Information Sciences 447 (2018) 117–139

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Capturing correlation changes by applying kernel change point detection on the running correlations☆

Jedelyn Cabrieto[a,*], Francis Tuerlinckx[a], Peter Kuppens[a], Frank H. Wilhelm[b], Michael Liedlgruber[b], Eva Ceulemans[a]

[a] *Quantitative Psychology and Individual Differences Research Group, KU Leuven – University of Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium*
[b] *Division of Clinical Psychology, Psychotherapy and Health Psychology, Department of Psychology, University of Salzburg, Salzburg, Austria*

ARTICLE INFO

Article history:
Received 17 November 2016
Revised 23 January 2018
Accepted 4 March 2018
Available online 8 March 2018

Keywords:
Non-parametric change point detection
Kernel change point detection
Correlation change
Multivariate time series
Running correlation
Psychophysiology

ABSTRACT

Change point detection methods signal the occurrence of abrupt changes in a time series. Non-parametric approaches, such as the Gaussian kernel based change point (KCP) detection (Arlot et al., 2012), are especially attractive because they impose less assumptions on the data. Yet, a drawback of these methods is that most of them are sensitive to changes in the mean, the variance, etc., making them less sensitive to specific kinds of changes. We show that KCP can be adapted to detect a particular type of change only. We focus here on correlation change, which has been put forward by different theories but proved hard to trace in multivariate time series. We propose KCP-corr, which boils down to applying KCP on the running correlations. To confirm that KCP-corr is more sensitive than merely applying KCP on the raw data (KCP-raw), a simulation study was conducted in which the number of (noise) variables and the size of the correlation change were varied. KCP-corr emerged as the better method especially in the more difficult but realistic settings where the correlation change is minimal and/or noise variables are present. KCP-corr also outperforms Cusum, a non-parametric method that specifically targets the detection of correlation changes.

1. Introduction

Change point detection methods have a long-established function in many scientific fields where industrial, geophysical and mechanical processes are monitored across time [5,7,28,30,33]. Nowadays, change point detection is applied in diverse areas such as genetics and medicine [12], psychopathology [32], financial market analysis [13,40] and computer networks intrusion prevention [36]. The goal is straightforward: detect the exact timing of abrupt changes in the distribution of the variables.

When a single variable is monitored, researchers are interested in abrupt changes in univariate features such as the mean, variance or autocorrelation. An extensive range of univariate change point detection techniques is available for this

☆ The research leading to the results reported in this paper was sponsored in part by a research grant from the Fund for Scientific Research-Flanders (FWO, Project No. G.0582.14 awarded to Eva Ceulemans, Peter Kuppens and Francis Tuerlinckx), by the Belgian Federal Science Policy within the framework of the Interuniversity Attraction Poles program (IAP/P7/06), and by the Research Council of KU Leuven (GOA/15/003).

* Corresponding author.
E-mail address: Jed.Cabrieto@kuleuven.be (J. Cabrieto).

https://doi.org/10.1016/j.ins.2018.03.010
0020-0255/

purpose (see Killick and Eckley [17] for a list of references and softwares). However, in the multivariate setting where multiple variables are tracked, aside from changes in univariate features, crucial events are often marked by abrupt correlational changes. For instance, in economics, increased dependency between supposedly diversified financial assets is typically linked to a financial crisis [13,40]. In EEG signal processing, excessive synchronization of brain electrical signals marks the onset of an epileptic seizure [16]. In developmental psychology, a change in the covariance structure of a set of tasks characterizes mastery of a competence by a child [1,18,37]. In emotion psychology, an emotion is considered to emerge from the synchronization of physiological, behavioral and experiential reactions in response to an emotion eliciting event in order for an organism to cope with threats and seize opportunities [24,25].

To capture such correlation changes in multivariate systems, non-parametric change point detection methods are flexible tools that do not impose strong assumptions on the data and thus are widely applicable in various fields. Most of these methods are general purpose methods which can signal change points caused by any type of distributional change in the data – DeCon [8], E-divisive [23], Multirank [22] and Gaussian Kernel[1] based Change Point detection (KCP) [2,3]. Yet, two major drawbacks of these general purpose methods were identified in a recent comparison [10]: first, they do not identify which specific distributional aspects changed. Second, compared to mean changes, correlation changes are harder to detect especially in the presence of noise variables (i.e., randomly fluctuating variables unaffected by the change). An alternative is to use Cusum,[2] a method that targets correlation changes by comparing the successive correlations (i.e., correlations based on the first to the ith time point) to the overall correlations. However, unlike KCP, which can locate multiple change points simultaneously, Cusum is a binary segmentation method that is constrained to locating one change point at a time. This approach is problematic, whenever the time series contains more than one correlation change point and particular correlation phases recur (i.e., because phases are pooled in the computation of the test statistic, the effect size is underestimated leading to a low detection power [9]).

We propose to address the correlation change detection problem by adapting KCP, such that it focuses on correlation changes. Specifically, we will implement KCP on a time series derived from the raw data which reflects fluctuations in correlations. To this end, we propose to inspect the running correlations which are the correlations computed in a time window that is slid across the time series. Through this adaptation, we focus the change point detection directly on the correlations. At the same time, we expect to have more detection power as correlation changes in the original data correspond to mean changes in the running correlations, which are easier to trace for KCP according to the results of Cabrieto et al. [10]. Thus, the main goal of this paper is to introduce KCP-corr as a multivariate non-parametric change point detection method dedicated to capturing correlation changes. Furthermore, we conduct an extensive simulation study to demonstrate that KCP-corr offers a substantial gain in performance over simply implementing KCP on the raw data. As a second goal, we aim to highlight that KCP-corr retains power in settings where Cusum breaks down.

The remainder of this paper is organized as follows: we present KCP-corr using a simulated toy example. Next, we evaluate its performance in a simulation study based on that of Matteson and James [23] which focuses on correlation change detection in settings where noise variables are present. We then apply KCP-corr to time series from an emotional reactivity study. Finally, we provide a discussion of our results and identify potential research directions.

2. Method

2.1. Simulated toy example

We begin by introducing the simulated toy example which will be used throughout this section. Let $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{300}\}$ denote a time series consisting of 300 time points, where $\boldsymbol{X}_i = \begin{bmatrix}\boldsymbol{X}_{i1}\\ \boldsymbol{X}_{i2}\\ \boldsymbol{X}_{i3}\end{bmatrix}$ is a vector of observed scores at time point, i, at which 3 variables are measured (Fig. 1, left panel). The variables follow a multivariate normal distribution with a variance of 1. Two change points are present: the first one is between the 100th and the 101st time points, and the second one is between the 200th and 201st time points, generating three phases with 100 subsequent time points each. In the first phase, $\boldsymbol{X}_{1:100} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{100}\}$, all variables have means equal to zero, $\boldsymbol{\mu}_{1:100} = \begin{bmatrix}0\\0\\0\end{bmatrix}$, and are uncorrelated, i.e., the covariance matrix $\boldsymbol{\Sigma}_{1:100} = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$. In the second phase, $\boldsymbol{X}_{101:200} = \{\boldsymbol{X}_{101}, \boldsymbol{X}_{102}, \ldots, \boldsymbol{X}_{200}\}$, a mean change

[1] The Gaussian kernel is a characteristic kernel implying that it can capture any change in the distribution. Note that non-characteristic kernels which are sensitive to specific changes exist and can be plugged in to the KCP algorithm, but to the best of our knowledge, there is no kernel proposed yet to specifically target correlation.

[2] Another technique focused on correlation changes is the matrix norm methods proposed by Barnett and Onnela [4]. Although the performance of this binary segmentation method was assessed in locating a single change point, the authors did not propose a way to account for multiple testing when the technique is applied sequentially to search for multiple change points. We deem that this is an important limitation since binary segmentation methods are prone to a high false detection rate [13]. For the Cusum method, for instance, aside from the correction for multiple testing, a refinement step is implemented to further prune the generated change points. Hence, we did not include the mentioned matrix norm methods in our comparison for this paper.

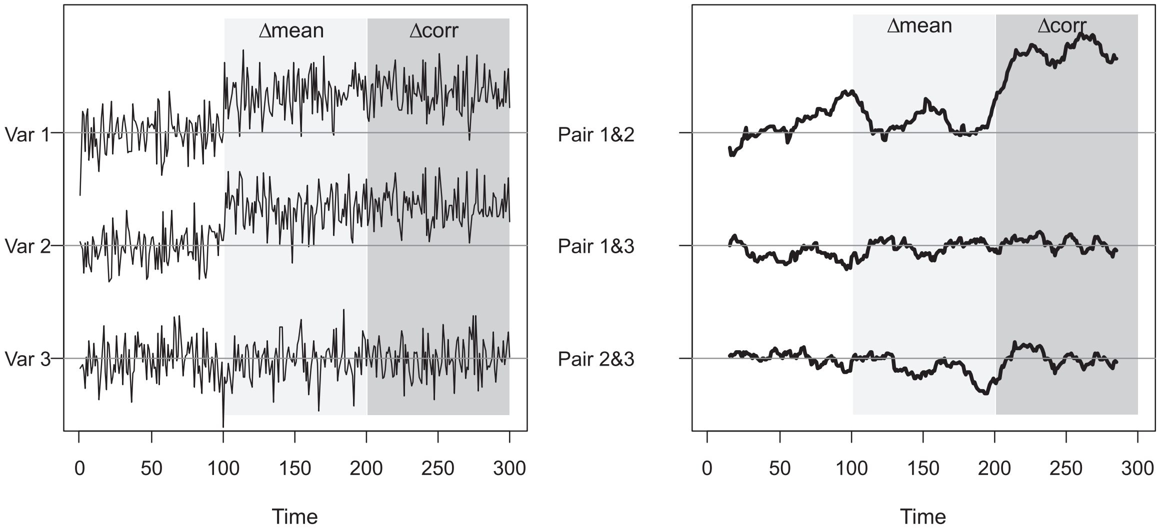

Fig. 1. The toy example: Raw data and running correlations. The left panel displays the simulated time series drawn from a multivariate normal distribution and composed of 3 variables measured at 300 time points. Two change points are introduced at time points $T = 101$ (mean change of 2 standard deviations for the first 2 variables) and $T = 201$ (correlation change of 0.9 for the first 2 variables), segmenting the series into 3 phases with the following distribution: $\boldsymbol{X}_{1:100} \sim MVN\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)$, $\boldsymbol{X}_{101:200} \sim MVN\left(\begin{bmatrix}2\\2\\0\end{bmatrix}, \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)$ and $\boldsymbol{X}_{201:300} \sim MVN\left(\begin{bmatrix}2\\2\\0\end{bmatrix}, \begin{bmatrix}1&0.9&0\\0.9&1&0\\0&1&1\end{bmatrix}\right)$. The underlying phases are indicated by varying background shading. In the right panel, Fisher's Z transformed running correlations for all possible pairs are plotted (window size = 30 time points). During the third phase (indicated by a dark gray background), a drastic increase in the running correlation values of the first and second variables is seen due to the correlation change introduced.

is introduced for the first two variables such that the new mean vector becomes $\boldsymbol{\mu}_{101:200} = \begin{bmatrix}2\\2\\0\end{bmatrix}$, while the covariance matrix, $\boldsymbol{\Sigma}_{101:200} = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$, remains unchanged. Finally, in the third phase, $\boldsymbol{X}_{201:300} = \{\boldsymbol{X}_{201}, \boldsymbol{X}_{202}, \ldots, \boldsymbol{X}_{300}\}$, the means, $\boldsymbol{\mu}_{201:300} = \begin{bmatrix}2\\2\\0\end{bmatrix}$, stay the same, but a correlation change is introduced for the first two variables such that the new covariance matrix becomes $\boldsymbol{\Sigma}_{201:300} = \begin{bmatrix}1&0.9&0\\0.9&1&0\\0&0&1\end{bmatrix}$.

Using this toy example, we will first introduce the original KCP approach, KCP-raw, as described in Arlot et al. [2,3]. Next, we will present KCP-corr, which boils down to applying KCP to the running correlations. We will show that for this particular case, KCP-raw will detect the first change point due to a mean change. However, as KCP-raw's performance in detecting correlation changes is severely affected by noise variables [10], it fails to detect the second change point. In contrast, KCP-corr exhibits a greater sensitivity and detects the second change point despite the noise variable.

We note here that our goal is to detect the change points between the generated phases. However, these change points are unobserved, as the first change point occurs for instance between time points, $T = 100$ and $T = 101$. As a way out, henceforward, we will denote the first observation from the new phase as the change point, implying that the change points in the toy example are at $T = 101$ and $T = 201$, respectively.

2.2. KCP-raw: kernel change point (KCP) detection on the raw data

KCP is a kernel based method which conducts change point detection by pooling the most similar subsequent observations into phases. This is carried out by using a kernel function to compute similarities of each observation with respect to all other observations in the time series. We employed the Gaussian kernel function, which is the most popular kernel in the literature [34]. Since this kernel is a characteristic kernel, this implementation of KCP is sensitive to any distributional change in the data.

The KCP method consists of two major steps: (1) estimation of change point locations, for different numbers of change points, and (2) choosing the number of change points. In the following subsections, we will describe these steps in detail.

