence in time series. *Technometrics* **59**, 262–270.

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

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

FRYZLEWICZ, P. & SUBBA RAO, S. (2014). Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 903–924.

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. & SMOLA, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research* **13**, 723–773.

HARCHAOUI, Z., VALLET, F., LUNG-YUT-FONG, A. & CAPPÉ, O. (2009). A regularized kernel-based approach to unsupervised audio segmentation. In *ICASSP*.

HAYNES, K., FEARNHEAD, P. & ECKLEY, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing* **27**, 1293–1305.

HUDECOVÁ, S. (2013). Structural changes in autoregressive models for binary time series. *Journal of Statistical Planning and Inference* **143**, 1744–1752.

HUSKOVA, M. & SLABY, A. (2001). Permutation tests for multiple changes. *Kybernetika* **37**, 605–622.

JEWELL, S. W., HOCKING, T. D., FEARNHEAD, P. & WITTEN, D. M. (2020). Fast nonconvex deconvolution of calcium imaging data. *Biostatistics* **21**, 709–726.

KILLICK, R., FEARNHEAD, P. & ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107**, 1590–1598.

KIRCH, C. & RECKRUEHM, K. (2024). Data segmentation for time series based on a general moving sum approach. *Annals of the Institute of Statistical Mathematics*, 1–29.

KORKAS, K. K. & FRYZLEWICZ, P. (2017). Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, 287–311.

LAVIELLE, M. & TEYSSIERE, G. (2007). Adaptive detection of multiple change-points in asset price volatility. In *Long Memory in Economics*. Springer, pp. 129–156.

LEUCHT, A. & NEUMANN, M. H. (2013). Dependent wild bootstrap for degenerate U- and V-statistics. *Journal of Multivariate Analysis* **117**, 257–280.

LI, S., XIE, Y., DAI, H. & SONG, L. (2019). Scan B-statistic for kernel change-point detection. *Sequential Analysis* **38**, 503–544.

MADRID PADILLA, C. M., XU, H., WANG, D., MADRID PADILLA, O. H. & YU, Y. (2023). Change point detection and inference in multivariate non-parametric models under mixing conditions. In *Advances in Neural Information Processing Systems*, vol. 36.

MADRID PADILLA, O. H., YU, Y., WANG, D. & RINALDO, A. (2021). Optimal nonparametric change point analysis. *Electronic Journal of Statistics* **15**, 1154–1201.

MADRID PADILLA, O. H., YU, Y., WANG, D. & RINALDO, A. (2022). Optimal nonparametric multivariate change point detection and localization. *IEEE Transactions on Information Theory* **68**, 1922–1944.

MATTESON, D. S. & JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association* **109**, 334–345.

MCGONIGLE, E. T. & CHO, H. (2023). *CptNonPar: Nonparametric Change Point Detection for Multivariate Time Series*. R package version 0.1.2.

MEIER, A., KIRCH, C. & CHO, H. (2021). mosum: A package for moving sums in change-point analysis. *Journal of Statistical Software* **97**, 1–42.

- MESSER, M., KIRCHNER, M., SCHIEMANN, J., ROEPER, J., NEININGER, R. & SCHNEIDER, G. (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *Ann. Appl. Stat.* **8**, 2027–2067.
- MOKKADEM, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications* **29**, 309–315.
- PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115.
- PREUß, P., PUCHSTEIN, R. & DETTE, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association* **110**, 654–668.
- RAMDAS, A., REDDI, S. J., PÓCZOS, B., SINGH, A. & WASSERMAN, L. (2015). On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. *Proceedings of the AAAI Conference on Artificial Intelligence* **29**.
- SAFIKHANI, A. & SHOJAIE, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *Journal of the American Statistical Association* **117**, 251–264.
- SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. & FUKUMIZU, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, 2263–2291.
- SZÉKELY, G. J., RIZZO, M. L. & BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769–2794.
- TECUAPETLA-GÓMEZ, I. & MUNK, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scandinavian Journal of Statistics* **44**, 346–368.
- TRUONG, C., OUDRE, L. & VAYATIS, N. (2020). Selective review of offline change point detection methods. *Signal Processing* **167**, 107299.
- VANEGAS, L. J., BEHR, M. & MUNK, A. (2022). Multiscale quantile segmentation. *Journal of the American Statistical Association* **117**, 1384–1397.
- WANG, D., YU, Y. & RINALDO, A. (2021). Optimal covariance change point localization in high dimensions. *Bernoulli* **27**, 554–575.
- WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102**, 14150–14154.
- XIE, L., XIE, Y. & MOUSTAKIDES, G. V. (2019). Asynchronous multi-sensor change-point detection for seismic tremors. In *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE.
- XU, H., WANG, D., ZHAO, Z. & YU, Y. (2024). Change point inference in high-dimensional regression models under temporal dependence. *Annals of Statistics (to appear)*.
- YOUSUF, K. & FENG, Y. (2022). Targeting predictors via partial distance correlation with applications to financial forecasting. *Journal of Business & Economic Statistics* **40**, 1007–1019.
- ZHAO, Z., JIANG, F. & SHAO, X. (2022). Segmenting time series via self-normalisation. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84**, 1699–1725.
- ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis* **33**, 438–457.
- ZOU, C., YIN, G., FENG, L. & WANG, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *The Annals of Statistics* **42**, 970–1002.

[Received on 2 January 2024. Editorial decision on 3 February 2025]