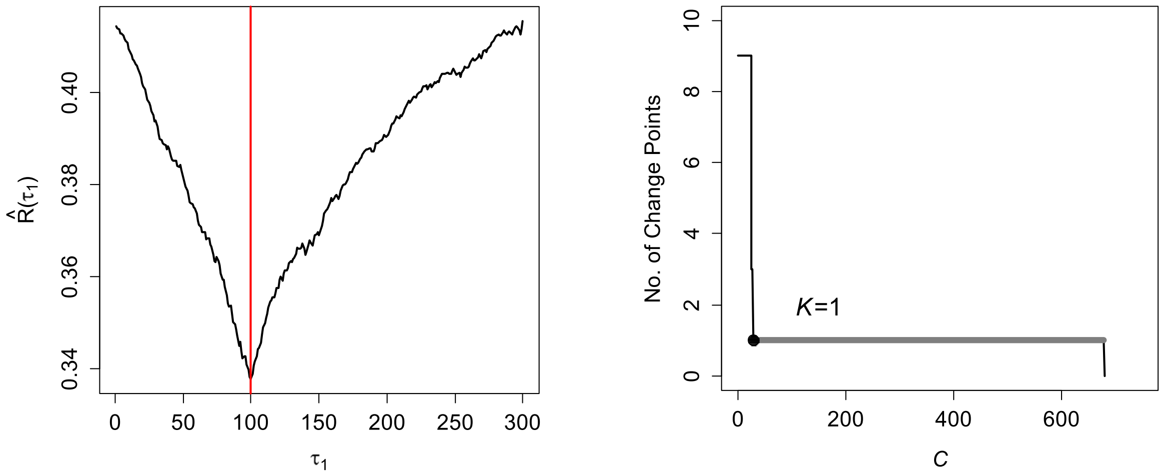

Fig. 2. KCP change point detection steps. In the left panel, $\hat{R}_{min,1}$ is minimized to find the change point location given that there is only one change point (K= 1) in the data. The minimum is generated at T= 100, indicating that the change point occurred at T= 101. In the right panel, the penalty coefficient, C, is tuned to derive the optimal number of change points, K= 1.

2.2.1. Estimation of change point locations

For a multivariate time series, $\boldsymbol{X} = \{\boldsymbol{X}_1, \ \boldsymbol{X}_2, \ \ldots, \boldsymbol{X}_n\}$, the similarities for all possible pairs, $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, are computed using the Gaussian kernel,

$$Gk\left(\boldsymbol{X}_i, \ \boldsymbol{X}_j\right) = \exp\left(\frac{-\parallel \boldsymbol{X}_i - \boldsymbol{X}_j \parallel^2}{2h^2}\right).$$

The value of this similarity approaches 1, when $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are similar and approaches 0 when they are dissimilar. The bandwidth, h, is a smoothing parameter reflecting how strict the user is when deciding if two observations are similar (the smaller, the stricter). The bandwidth parameter has to be fixed in advance. In line with Arlot et al. [3], we set h equal to the median Euclidean distance between all observations as a classical heuristic choice for Gaussian kernels [14].[3]

Given a specific number of change points, K, KCP estimates their location by minimizing a criterion based on intra-phase scatter values,

$$\hat{V}_{p, \ (\tau_1, \ \tau_2, \ \ldots, \ \tau_K)} \ = \ \left(\tau_p - \tau_{p-1}\right) - \frac{1}{\tau_p - \tau_{p-1}} \sum_{i=\tau_{p-1}+1}^{\tau_p} \sum_{j=\tau_{p-1}+1}^{\tau_p} Gk\left(\boldsymbol{X}_i, \ \boldsymbol{X}_j\right),$$

where $p \in \{1, 2, 3, \ldots, K+1\}$ indicates the phase, τ_p denotes the last time point of phase p, and τ_0 and τ_{K+1} the first and last observations in the time series, respectively. The second subscript $(\tau_1, \ \tau_2, \ \ldots, \ \tau_K)$ of $\hat{V}_{p, \ (\tau_1, \ \tau_2, \ \ldots, \ \tau_K)}$ codes the locations of the phase boundaries and thus the set of change points on which the computations are based. The value obtained for $\hat{V}_{p, \ (\tau_1, \ \tau_2, \ \ldots, \ \tau_K)}$ indicates how homogeneous phase p is, when running from $\boldsymbol{X}_{\tau_{p-1}+1}$ to $\boldsymbol{X}_{\tau_p}$. Indeed, the more similar the observations in this phase are, the larger the sum that is subtracted by the rightmost term of $\hat{V}_{p, \ (\tau_1, \ \tau_2, \ \ldots, \ \tau_K)}$ and thus the smaller the intra-phase scatter. These $K+1$ intra-phase scatter values are then summed yielding the KCP criterion,

$$\hat{R}(\tau_1, \ \tau_2, \ \ldots, \ \tau_K) = \frac{1}{n} \sum_{p=1}^{K+1} \hat{V}_{p, \ \tau_1, \ \tau_2, \ \ldots, \ \tau_K},$$

which is minimized across all possible combinations of phase boundaries, $\tau_1, \ \tau_2, \ \ldots, \ \tau_K$. To clarify this step, let us consider a scenario where there is only one change point (i.e., K= 1), and thus two phases. To find the optimal change point location, $\hat{\tau}_1 + 1$, with $\tau_1 \in \{1, 2, 3, \ldots, n\}$, the trick is to minimize the total of the intra-phase scatters, $\hat{V}_{1, \ \tau_1}$ and $\hat{V}_{2, \ \tau_1}$ for the first and second phases, respectively:

$$\hat{\tau}_1 = \arg\min \hat{R}(\tau_1) = \frac{1}{n}\left(\hat{V}_{1, \ \tau_1} + \hat{V}_{2, \ \tau_1}\right)$$

In Fig. 2 (left panel), this minimization of the KCP criterion for the toy example is shown, where the time point, T= 100 generated the smallest $\hat{R}(\tau_1)$ implying a change point at T= 101.

[3] In a previous study [10] we set $2h^2 = med \parallel \boldsymbol{X}_i - \boldsymbol{X}_j \parallel^2$ in computing the Gaussian kernel as described in Arlot et al. [2]. In the present paper, however, we changed this to $h^2 = med \parallel \boldsymbol{X}_i - \boldsymbol{X}_j \parallel^2$ in line with the updates in the recent KCP paper [3].

Table 1
Obtained $\hat{R}_{min,K}$-value and change point locations for different values of K for the toy example.

K	$\hat{R}_{min,K}$	Change points
1	0.5046	101
2	0.4997	101, 255
3	0.4939	101, 238, 255
⋮	⋮	⋮
10	0.4671	101, 129, 131, 133, 152, 238, 255, 259, 262, 268

To extend to more than one change point, we generalize the notation and denote the smallest $\hat{R}(\tau_1)$–value for K= 1 as $\hat{R}_{min,1}$. The minimization procedure described above is conducted for every $K \in \{1, 2, 3, \ldots, K_{max}\}$, producing for each K a single $\hat{R}_{min,K}$-value and the corresponding optimal change point estimates, $\hat{\tau}_1 + 1$, $\hat{\tau}_2 + 1$, ..., $\hat{\tau}_K + 1$, are determined by

$$\hat{\tau}_1, \hat{\tau}_2, \ \ldots, \ \hat{\tau}_K = \arg\min \hat{R}(\tau_1, \ \tau_2, \ \ldots, \ \tau_K).$$

For the toy example, this procedure yields the change points tabulated in Table 1.

2.2.2. Choosing the number of change points

In the previous step, the change points were located conditional on the number of change points, K. In practice, the user often does not know the number of change points present in the data and has to estimate this as well. This decision cannot be based on $\hat{R}_{min,K}$'s, though, since this criterion decreases as K is increased (see Table 1). However, constructing a too complex model by adding superfluous change points will lead to negligible decrease in the criterion. Therefore, Arlot et al. [2] proposed to estimate the number of change points by choosing the K-value that minimizes the $\hat{R}_{min,K}$'s and a penalty that penalizes for the complexity introduced by adding extra change points,

$$\hat{K} = \arg\min \hat{R}_{min,K} + pen_K,$$

where $pen_K = C\frac{V_{max}\ (K+1)}{n}[1 + \log(\frac{n}{K+1})]$. The penalty coefficient, C, controls the influence of the penalty term and can be tuned as described below. The remaining constant, v_{max}, is derived by computing the trace of the covariance matrix for the first (also for the last) 5% observations in the time series, and choosing whichever is larger [2].

The choice of C directly affects the estimated number of change points. The larger C is, the stronger the effect of the penalty term, promoting less change points (smaller K's). To generate stable solutions, in line with Lavielle [21], we implemented a grid search,[4] which proceeds by setting C= 1, generating the maximum number of change points, K= $K_{\max}$, and then incrementing C, linearly, until it becomes too large such that the generated estimate for K becomes 0. We then choose the most stable K-value, which is the mode of all obtained K's across the considered C values. When the grid plot exhibits only two K-values, $K_{\max}$ and 0, we declare that there is no change point present in the data. This rule will lead to many type 1 errors, however, because often an intermediate K-value is observed for at least one C-value. We will address this issue by proposing a preliminary permutation test.[5]

In Fig. 2 (right panel), we show the grid search for the toy example, where K= 1 was the most often selected number of change points as indicated by the longest horizontal line. Given that the change point location for K= 1 is at T= 101 (see Table 1), we conclude that implementing KCP on the raw data successfully captured the first change point due to the mean change introduced. However, the second change point which reflects a correlation change was missed. Note also that even if K is fixed to two, the second estimated change point was located in the middle of the third phase which is far from the true correlation change point (Table 1).

2.3. KCP-corr: KCP on the running correlations

The Gaussian kernel employed in KCP-raw is a characteristic kernel, which by definition, is sensitive to any type of distributional change in the data. However, one can also plug in a non-characteristic kernel, which is designed for capturing specific types of changes [2,3]. To focus on correlation changes, therefore, the most intuitive approach is to find a non-characteristic kernel which is sensitive to correlation changes only. However, as far as we know, no kernel is proposed yet for this specific purpose. As a solution, we propose to first pre-process the time series so that only the information on correlational fluctuation is retained, and then implement KCP on this derived time series. Specifically, we propose to apply KCP on the running Fisher's Z transformed correlations, a multivariate time series with a dimension equal to $\frac{V(V-1)}{2}$, which is the total number of possible pairwise correlations when there are V variables in the raw data. However, other

[4] We note that Arlot et al. [3] recently proposed several ways of tuning C (e.g. slope heuristic), which all rely on the independence assumption. The grid search based on Lavielle [21], however, can also handle dependent cases and is therefore preferred in this paper.

[5] In Lavielle [21], it is mentioned that one can compare the maximum second derivative of $\hat{R}_{min,K}$ to a data dependent threshold to judge if there is no change point present in the data. Our proposed permutation test employs a very similar approach.

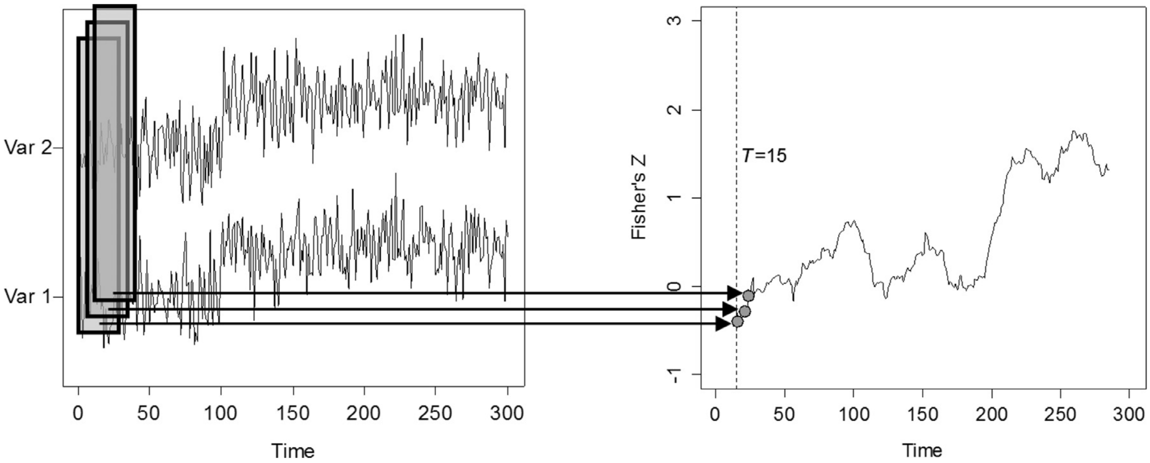

Fig. 3. Computation of the running correlation between the first and the second variables by sliding a window of 30 time points across the time series. The time point corresponding to the computed correlation is the midpoint of the window, such that for the first window, $\boldsymbol{X}_{1:30}$, the obtained running correlation is shown at $T= 15$.

correlation measures (e.g., distance correlation [35]) can also be plugged into the approach, although this may imply that a more appropriate kernel function has to be used. This choice, of course, depends on the specific research question and data at hand.

2.3.1. Computation of running correlations

The first step of KCP-corr is to derive the running correlations. This is done by sliding a window across the time series which moves one time point at a time. Per move, we compute the correlations using only the time points within the window. We will illustrate this procedure by picking the first and second variables from the toy example and using a window size of 30 time points. In Fig. 3, the first running correlation value is obtained from the first window, $\boldsymbol{X}_{1:30} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{30}\}$. The time point corresponding to this correlation value is set at the midpoint of the current window (time point before the midpoint if the window size is even), in this case $T= 15$ (Fig. 3, right panel).[6] The window is then moved one time point at a time, until the end of the series to generate the complete running correlation. The same process is replicated for all remaining variable pairs in the time series. For the toy example, there are three pairs of variables, thus generating three running correlations (Fig. 1, right panel).

Regarding window size, smaller window sizes will pick up smaller changes, but will also be more prone to noise. Larger window sizes, on the other hand, will lead to less noisy running correlations, yet smaller changes (low level of correlation change and/or small phase size) might be missed and change point locations might be harder to pinpoint. In the simulation study, we will investigate the optimal tuning of the window size.

2.3.2. Testing for the presence of at least one correlation change point and multiple change point detection

The next step is applying KCP on the obtained running correlations. We remark, however, that as stated in Section 2.2.2, the grid search is not optimal for cases when there are no change points because it yields too many false positives. We therefore propose to first test if at least one correlation change point is present in the data. To this end, we conduct a permutation test, that outperformed competitor methods in an extensive simulation study [9]. This test looks at how the obtained $\hat{R}_{min,K}$-values compare to those of a large number of permuted data sets (i.e., data sets in which the time points are randomly shuffled before the running correlations are computed). Specifically, we combine the results of two comparisons. We first run the variance test, which focuses on $\hat{R}_{min,K=0}$, a value that equals the variance of the running correlations. If there is a correlation change point, we expect the variance of the original running correlations to be significantly larger than the variances obtained from the running correlations for the permuted data sets. The p-value is therefore determined by computing the proportion of permuted data sets for which $\hat{R}_{min,K=0,perm}$ exceeds that of the original data.

The second aspect we examine is the maximum decrease in the intra-phase variance of the running correlations as a consequence of allowing for additional change points. We do this by conducting the variance drop test, which looks at the maximum $\hat{R}_{min,K} - \hat{R}_{min,K-1}$ across all $K \geq 1$. If a correlation change is present, we expect the maximum drop in the intra-phase variance of the running correlations to be significantly larger than that of permuted data. Thus, the p-value for this test is obtained by computing the proportion of permuted data sets for which the maximum intra-phase variance drop exceeds that of the original data.

[6] Other choices may also be possible here, but setting the correlation at the midpoint of the interval has an advantage that in case of a change point, the time point that is equally strongly affected by the two phases, is exactly the time point of the change point.

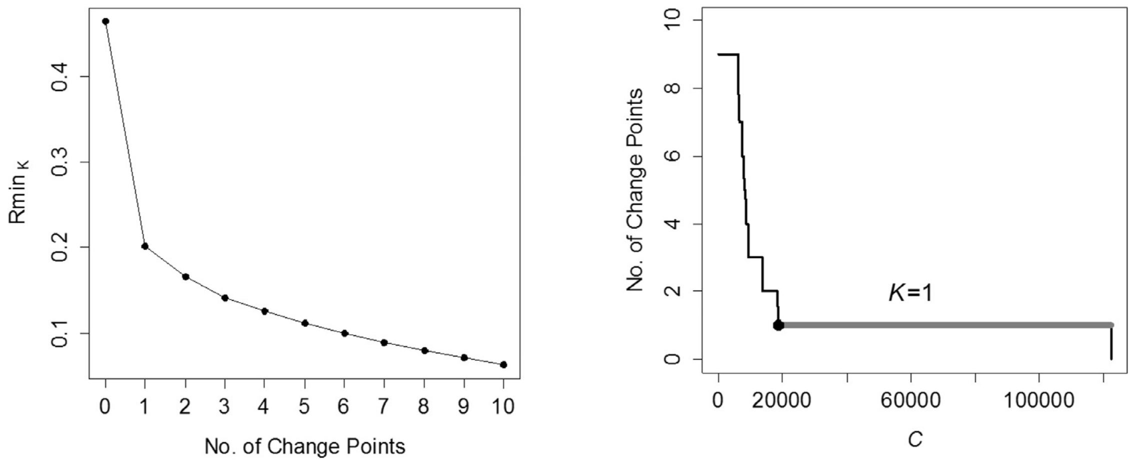

Fig. 4. KCP applied on the running correlations of the toy example. In the left panel, the $\hat{R}_{min,K}$-curve exhibits the expected downward trend for the total intra-phase variance as more change points are extracted from the running correlations. In the right panel, the tuning of the penalty coefficient is shown, implying that only one change point is present in the running correlations.

The test declares that there is a correlation change in the data if at least one of the two sub-tests is significant. Since there are two sub-tests for the KCP-based permutation test, we use the Bonferroni correction to account for the multiple testing. Note that the variance test would not be meaningful for KCP-raw, because, in contrast to the variance of the running correlations, the variance of the raw time series is unaffected by permutations of the time ordering.

To illustrate the significance testing, we first conduct the variance test on the running correlations. The $\hat{R}_{min,K=0}$ -value for the running correlations of the original data equals 0.4647 (Fig. 4, left panel), which after comparison with the values for 1000 permuted data sets yields a p-value of 0. For the variance drop test, we observe the maximum variance drop at K= 1 (Fig. 4, left panel), amounting to 0.2627. Using the same set of permuted data sets, the p-value for this test was also 0. Thus, the permutation test clearly suggests that there is at least one correlation change point present in the data. We thus proceed to locating the change points and choosing the optimal K for the data. The most stable K generated from the grid search is K= 1 (Fig. 4, right panel), and the change point corresponding to this solution is at T= 208. Thus, KCP-corr was able to capture the correlation change occurring at T= 201 (with a delay of 7 time points) which was not detected by KCP-raw. Since KCP-corr looks at the running correlations, it only signaled the correlation change point, while the mean change point, which we are not interested in, is not detected.

2.4. Cusum

The last method we consider is the Cusum based technique of Galeano and Wied [13], which is proposed for the same goal: detection of multiple correlation change points in a multivariate time series. This method compares the successive correlations, $\boldsymbol{p}_{1:i} = \begin{bmatrix} p_{1:i,1} \\ p_{1:i,2} \\ \vdots \\ p_{1:i,\frac{V(V-1)}{2}} \end{bmatrix}$, computed on the basis of time points 1 to i with $i \in 2{:}\ n$, to the overall correlations, $\boldsymbol{p}_{1:n} = \begin{bmatrix} p_{1:n,1} \\ p_{1:n,2} \\ \vdots \\ p_{1:n,\frac{V(V-1)}{2}} \end{bmatrix}$, computed using all time points. If there is no correlation change, then the successive correlations will not differ substantially from the overall correlations, such that their differences will fluctuate around zero. If there is a correlation change point, however, then its optimal location will be at the time point, $\hat{\tau}_1$, where the norm of the difference vector is maximal,

$$\hat{\tau}_1 = \arg\max \ \| \boldsymbol{P}_{\tau_1,\ n} \| = \arg\max \ \| \boldsymbol{p}_{1:\tau_1} - \boldsymbol{p}_{1:n} \| \ .$$

To test whether a change point is significant, the method therefore compares the standardized norm of the difference vector obtained at the optimal change point location to the distribution of the supremum of the sum of the absolute value of $\frac{V(V-1)}{2}$ independent Brownian bridges, where $\frac{V(V-1)}{2}$ is the total number of pairwise correlations in the data (for more details see Wied [40]).

To detect multiple change points, the significance testing is implemented sequentially. The optimal change point is first located and tested, and if found significant, the method proceeds to search for more change points within the generated phases. The search is terminated when no further significant change point is found. The method corrects for multiple testing through a Bonferroni-type correction, and as a final step, implements a refinement procedure to further prune the obtained change points.

When applied to the toy example, Cusum detects two change points located at T= 101 and T= 211. The first estimated change point corresponds exactly to the mean change point at T= 101, while the second successfully signaled the correlation change point T= 201 with a delay of 10 time points. Since the computed correlations can be influenced by mean changes, it is not surprising that Cusum is also sensitive to the mean change when the data is not centered. This, of course, also applies to KCP-corr to some extent, and we will further investigate this in the following section.

3. Simulations

To assess the performance of KCP-corr, we conducted two simulation studies based on the noise settings of Matteson and James [23]. In these settings only a subset of the simulated variables change in correlation; the rest are noise variables, which are unaffected by the change. We extended the settings by including more levels for (1) the number of variables, (2) number of noise variables and (3) strength of correlation change. We simulated data with noise variables because this is a realistic setting and because in previous studies it is shown that this is a challenging setting for most change point detection methods [9,10].

For the first simulation study, we aim to confirm that KCP-corr indeed outperforms KCP-raw in terms of capturing correlation changes. We also investigated whether the window size used influences KCP-corr's performance, and if so, which value yields the best results. We further compare KCP-corr to Cusum to assess whether KCP-corr can actually perform at par with a recently proposed method for correlation change detection. We expect that our settings will challenge the sensitivity of Cusum, since the first phase occurs again as the third phase, and recent simulations already revealed that the pooling of phases implied in the computation of the Cusum test statistic leads to the underestimation of the real magnitude of correlation change occurring in the middle phase [9].

3.1. Simulation study 1

For simulation study 1, the simulated data comprise 300 time points drawn from a multivariate normal distribution with all means equal to 0 and all variances equal to 1. Two change points are introduced to generate three phases of 100 observations each: in the first phase, all variables are uncorrelated, in the middle phase, a subset of variables becomes correlated, and in the third phase, all variables become uncorrelated again. The following factors were manipulated with 100 replicates per cell:

1 Number of variables V: 3, 5, 7, 9.
2 Number of correlating variables S: ranges from 2 until V-1 (the number of noise variables is V-S).
3 Strength of correlation change $\Delta\rho$: 0.3, 0.5, 0.7 and 0.9.
4 Window size w to calculate the running correlation: 25, 50, 75, 100.

We applied both KCP-raw and KCP-corr on the simulated data. The final number of change points for both methods was determined using the grid search described in Section 2.2.2. For the permutation test employed in KCP-corr, we used 1000 permuted data sets. We also applied Cusum, where 1000 bootstrap samples were used to estimate the empirical covariance matrix needed for the standardization of its test statistic and 10,000 sets of Brownian bridges were drawn to approximate the null distribution.

To evaluate the detection performance, we employed the Rand Index (RI) [31], which measures agreement between the true and the recovered phases. Specifically denoting the true time segmentation as $U = \{u_1, \ldots u_y\}$ and the recovered segmentation as $R = \{r_1, \ldots r_z\}$, where y and z are the respective number of phases, two agreement quantities are computed: the number of time point pairs that belong to the same phase in both U and R (which we denote as A) and the number of time point pairs that belong to different phases in U as well as in R (which we denote as B). Next, we compute RI by dividing the sum of A and B by the number of possible time point pairs:

$$Rand\ Index = \frac{A + B}{\binom{n}{2}}.$$

The value of RI lies between 0 and 1, implying non-agreement when its value is near 0 and perfect agreement as it approaches 1.

Fig. 5 displays the obtained RI's for all settings with correlation change, revealing that KCP-corr almost always strongly outperforms KCP-raw in these settings. The same conclusion follows from Fig. 6 which displays the location estimates of the obtained change points. The results of both methods clearly improve with increasing values of S and increasing strength of the correlation change (Fig. 5). KCP-corr benefits from using a smaller window size, in that most of the location estimates are closer to the real change points (see higher frequencies near the real change points in Fig. 6).[7] Detections resulting from

[7] In the Appendix (Fig. A1), we provided a plot of the Mean Absolute Deviation (MAD) between the estimated and true change point locations when the number of change points is set to two. MAD was obtained by first computing the mean absolute deviations per data set and then obtaining the mean

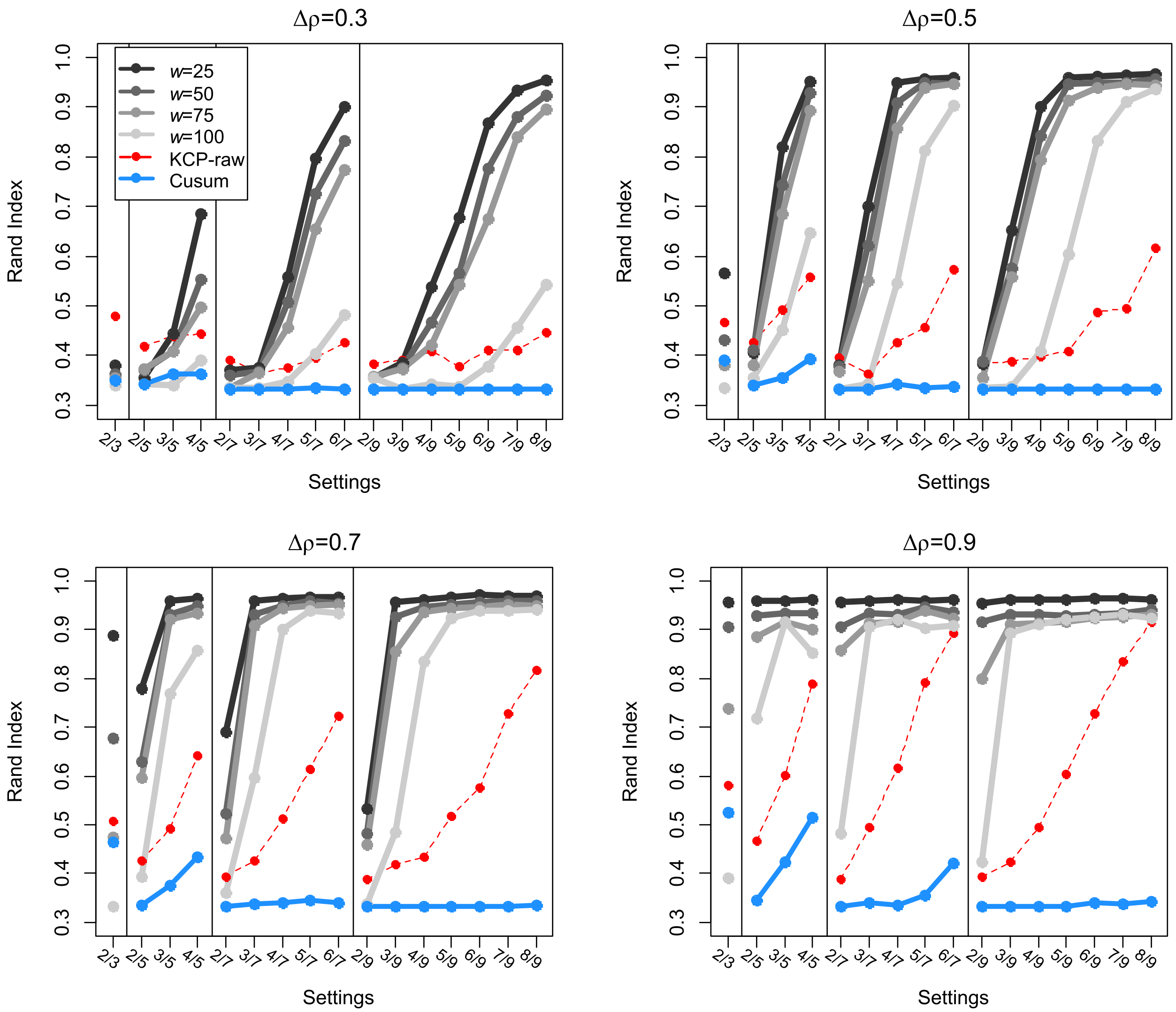

Fig. 5. Mean Rand indices (RI) for settings with correlation change in simulation study 1: the four panels correspond to the level of correlation change in the simulated data: 0.3, 0.5, 0.7 and 0.9. In every panel, mean RI's obtained using KCP-corr (employing varying window sizes: 25, 50, 75 and 100; in gray), KCP-raw (in red) and Cusum (in blue) were plotted across the simulation settings on the *x*-axis, where each setting is written as (no. of correlating variables)/(total no. of variables). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

larger window sizes have a larger bias (see more delayed detections after the first change point and more too early detections before the second change point for larger window sizes in Fig. 6). This bias can be observed in two cases: first, when the correlation change is small, it takes more observations from the new phase to be included in the moving window in order for the running correlations to exhibit the change, therefore resulting to delayed detections. Second, when the correlation change is large, it takes only a few new-phase observations to be included in the window for the running correlations to change, and this can result in too early detections. We note, however, that using a small window size is not a full-proof solution, since they are prone to outliers, which can easily influence the values of the running correlations. This might be the case for $w=$ 25, where some detections lie in the middle of the time series. Furthermore, the number of false detections at the beginning and at the end of the time series may appear to be higher for small window sizes, but this is, to a large extent, a methodological artefact. Of course, larger window sizes will not have many false detections in these regions simply because it is impossible for them to generate detections there: The midpoint of larger windows starts much later (earlier) at the beginning (end) of the time series.

Note that, next to the grid search, other procedures are available to choose the number of change points. Specifically, for KCP-raw, where the time points are independent, the slope heuristic approach [3] proposed by the KCP authors could be

for all data sets in each setting. For settings where the mean RI is greater than .80, implying that KCP-corr had a good recovery of the change points, the lowest MAD is generated by small window sizes.

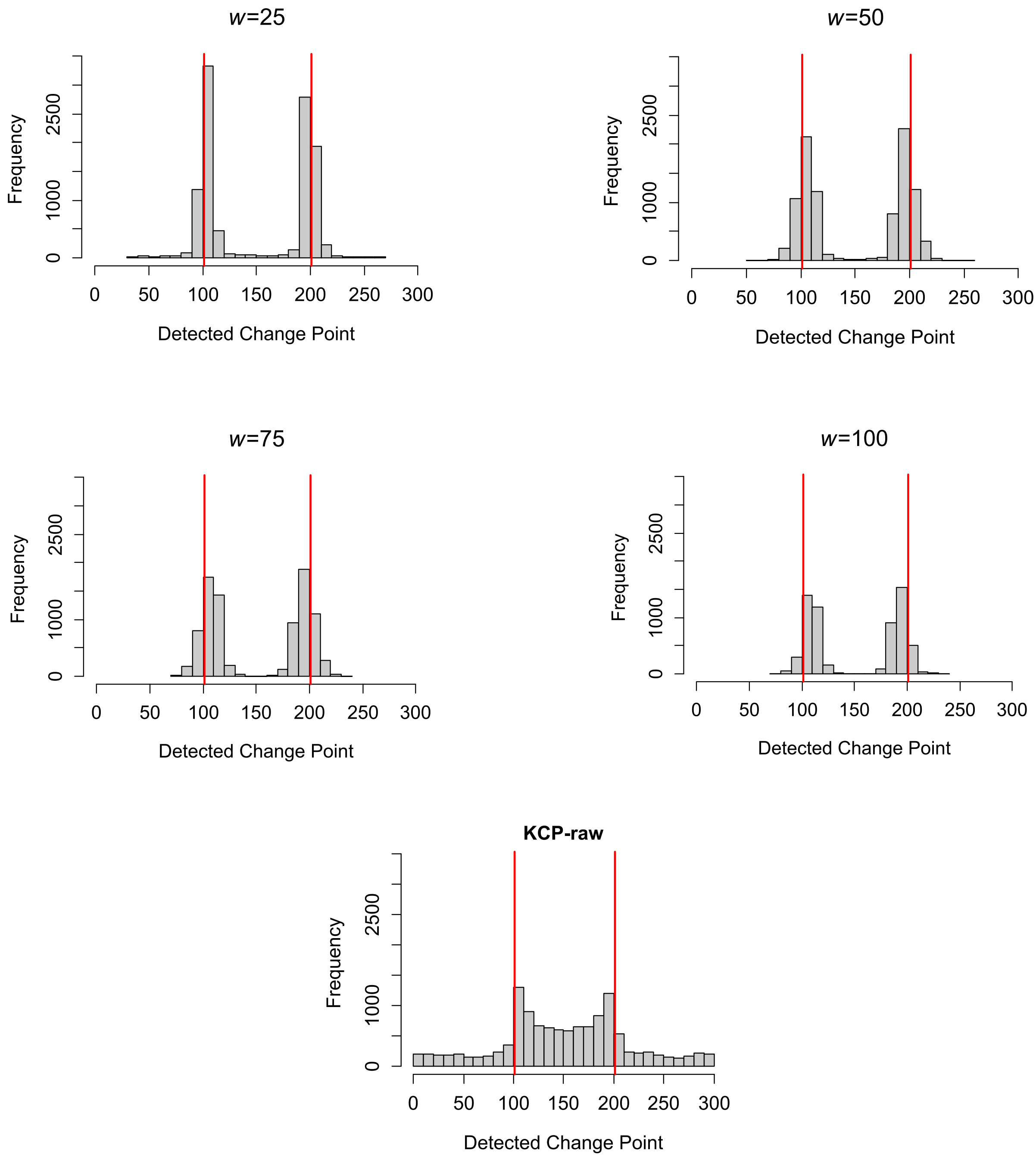

Fig. 6. Histogram of KCP-corr (employing different window sizes) and KCP-raw change points in simulation study 1. The upper four panels show the frequencies of detection using 4 window sizes: 25, 50, 75 and 100 time points for KCP-corr, while the lowermost panel exhibit those of KCP-raw. The change points are plotted for all settings in simulation study 1.

considered. We emphasize, however, that in this paper, it is not our goal to investigate which rule is optimal in choosing the number of change points, but rather, we aim to exhibit that KCP-corr locates the correlation change points better than KCP-raw. To provide further evidence for this claim, we also examined where the change points are located if the number of change points is fixed to two. Results showed that in all settings, RI's for KCP-corr are sizeably higher than those of KCP-raw (see Fig. A2 in the Appendix). When the location of change points is further scrutinized (see Fig. A3 in the Appendix), it

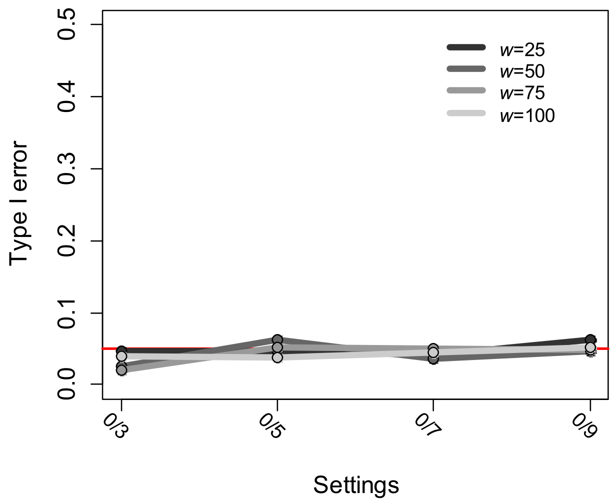

Fig. 7. The type I error rate for settings without correlation change: KCP-corr (with window sizes 25, 50, 75 and 100).

is clear that KCP-corr's detections are closer to the real correlation change points, while KCP-raw estimates are scattered throughout the time series.

In comparison with Cusum, KCP-corr proves to be the more powerful method for detecting correlation changes. It is evident that Cusum simply does not have enough power to detect the change points since its RI's remained around 0.33 across all settings, implying that it declares that the time series have no correlation change point. As expected, the recurrence of the zero correlations in the third phase, proved to be a difficult setting for Cusum. We note that if the last phase is discarded, Cusum will not perform as bad since the pooling of phases will not be too detrimental in estimating the effect sizes, and therefore would result to a more powerful significance test. We further remark that when Galeano and Wied [13] proposed this method, the minimum sample size they used in their simulations was N= 500 and performance quickly improved as they increased the sample size. In our simulations, however, the sample size was relatively small at N= 300.

To evaluate the performance of KCP-corr in settings without correlation change ($\Delta\rho = 0$), we looked at the type I error rate (using 500 replicates[8]) of the proposed permutation test when the nominal rate is set to 0.05. From Fig. 7, we conclude that KCP-corr's type I error rate is controlled around 0.05.

From these results, we conclude that smaller window sizes should be preferred when running KCP-corr, as they hold the nominal size and at the same time, yield the highest RI's. In the next simulation study, therefore, we will only employ window size w= 25.

3.2. Simulation study 2

In the second simulation study, we further assess the performance of KCP-corr in a wider array of settings to identify its strengths as well as its limitations when used in detecting correlation change points. In the previous simulation study, we only considered time series with change points exhibiting changes with equal effect sizes. In this section, we will examine the performance of KCP-corr in detecting change points when the effect sizes differ. Specifically, we want to investigate whether the locations of the change points are correctly retrieved by KCP-corr and whether the grid search used in determining the number of change points will be able to retain all relevant change points.

The simulated data are again comprised of 300 time points which were drawn from a multivariate normal distribution with all means equal to 0 and all variances equal to 1. In the first phase, all variables are uncorrelated implying $\rho_1 = 0$. In the middle phase, a subset of these variables exhibit a correlation of ρ_2, and finally in the third phase, the same subset change their correlations with a magnitude of ρ_3. The following factors were crossed with 100 replicates per cell:

1 Number of variables V: 3, 5, 7, 9.
2 Number of correlating variables S: ranges from 2 until V-1 (the number of noise variables is V-S).
3 Strength of correlation changes:

Easy setting: $\rho_1 = 0$, $\rho_2 = 0.9$, $\rho_3 = 0.3$ ($\Delta\rho_{12} = 0.9$ and $\Delta\rho_{23} = 0.6$)
Difficult setting: $\rho_1 = 0$, $\rho_2 = 0.3$, $\rho_3 = 0.9$ ($\Delta\rho_{12} = 0.3$ and $\Delta\rho_{23} = 0.6$).

[8] Since a few data sets can greatly influence the computation of the type I error rate, we used more replicates for settings with no correlation change.

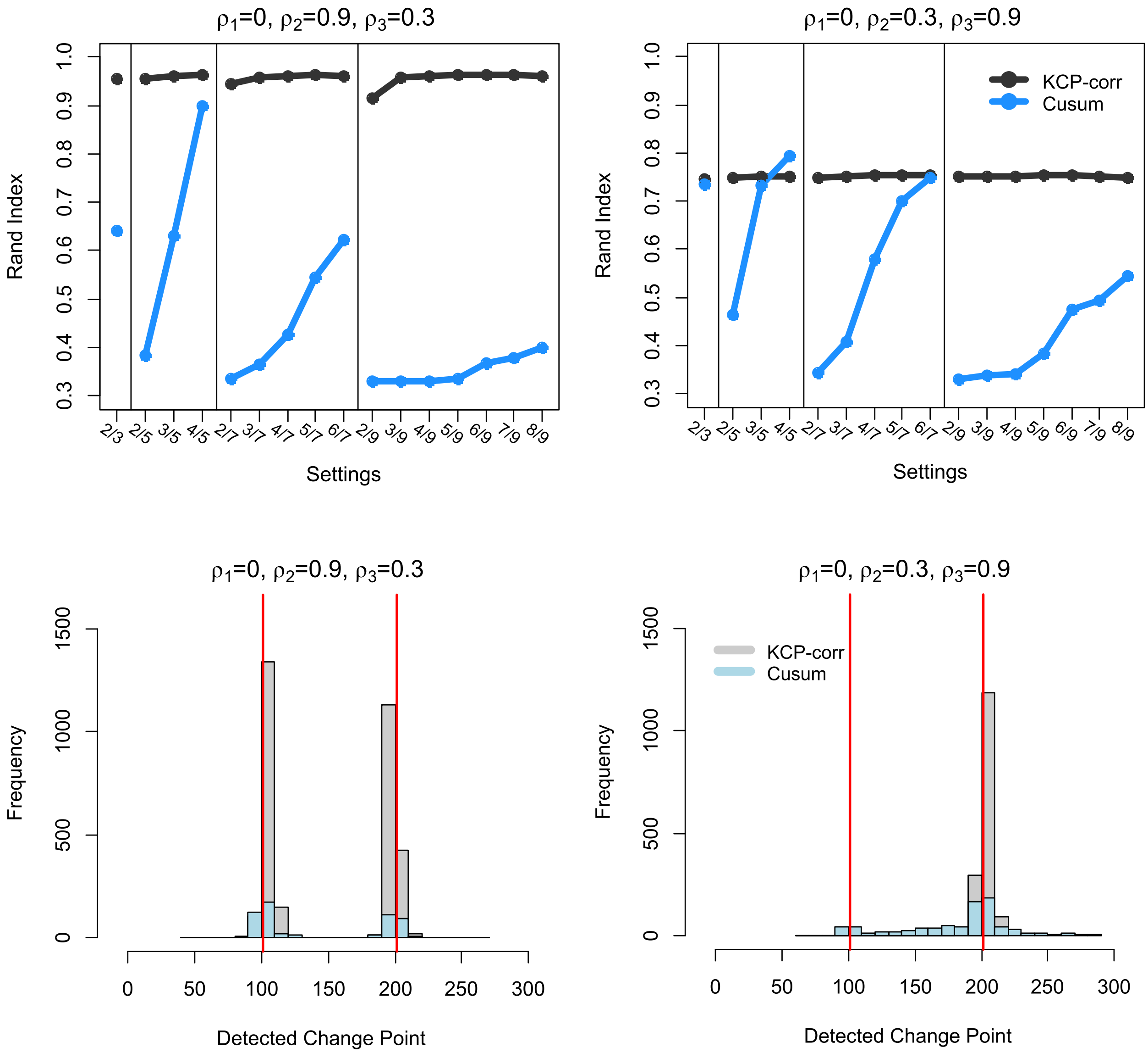

Fig. 8. Mean Rand indices (RI) and detected change points for settings with unequal effect sizes. In the upper panels, mean RI's obtained using KCP-corr (w= 25, in dark gray) and Cusum (in blue) were plotted across the simulation settings on the x-axis, where each setting is written as (no. of correlating variables)/(total no. of variables). In the lower panels, the frequency distribution of the detected change points is displayed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Since in these simulation settings the correlations do not return to zero in the third phase, it is interesting to investigate whether Cusum catches up in terms of performance with KCP-corr. Hence, we also applied Cusum in this simulation study and assessed how the two methods perform relative to each other.

When the effect sizes are unequal but large, KCP-corr is able to detect the change points. This is exhibited by the high RI's for the easy setting (Fig. 8, upper left panel) where both effect sizes are substantial ($\Delta\rho_{12} = 0.9$ and $\Delta\rho_{23} = 0.6$). For the difficult setting, however, where the effect size for the first change point is small, KCP-corr's RI was consistently around 0.75 (Fig. 8, lower left panel). When the detections are examined (Fig. 8, lower right panel), we see that only the second change point with the larger effect size ($\Delta\rho_{23} = 0.6$). is detected, while the first one with the smaller effect size ($\Delta\rho_{12} = 0.3$) is always missed. This is rather expected since in the first simulation study, an effect size of 0.3 was indeed a difficult setting, where KCP-corr mostly yielded low RI's (Fig. 5, upper left panel). A closer look at the results for these settings, however, reveals that two change points can be detected even if the effect size is as small as 0.3 as long as many variables are correlating (e.g., 6 out 7 variables, 7 or 8 out of 9 variables). This is not the case for the difficult setting in the second simulation study, where KCP-corr always yielded one instead of two change points. We note, however, that if we do not employ the grid search and the number of change points is fixed to two, KCP-corr's detections are close to the two

Table 2
Detection rate of KCP-corr on data sets with change points caused by changes in mean, variance and autocorrelation. The simulated data is comprised of 5 variables with 200 time points drawn from a multivariate normal distribution. In the first phase, $X_{1:100} = \{X_1, X_2, \ldots X_{100}\}$, the variables have zero means and unit variances. In the second phase, $X_{101:200} = \{X_{101}, X_{102}, \ldots X_{200}\}$, a change point is introduced while varying two factors: type of change and amount of change. We generated 100 replicates per cell.

Type of change	Amount of change	Detection rate
Mean	New mean (from 0)	
	1sd	7%
	3sd	100%
	5sd	100%
Variance	New variance (from 1)	
	4	2%
	9	9%
	25	25%
Autocorrelation	New autocorrelation (from 0)	
	.25	5%
	.50	4%
	.75	18%

Table 3
Detection rate of KCP-corr on data sets with induced outliers. The simulated data is comprised of 5 variables with 200 time points drawn from a multivariate normal distribution. For settings without correlation change, all time points have zero means and unit variances. For settings with correlation change, the variables have zero means and unit variances in the first phase, $X_{1:100} = \{X_1, X_2, \ldots X_{100}\}$. A correlation change point is introduced in the second phase, $X_{101:200} = \{X_{101}, X_{102}, \ldots X_{200}\}$, such that all variables have a correlation equal to 0.7. We introduced outliers, systematically varying the magnitude of outlyingness and the contamination rate, and generated 100 replicates per cell.

Correlation change point	Outlyingness	Contamination rate (%)	Detection rate (%)
0	3sd	5	1
		10	5
	5sd	5	3
		10	5
1	3sd	5	100
		10	92
	5sd	5	91
		10	50

real change points and the RI's become impressively high again (see Fig. A4, right panels, in the Appendix). These results therefore suggest that the grid search may miss change points signaling small correlation changes in the data for some settings with unequal effect sizes.

Although Cusum yielded higher RI's than in the first simulation study, these were generally inadequate (most RI's < 0.80) and consistently lower than those of KCP-corr in all but one simulation setting (difficult setting: 4 out 5 variables). Hence, the pooling of phases still causes problems for CUSUM, as both the successive correlation and the overall correlation are expected to be weak. This leads to a small Cusum test statistic and thus low power to detect a significant change point. We also remark that the number of variables seems to have a systematic influence on Cusum's performance. The RI's reach their maximum when there are 4 out of 5 variables correlating. When there are more variables in the system, say 7 or 9, the RI's are generally lower even if the majority of the variables exhibits correlation change.

Finally, we have proposed KCP-corr to specifically target correlation changes. However, the method involves computation of the running correlations which might actually also reflect changes in other parameters. We tested a number of different settings, revealing that KCP is robust to small mean (variance and autocorrelation) changes, but will be influenced by considerably large changes in them (Table 2). Such changes would also be problematic for competing methods however, as they assume constancy of mean and variance to ensure that the change detected can be attributed to correlation change [4,13,40]. In terms of robustness to outliers, KCP-corr holds the nominal size despite high contamination with outliers (see settings with no change points in Table 3). However, in terms of power, performance may decrease if the outliers are extremely outlying and/or densely present in the data set (see settings with a correlation change point in Table 3).

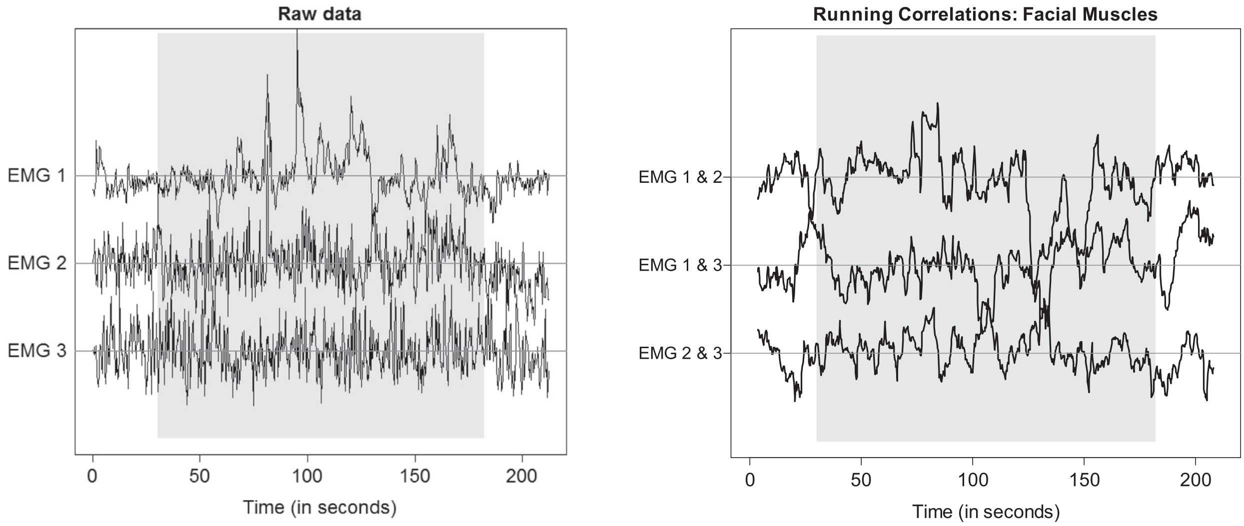

Fig. 9. Scaled electromyography (EMG) variables (*m. corrugator supercilii, m. zygomaticus major* and *m. orbicularis oculi)* of a participant viewing a fear eliciting film clip and their running correlations. The left panel shows the standardized data (horizontal lines indicate the overall mean), comprised of 848 time points (212 s with 4 measurements per second). The shaded area is the duration of the film clip (152.25 s), while the unshaded areas before and after are the baseline periods (30 s starting baseline and 29.75 s ending baseline). The right panel plots the running correlations (horizontal lines indicate zero correlation level) for all possible EMG variable pairings, computed using a window size of 30 time points (equal to a length of 7.5 s).

4. Illustrative application

To further illustrate the use of KCP-corr and the difference with KCP-raw and Cusum, we analyzed empirical data collected by Wilhelm et al. [41] on emotional reactivity. In their study, 52 participants were shown film clips expected to elicit different emotions such as fear, sadness, pride, and serenity (compared to neutral). To assess the different facets of participants' emotional reactions, experiential (valence and arousal), behavioral (facial muscle activity indicative of facial emotion expression), autonomic (cardiovascular activity, skin conductance), and respiratory signals were collected during the experiment and analyzed using ANSLAB [6]. Only 44 out of the 52 participants have yielded time series with valid measurements. The resulting sets of measures represent different systems, which are postulated by emotion psychologists to synchronize in response to emotional stimuli [25]. However, aside from synchronization across systems, (de)synchronization within systems can also be hypothesized. Specifically, stronger positive correlations can be exhibited by variables involved in the emotional response, while zero correlations might be observed between involved and uninvolved variables. Previous studies, however, indicate that response synchronization is hard to find in empirical data [8,24] and may manifest in different ways across individuals [15]. In this illustrative application, therefore, we will examine whether we can find empirical evidence for the presence of synchronization and whether it manifests in different ways and different time points across persons. We will focus on the within-system synchronization in the behavioral system. Based on visual inspection, level changes in this system are not very pronounced compared to other systems. This characteristic makes it a perfect context to observe differential performance of KCP-corr and KCP-raw since there might be change points caused by pure correlation changes (without mean changes), which are more difficult to detect for KCP-raw than KCP-corr.

We analyzed the time series obtained from watching a film clip (*Copycat*) which is expected to elicit fear. We will first discuss the results for one participant in detail and then give a summarizing overview of the main findings for all participants. For the chosen participant, the analyzed time series is 212 s in length and consists of a 30 s starting baseline preceding the clip, a 152.25 s film clip on a perpetrator imposing physical threat to the protagonists, and a 29.75 s ending baseline after the clip. The behavioral response was assessed by monitoring the electrical activity of 3 facial muscles: *m. corrugator supercilii* (EMG1), *m. zygomaticus major* (EMG2) and *m. orbicularis oculi* (EMG3), considered to be involved in the expression of emotion. Electromyography (EMG) amplitude (in μV) for these 3 muscles were filtered, rectified and exported at a sampling rate of 4 measurements per second, obtaining a multivariate time series comprised of 3 variables with 848 time points. The scaled EMG data are plotted in Fig. 9 (left panel), where the shaded area corresponds to the duration of the film clip. The right panel of this figure shows the derived running correlations, using a window size of 30.

When applying KCP to the scaled EMG time series, three change points (see Fig. 10, left panel) are detected at 94 s, 129.5 s and 182.5 s, segmenting the time series into 4 phases. The first change point at 94 s is detected during a scene where one of the protagonists was suddenly startled by falling objects from the cubicle she opened. The second change point at 129.5 s is detected about 4 seconds after the killer captures this woman. The last change point almost perfectly coincides with the ending of the film at 182.5 s. From Fig. 11 (left panel), it is noticeable that the detected changes are level or mean changes, whereas none of them reflect significant differences in correlations. Specifically, EMG1 elevated in the second phase

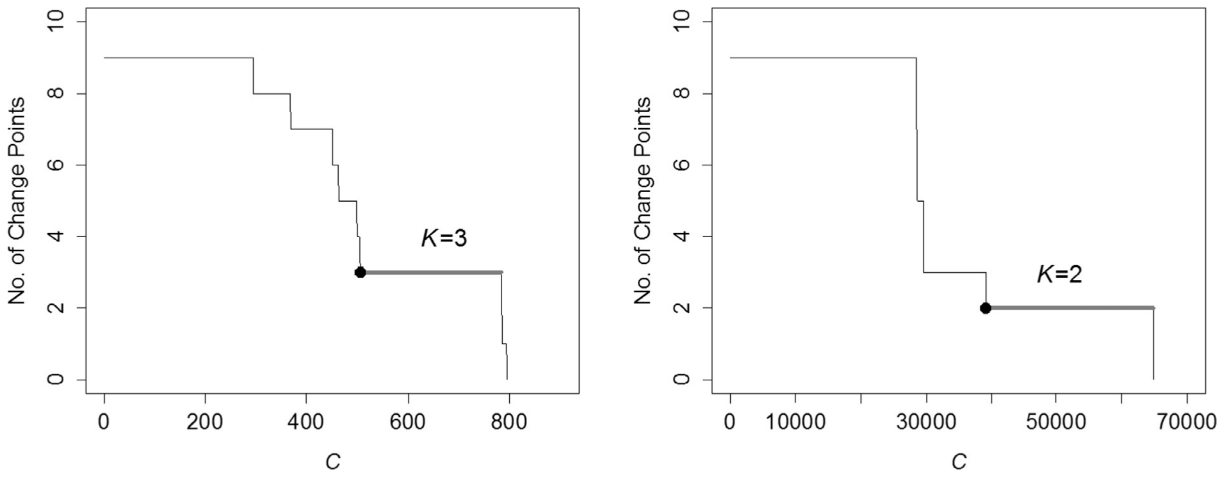

Fig. 10. Tuning of the penalty coefficient when implementing KCP-raw (left panel) and KCP-corr (right panel).

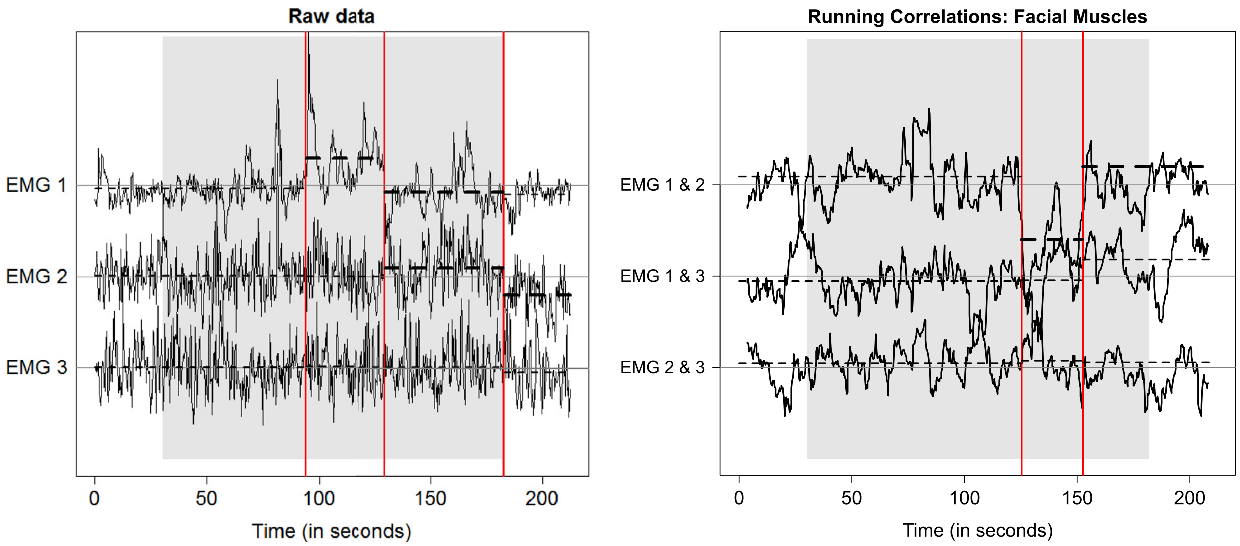

Fig. 11. Change points (in red) generated using KCP-raw (left) and KCP-corr (right). The broken lines indicate the phase means and correlations, respectively. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

and then reverted back to a lower level in the third phase. EMG2, on the other hand, slightly increased in the third phase and then declined in the fourth phase. All these mean changes were found significant using the Mann-Whitney test with a Bonferroni correction for multiple testing (overall $\alpha = 0.05$). It is noteworthy that EMG1 and EMG3, pertain to muscles situated near the eye and are responsible for the wrinkling of eyebrows (EMG1) and narrowing of the eyes (EMG3), while EMG2 measures activity of the muscles typically responsible for smiling, but in this context may also be activated due to intensely fearful facial expressions being characterized by a pulling back of the mouth angles [11].

For KCP-corr, we used a range of window sizes $w \in \{30, 40, 50, \ldots, 100\}$. The smaller window sizes (up to $w = 70$) consistently resulted in two change points at very similar locations (see Fig. 10, right panel, for w= 30), while for larger window sizes the permutation test was non-significant. Since it was concluded from the simulations that smaller window sizes can capture change point locations better, we maintained the solution for the smallest window size, $w = 30$, shown in Fig. 11 (right panel). The change points are located at 125.50 s and 152 s (indicated by the red vertical lines). Interestingly, the second phase corresponds to the exact instant when the police woman is seized physically by the killer until the moment when she attempted to break away from him. These scenes are further marked by a sudden shift in the musical score indicating that something eventful is taking place.

The mean correlations per phase are indicated by the broken horizontal lines. In the first phase, all variables are almost uncorrelated, such that all pairwise correlations lie close to the zero-line. In the second phase, EMG1 and EMG2 exhibited a strong negative correlation, which became weakly positive in the third phase. It is noteworthy that these significant changes involve muscles that have been found to be related to affect, with EMG1 being strongly linearly related to negative emotions

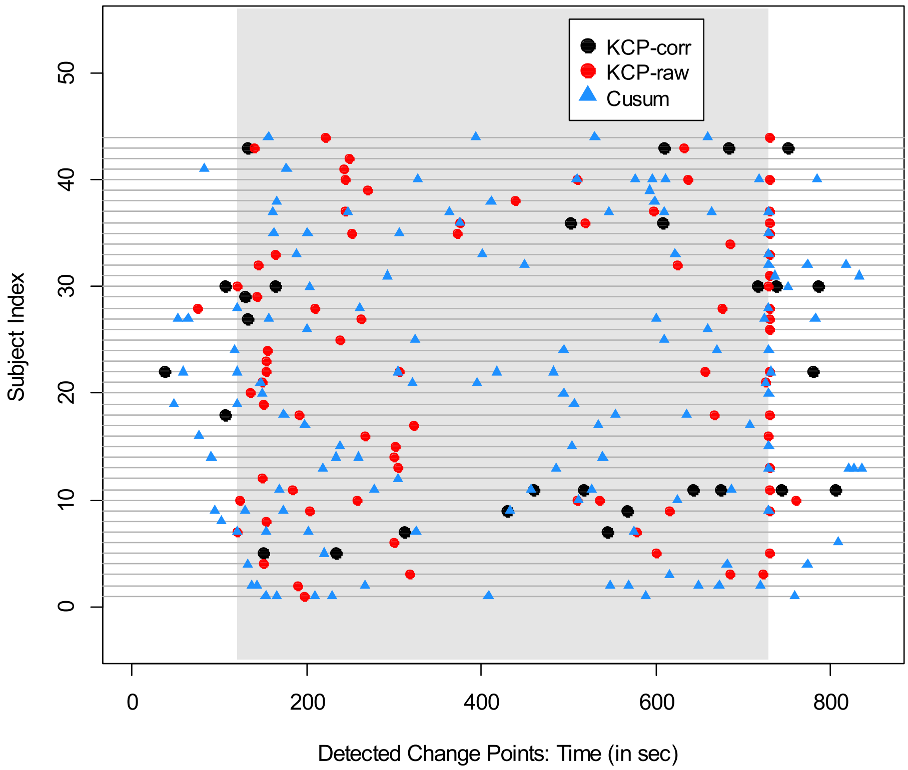

Fig. 12. Change points detected by KCP-corr, KCP-raw and Cusum in the emotional reactivity data of 44 subjects. The shaded area is the duration of the film clip (152.25 s), while the unshaded areas before and after are the baseline periods (30 s starting baseline and 29.75 s ending baseline). KCP-corr, KCP-raw and Cusum change points are indicated by black dots, red dots and blue triangles, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Correlation changes for KCP-corr generated phases.

Pair	Phase 1 to Phase 2		Phase 2 to Phase 3	
	$\Delta\rho$	*p*-value	$\Delta\rho$	*p*-value
EMG 1 and 2	−0.68	<0.0001	0.80	<0.0001
EMG 1 and 3	0.01	0.0931	0.23	0.0430
EMG 2 and 3	0.02	0.0837	−0.02	0.0886

(see Larsen, Norris, and Cacciopo [20] and references therein). For pairs associated with EMG3, no significant correlation changes were found (see Table 4).

Enumerating the insights gained from applying both KCP-raw and KCP-corr to these data, we first conclude that changes in mean level and correlation occurred within the film clip and could be meaningfully related to particular scenes within the clip for the chosen participant. Second, the solutions from the two approaches differ, with KCP-raw signaling subtle level changes rather than correlation changes. Finally, KCP-corr reveals correlation changes, which were not captured by KCP-raw at all. KCP-corr, therefore, enabled us to obtain evidence supporting (de)synchronization of behavioral measures in response to emotional stimuli.

Turning to the results for all participants (Fig. 12), the KCP-raw and the KCP-corr change points are not situated in the same locations, again evidencing that KCP-corr can unravel correlation changes which were not detected by KCP-raw, as was already shown in the chosen participant. In terms of the occurrence of synchronization, only 11 out of 44 subjects had significant correlation changes according to KCP-corr. For the remaining 33 participants, KCP-corr did not detect any change point, implying that synchronization might have been weaker or even absent in most subjects. This confirms the results of previous studies [8,24] where response synchronization during an emotional episode was found elusive and rather hard to capture. On the other hand, we do not rule out the possibility that KCP-corr might have insufficient power to detect the probably subtle correlation changes in this data set. For Cusum, a somehow different picture emerges. It recovered correlation change points for 39 out of 44 participants, and most of these have multiple change points which are scattered along the time series. Despite the refinement step that the method employs, there are extremely close change points, which should be further scrutinized as it seems unrealistic to observe correlation changes occurring so closely to each other. We note, however, that we did not correct for any mean, variance, or other parameter changes in the data, and Cusum might be highly sensitive to these.

Focusing on the cases where KCP-corr detected change points, the location and the number varied across participants. This is not entirely unexpected since response synchronization has been reported to manifest in varied ways across individuals [8,15,24]. We also note that the stimulus used in this experiment is a film clip comprised of multiple events which can have differential effects on the participants. To validate our results, we would therefore need to know whether or not participants experienced some emotion(s) during the film clip, and if so, during which event(s) exactly. This way, the location of the subject-specific emotion experiences would be known, and we could examine whether the obtained correlation change points coincide. In addition, this information would allow identifying possible spurious change points (e.g. change points because of random facial movements in the baseline). Unfortunately, this information is not available in this data set.

Of course, it is also possible that the large differences across participants were caused by methodological drawbacks of KCP-corr. In terms of biases in change point location, we have shown in the simulations that KCP-corr yields estimates which are around the true change point, hence this seems less of a concern. However, in terms of retaining the correct number of change points, two issues should be pointed out. First, change points might be missed by the grid search if there are correlation changes of unequal magnitude (see simulation study 2). It is reasonable to assume that this might be the case for some participants. Second, we cannot completely rule out that, similar to what happens for CUSUM, some KCP-corr detections could have been caused by other parameter changes (e.g. variance, autocorrelation, etc.) during the experiment (see simulation study 2).

5. Discussion and conclusion

We have proposed a new non-parametric approach for capturing correlation changes, called KCP-corr, which builds on the Gaussian kernel based change point detection method proposed by Arlot et al. [2,3]. Given that the Gaussian kernel is a characteristic kernel, the original KCP approach, KCP-raw, may reveal abrupt changes in different statistics: mean, variance, autocorrelation, correlation and even higher moments. Thus, KCP-raw can be considered a rather exploratory tool, requiring additional analyses to verify which statistics exhibited the change. In contrast, KCP-corr applies KCP on the running correlations to focus the change point detection on correlation changes. Thus, it is a more confirmatory tool to test whether or not a correlation change occurred.

Applying both methods to emotional reactivity data revealed that, unlike for KCP-corr, KCP-raw results may be dominated by level changes (and possibly, changes in other statistics which were not further investigated here), shifting attention away from correlation changes at other time points. Unexpectedly, Cusum, a competing method which has been shown to have less power for detecting correlation changes than KCP-corr in simulations (see also Cabrieto et al. [9]), also yielded more change points than KCP-corr, suggesting that it can be highly sensitive to changes other than correlation. For the majority of the participants we did not find evidence of correlation change points with KCP-corr, and for those participants where change points were detected, the number and location differed across individuals. This finding reveals that synchronization is rather elusive in this specific application, which is in line with the earlier results of Bulteel et al. [8] and studies included in the review of Mauss and Robinson [24]. Of course, the low number of participants with detected change points might also indicate low sensitivity for KCP-corr. However, we remark that in a previous study where we re-analyzed empirical data with substantial correlation changes during a well-defined event (stocks data [13], EEG epileptic seizure data [38], critical slowing down data [39]), KCP-corr was the most sensitive[9] among competing methods, in general [9].

Based on our simulation results, KCP-corr strongly outperforms KCP-raw when the level of correlation change is lower and noise variables are present. For KCP-raw, even with very high correlation changes ($\Delta\rho = 0.9$), detection performance drastically dropped as the number of noise variables increased. For KCP-corr, on the other hand, as long as the correlation changes are strong ($\Delta\rho \geq 0.7$) the number of noise variables did not severely hamper sensitivity. It is only for moderate correlation changes $\Delta\rho = 0.5$), that presence of noise variables started to weaken the detection performance of KCP-corr. Yet as long as the majority of the variables are involved during a correlation change, one can still expect a reasonable detection performance. For low correlation changes ($\Delta\rho \leq 0.3$), it requires 5 variables or more to correlate for KCP-corr to yield an acceptable RI, while KCP-raw did not recover the change points adequately, revealing that detection of minimal correlation changes remains a difficult task. Cusum, on the other hand, performed very badly in our simulation settings. This is probably due to the pooling of phases that is implied by Cusum, which is detrimental in case particular correlation phases recur across time.

Given that KCP-corr outperforms KCP-raw and Cusum, we now have a competitive non-parametric method focused on correlations, which can locate multiple change points simultaneously. It is worthwhile, therefore, to work on further improving KCP-corr. First, future research can investigate whether the use of a different kernel might make the method more sensitive to correlation changes. The present implementation looks at the similarities between running correlation vectors using a Gaussian kernel. An alternative approach is to look at running correlation matrices and use an appropriate kernel which is able to measure similarities between them. If one is willing to assume normality and constant mean for instance,

[9] This study introduced the KCP permutation test for the running correlations, hence focusing on whether change points are present in the time series or not and not on their number and location [19].

the Kullback–Leibler based kernel suggested in Moreno et al. [26] seems a reasonable choice. We expect, however, that tuning of some parameters would be needed to achieve a good performance. We also stress that for now, the pre-processing step which generates the derived time series (e.g. running correlation vectors, running correlation matrices) is needed even if a new kernel is used. To the best of our knowledge, there is no kernel proposed yet which can directly capture correlation differences only without transforming the raw time series data first. The pre-processing step of KCP-corr, however, may yield some computational challenges. While KCP-raw looks at V variables, KCP-corr computes and inspects $\frac{V\ (V-1)}{2}$ variables (all pairwise correlations). We further remark that the permutation test we introduced for KCP-corr requires additional running time because each permuted data set has to be analysed as well. This may render running KCP-corr infeasible in case of a large number of variables. Both limitations, though, also hold for the techniques of Galeano and Wied [13] and Barnett and Onnela [4], where the correlation matrix is computed for each time point, successively, and some form of resampling procedure is carried out to derive the test statistic and/or null distribution. When a sufficient number of time points is available, therefore, one may try to reduce computation time for KCP-corr by decreasing the overlap of the moving windows. Further study, however, should investigate how this would influence the accuracy of the estimated change point locations. In this regard, we also note that similar to competing methods [13,4], the current implementation of KCP-corr cannot provide confidence intervals around change point locations. Developing ways to estimate such intervals would be extremely useful, especially for applied researchers.

A second direction is to explore other approaches to optimally choose the number of change points. This development would be crucial for two reasons. First, as was seen in simulation study 2, change points can be missed by KCP-corr when correlation changes have unequal magnitude. Second, as observed in the illustrative example, too many (and in some cases, too close) detections might also be obtained. One possibility, given that the use of Gaussian kernel is maintained, is to further optimize the tuning of the bandwidth parameter as this choice can influence the number of retained change points. Another option is to improve the grid search we are currently using, as the results of the second simulation study revealed that it is prone to miss the small effect sizes. Lavielle [21] already suggested to also look at the next best solutions in the grid search as they usually yield more change points and thus may include smaller changes as well. It would be useful to devise a strategy on how to do this optimally, such that there is a balance between retaining smaller changes and controlling the rate of false detections. Other approaches such as turning to techniques for choosing the elbow in a plot of the average intra-phase variance versus the number of change points can also be considered. These approaches, though, should be able to handle the dependency inherent in the running correlations.

The third direction is to further investigate how to focus KCP-corr on correlation changes even more. Indeed, in simulation study 2, we saw that KCP-corr can be influenced by large changes in mean, variance and autocorrelation. Furthermore, in the illustrative example, there were also really close change points, which might be due to dramatic changes in parameters other than the correlation. The user can tackle this by ensuring that changes from these sources are duly filtered out (e.g. centering and scaling techniques, using residuals from a fitted AR model). However, we also acknowledge that some of these changes are difficult or even impossible to disentangle from correlation change. Take for instance a simultaneous change in mean for multiple variables. Of course, when one would look at the whole time series, one would judge this change as a mean change. However, this mean change induces a correlation change as well because several variables change in their levels at the same time. Moreover, the presence of outliers may influence KCP-corr's power, implying that it is recommended to remove excessively outlying values from the time series before implementing KCP-corr.

Given that we obtained promising results when forcing KCP to focus on capturing correlation changes, we can further attempt to implement the same approach to detect changes in other statistics such as eigenvalues, cross-lags, or autocorrelations. In this way, more interpretable phases can be obtained since one already knows beforehand what type of change KCP is targeting. To this end, the running correlations could be replaced by running eigenvalues, running cross-lags, or running autocorrelations. For instance, it has been shown by Müller et al. [27] that changes in the running eigenvalues can be tracked to study the collective behavior of complex quantum systems such as atomic nuclei and molecules. Another potential field of application is emotion psychology, where exact timing of changes in autocorrelation of affective states (emotional inertia or the amount of carry-over of an emotion from one moment to the next [18]) or cross-lags of emotions (augmentation and blunting of one emotion over another emotion at a previous time point [29]) can be captured to provide a more informative picture of the temporal dynamics of the emotion and consequently, the adaptive or maladaptive emotional functioning of the individual. If this adaptation is pursued, the user should be well aware however that different kernels may work better for different derived time series and/or different statistics. Finally, we stress that in our implementation where we use the Gaussian kernel to compare vectors of running correlations, we in fact converted the problem of detecting changes in specific features to detecting changes in their running mean. Thus, a reasonable alternative approach is to apply methods that directly target mean changes to these derived time series. Preferably, these methods should be able to handle the inherent dependency in these time series.

In conclusion, we have demonstrated that a general purpose non-parametric change point detection method such as the Gaussian kernel based KCP can be adapted to focus on a single multivariate statistic, by applying it to a derived time series reflecting fluctuations in the correlation. This adapted KCP method is more sensitive to changes in the statistic concerned than merely applying KCP on the raw data, particularly in difficult cases when changes are minimal and/or noise variables are present. It also outperforms Cusum, a method which explicitly focuses on detecting multiple correlation change points.

Appendix

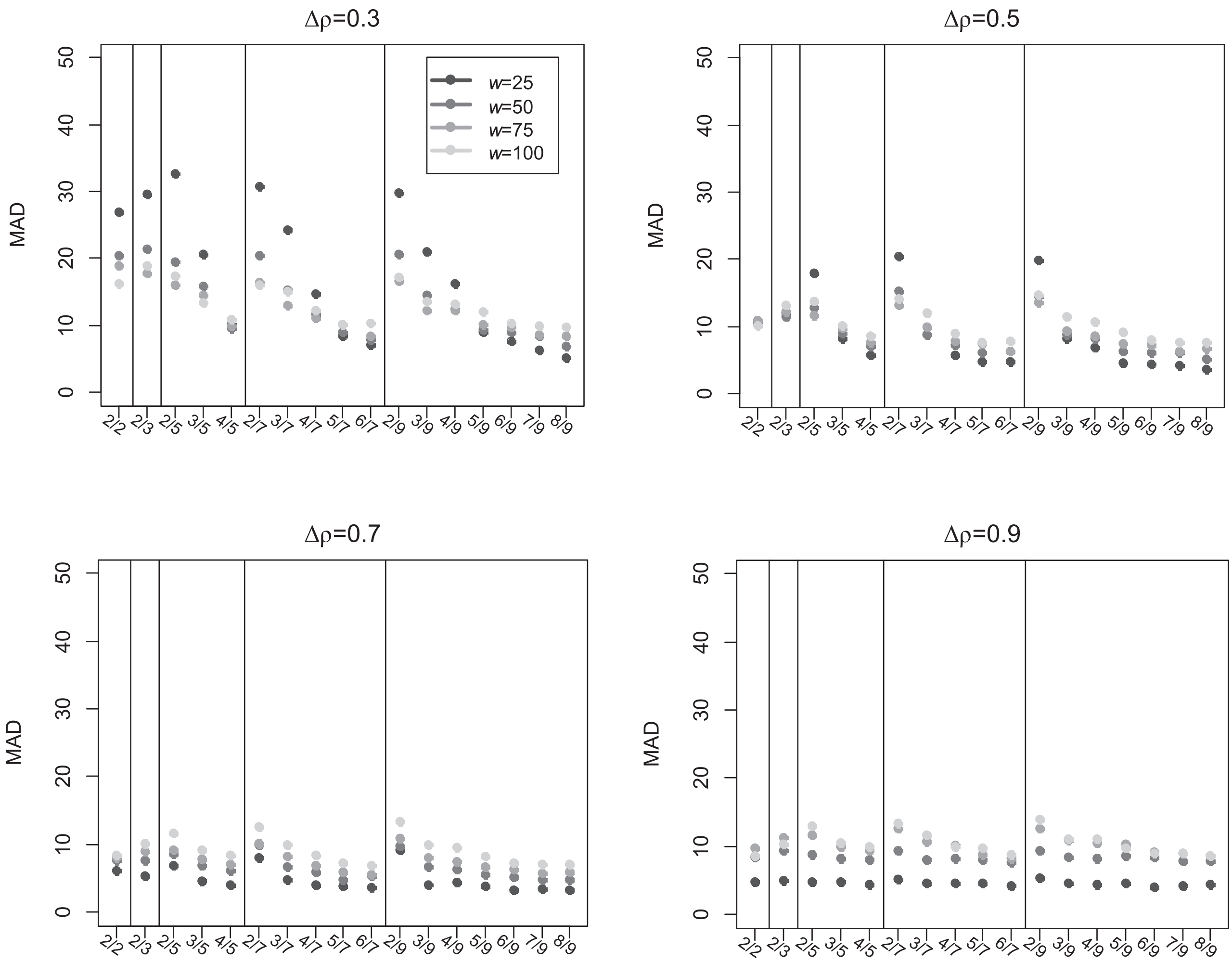

Fig. A1. Mean Absolute Deviation (MAD) for settings with correlation change in simulation study 1 (number of change points is fixed to 2): The four panels correspond to the level of correlation change in the simulated data: 0.3, 0.5, 0.7 and 0.9. In every panel, MAD obtained using KCP-corr (employing varying window sizes: 25, 50, 75 and 100) were plotted across the simulation settings on the *x*-axis, where each setting is written as (no. of correlating variables)/(total no. of variables).

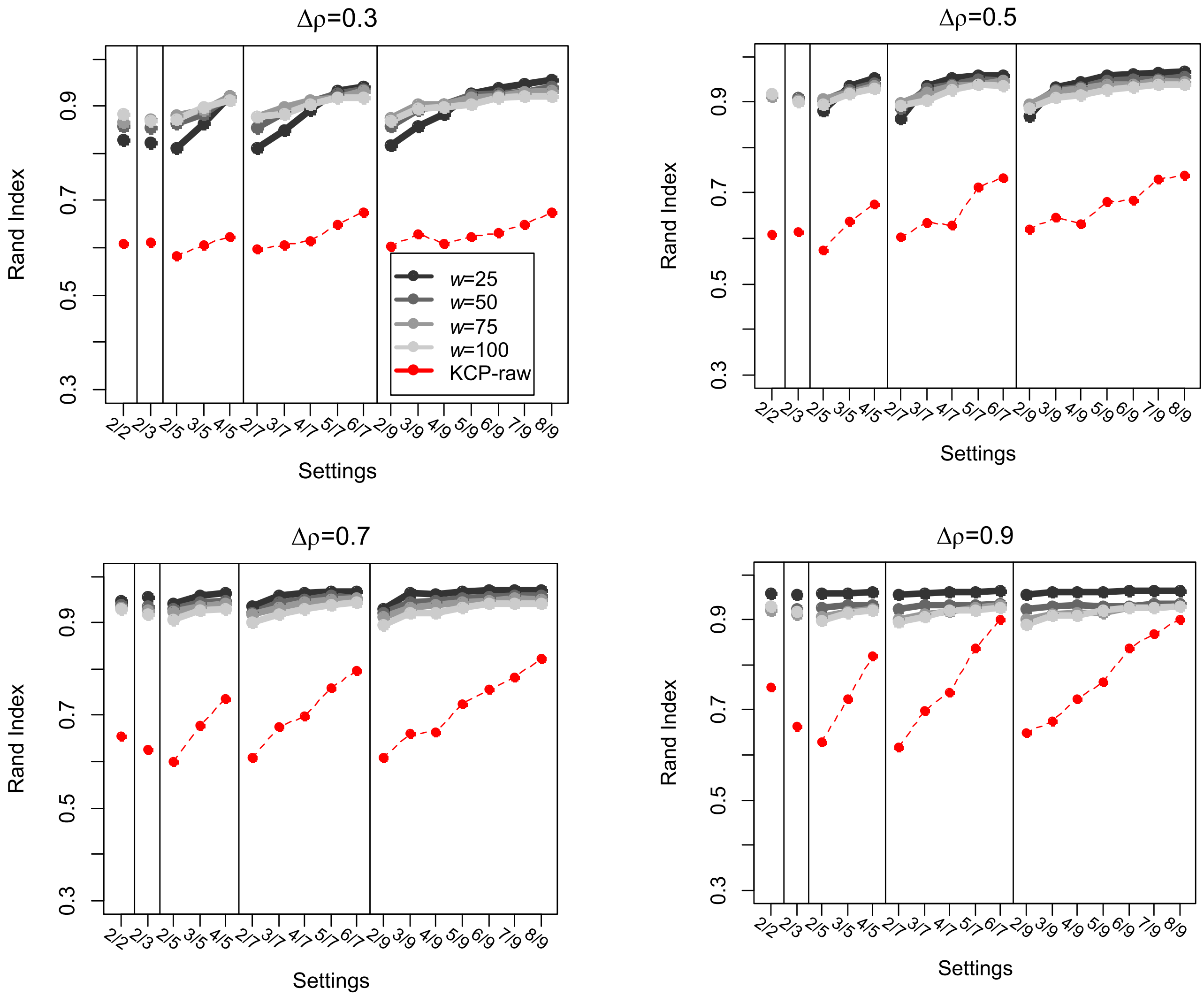

Fig. A2. Mean Rand indices (RI) for settings with correlation change in simulation study 1 (number of change points is fixed to 2): The four panels correspond to the level of correlation change in the simulated data: 0.3, 0.5, 0.7 and 0.9. In every panel, mean RI's obtained using KCP-corr (employing varying window sizes: 25, 50, 75 and 100; in gray) and KCP-raw (in red) were plotted across the simulation settings on the *x*-axis, where each setting is written as (no. of correlating variables)/(total no. of variables). (For the interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

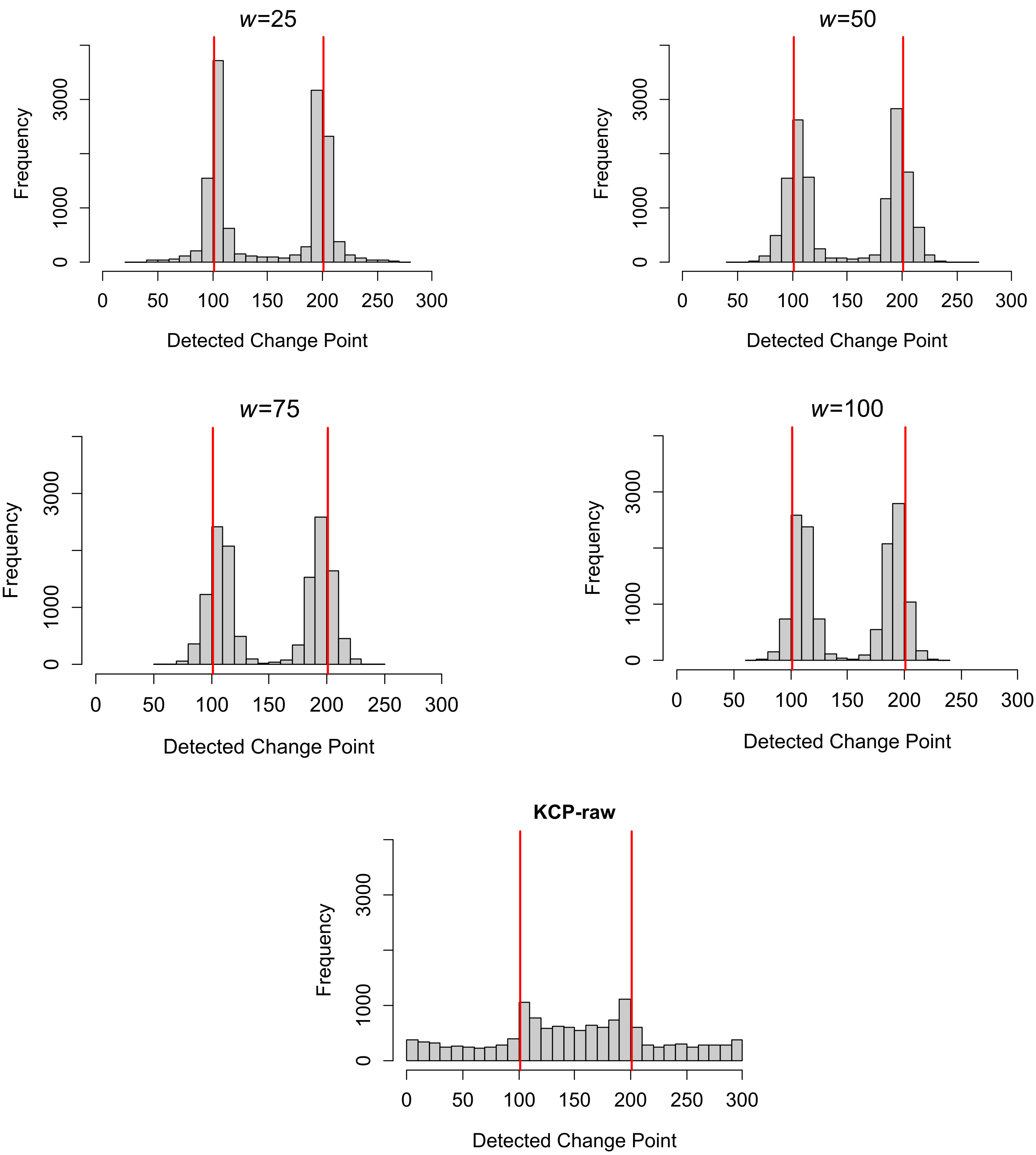

Fig. A3. Histogram of KCP-corr (employing different window sizes) and KCP-raw change points in simulation study 1 (number of change points is fixed to 2). The upper four panels show the frequencies of detection using 4 window sizes: 25, 50, 75 and 100 time points for KCP-corr, while the lowermost panel exhibit those of KCP-raw.

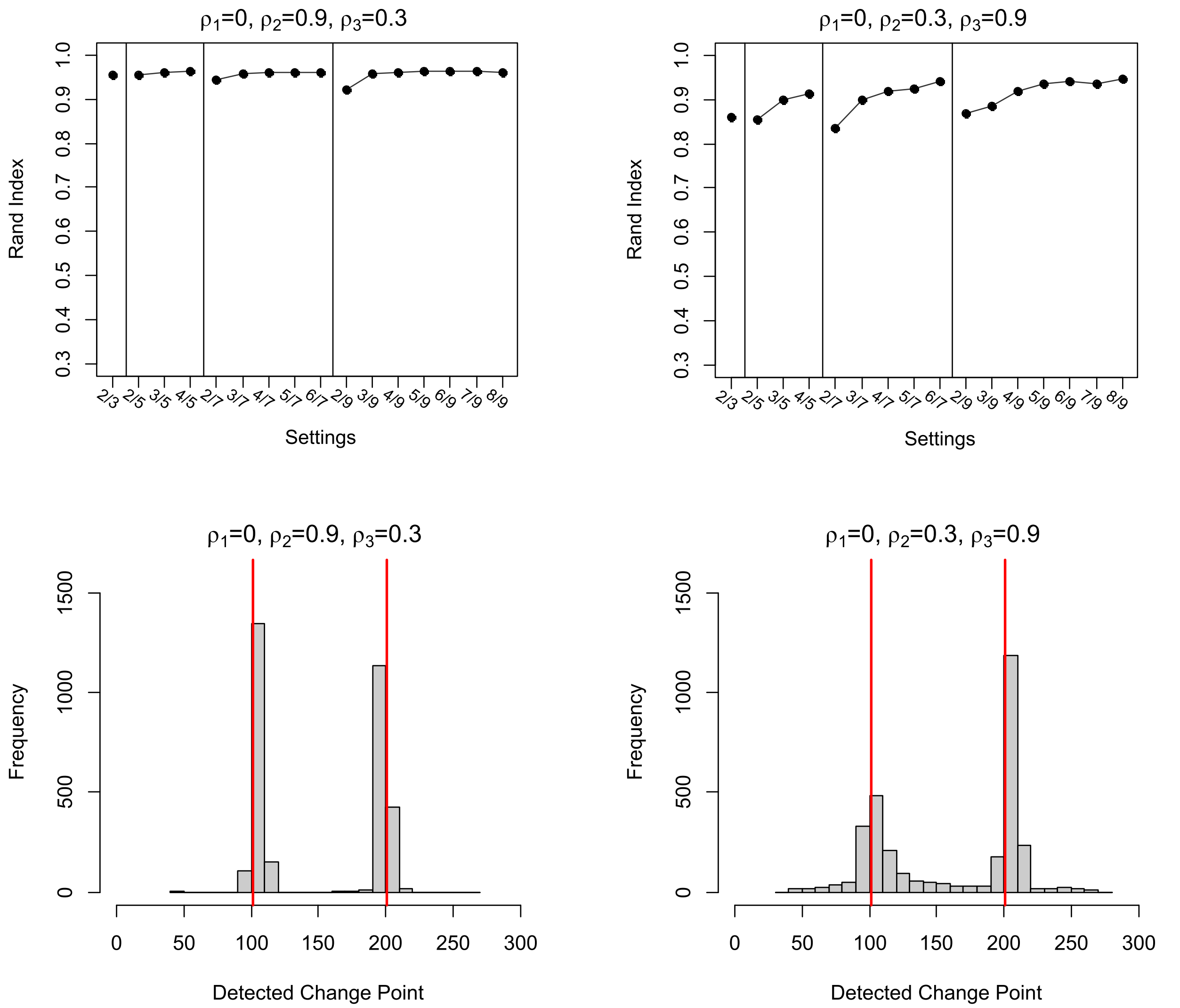

Fig. A4. Mean Rand indices (RI) and detected change points for settings with unequal effect sizes (number of change points is fixed to 2). In the upper panels, mean RI's obtained using KCP-corr (*ws*= 25) were plotted across the simulation settings on the *x*-axis, where each setting is written as (no. of correlating variables)/(total no. of variables). In the lower panels, the frequency distribution of the detected change points are displayed.

References

[1] E. Amsel, K. Renninger, Change and Development: Issues of Theory, Method and Application, Lawrence Erlbaum Associates, 1997.
[2] Arlot, S., Celisse, A., & Harchaoui, Z. (2012). Kernel change-point detection. Retrieved from https://arxiv.org/pdf/1202.3878v1.pdf.
[3] Arlot, S., Celisse, A., & Harchaoui, Z. (2016). A kernel multiple change-point algorithm via model selection. Retrieved from https://arxiv.org/pdf/1202.3878.pdf.
[4] I. Barnett, J. Onnela, Change point detection in correlation networks, Sci. Rep. 6 (2016) 18893, doi:10.1038/srep18893.
[5] M. Basseville, I. Nikiforov, Detection of Abrupt Changes: Theory and Application, Prentice-Hall, New Jersey, 1993.
[6] J. Blechert, P. Peyk, M. Liedlgruber, F. Wilhelm, ANSLAB: Integrated multichannel peripheral biosignal processing in psychophysiological science, Behav. Res. Methods 48 (4) (2016) 1528–1545, doi:10.3758/s13428-015-0665-1.
[7] B. Brodsky, B. Darkhovsky, Non Parametric Methods in Change-Point Problems, Springer, Netherlands, 1993.
[8] K. Bulteel, E. Ceulemans, R. Thompson, C. Waugh, I. Gotlib, F. Tuerlinckx, P. Kuppens, DeCon: A tool to detect emotional concordance in multivariate time series data of emotional responding, Biol. Psychol. 98 (1) (2014) 29–42, doi:10.1016/j.biopsycho.2013.10.011.
[9] J. Cabrieto, F. Tuerlinckx, P. Kuppens, B. Hunyadi, E. Ceulemans, Testing for the presence of correlation changes in a multivariate time series: A permutation based approach, Sci. Rep. 8 (2018) 769, doi:10.1038/s41598-017-19067-2.
[10] J. Cabrieto, F. Tuerlinckx, P. Kuppens, M. Grassmann, E. Ceulemans, Detecting correlation changes in multivariate time series: a comparison of four non-parametric change point detection methods, Behav. Res. Methods 49 (3) (2016) 988–1005, doi:10.3758/s13428-016-0754-9.
[11] J. Cacioppo, R. Petty, M. Losch, H. Kim, Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions, J. Personal. Soc. Psychol. 50 (2) (1986) 260–268, doi:10.1037/0022-3514.50.2.260.
[12] J. Chen, A. Gupta, Parametric Statistical Change Point Analysis With Applications to Genetics, Medicine and Finance, Second ed., Springer, New York, 2012.
[13] P. Galeano, D. Wied, Dating multiple change points in the correlation matrix, TEST 26 (2) (2016) 331–352, doi:10.2139/ssrn.2431468.
[14] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.

[15] J. Gross, The future's so bright, I gotta wear shades, Emot. Rev. 2 (3) (2010) 212–216, doi:10.1177/1754073910361982.
[16] P. Jiruska, M. de Curtis, J. Jefferys, C. Schevon, S. Schiff, K. Schindler, Synchronization and desynchronization in epilepsy: controversies and hypotheses, J. Physiol. 591 (Pt 5) (2012) 787–797, doi:10.1113/jphysiol.2012.239590.
[17] R. Killick, I. Eckley, Changepoint: An R package for changepoint analysis, J. Stat. Softw. 58 (3) (2014) 1–19, doi:10.18637/jss.v058.i03.
[18] H. Klausmeier, T. Sipple, Factor structure of the Piagetian stage of concrete operations, Contemp. Educ. Psychol. 7 (1982) 161–180, doi:10.1016/0361-476X(82)90041-8.
[19] P. Koval, P. Kuppens, Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia, Emotion 12 (2012) 256–267, doi:10.1037/a0024756.
[20] J. Larsen, C. Norris, J. Cacciopo, Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii, Psychophysiology 40 (2003), doi:10.1111/1469-8986.00078.
[21] M. Lavielle, Using penalized contrasts for the change-point problem, Signal Process. 85 (4) (2005) 1501–1510.
[22] Lung-Yut-Fong, A., Lévy-Leduc, C., & Cappé, O. (2012). Homogeneity and change-point detection tests for multivariate data using rank statistics. Retrieved from http://arxiv.org/abs/1107.1971.
[23] D. Matteson, N. James, A nonparametric approach for multiple change point analysis of multivariate data, J. Am. Stat. Assoc. 109 (505) (2014) 334–345, doi:10.1080/01621459.2013.849605.
[24] I. Mauss, M. Robinson, Measures of emotion: a review, Cognit. Emot. 23 (2) (2009) 209–237, doi:10.1080/02699930802204677.
[25] I. Mauss, R. Levenson, L. McCarter, F. Wilhelm, J. Gross, The tie that binds? Coherence among emotion experience, behavior, and physiology, Emotion 5 (2005) 175–190, doi:10.1037/1528-3542.5.2.175.
[26] P. Moreno, P. Ho, N. Vasconcelos, A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications, Advances in Neural Information Processing Systems, 16, MIT Press, 2004.
[27] M. Müller, G. Baier, A. Galka, U. Stephani, H. Muhle, Detection and characterization of changes of the correlation structure in multivariate time series, Phys. Rev. E 71 (2005) 046116, doi:10.1103/PhysRevE.71.046116.
[28] E. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115, doi:10.1093/biomet/41.1-2.100.
[29] M.L. Pe, P. Kuppens, The dynamic interplay between emotions in daily life: augmentation, blunting, and the role of appraisal overlap, Emotion 12 (2012) 1320–1328, doi:10.1037/a0028262.
[30] M. Pollak, Optimal detection of a change in distribution, Annal. Stat. 12 (1985) 206–227, doi:10.1214/aos/1176346587.
[31] W. Rand, Objective criteria for the evaluation of clustering methods, J. Am. Stat. Assoc. 66 (336) (1971) 846–850, doi:10.2307/2284239.
[32] D. Rosenfield, E. Zhou, F. Wilhelm, A. Conrad, W. Roth, A. Meuret, Change point analysis for longitudinal physiological data: detection of cardio-respiratory changes preceding panic attacks, Biol. Psychol. 84 (1) (2010) 112–120.
[33] A. Shiryaev, On optimum methods in quickest detection problems, Theory Probab. Appl. 8 (1) (1963) 22–46, doi:10.1137/1108002.
[34] B. Sriperumbudur, A. Gretton, K. Fukumizu, G. Lanckriet, B. Scholkopf, Hilbert Space embeddings and metrics on probability measures, J. Mach. Learn. Res. 11 (2010) 1517–1561.
[35] G. Szekely, M. Rizzo, N. Bakirov, Measuring and testing dependence by correlation of distances, Annal. Stat. 35 (6) (2007) 2769–2794, doi:10.1214/009053607000000505.
[36] A. Tartakovsky, B. Rozovskii, R. Blazek, H. Kim, A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-pointdetection methods, IEEE Trans. Signal Process. 54 (9) (2006) 3372–3382, doi:10.1109/TSP.2006.879308.
[37] H. Van der Maas, P.C. Molenaar, Stagewise cognitive development: an application of catastrophe theory, Psychol. Rev. 99 (3) (1992) 395–417, doi:10.1037/0033-295X.99.3.395.
[38] A. Vergult, W. De Clercq, A. Palmini, B. Vanrumste, P. Dupont, S. Van Huffel, W. Van Paesschen, Improving the interpretation of ictal scalp EEG: BSS-CCA algorithm for muscle artifact removal, Epilepsia 48 (5) (2007) 950–958, doi:10.1111/j.1528-1167.2007.01031.x.
[39] M. Wichers, P. Groot, Critical slowing down as a personalized early warning signal for depression, Psychother. Psychosom. 85 (2016) 114–116, doi:10.1159/000441458.
[40] D. Wied, A nonparametric test for a constant correlation, Econometr. Rev. 36 (10) (2017) 1157–1172, doi:10.1080/07474938.2014.998152.
[41] F.H. Wilhelm, J. Rattel, M. Wegerer, M. Liedlgruber, S. Schweighofer, S.D. Kreibig, V. Kolodyazhniy, J. Blechert, Attend or Defend? Sex Differences in Behavioral, Autonomic, and Respiratory Response Patterns to Emotion-Eliciting Films, Biol. Psychol. 130 (2017) 30–40, doi:10.1016/j.biopsycho.2017.10.006.