Supplementary Material for

An objective, comprehensive and flexible statistical framework for detecting early warning signs of mental health problems

Jedelyn Cabrieto[1]*†, Janne Adolf[1]†, Francis Tuerlinckx[1], Peter Kuppens[1], Eva Ceulemans[1]

[1] Research Group of Quantitative Psychology and Individual Differences, KU Leuven - University of Leuven, Leuven, Belgium

* Corresponding author

† These authors contributed equally to this work

I. KCP on the running statistics

Step 1. Searching for change points.

The similarity measure

After obtaining the running statistic of interest (e.g., running mean, running variance, running autocorrelation or running correlation) by sliding a time window across the time series and computing the statistic in each window, KCP looks for change points in this running statistic, by examining the similarities between windows. The key idea is straightforward: When there is no change point, there will only be small fluctuations in the similarities. However, when statistics in subsequent windows are very similar, while those that are distant in time are dissimilar, there is more evidence for (a) possible change point(s). KCP employs a kernel function to quantify these similarities. Specifically, we use the Gaussian kernel, such that if $\boldsymbol{RS}_i$ and $\boldsymbol{RS}_j$ denote the running statistic computed in windows *i* and *j*, respectively, we compute the following similarity,

$$Gk\left(\boldsymbol{RS}_i, \boldsymbol{RS}_j\right) = \exp\left(\frac{-\| \boldsymbol{RS_i} - \boldsymbol{RS}_j \|^2}{2h_{RS}^2}\right), \tag{1}$$

where h_{RS} is the median Euclidean distance between all the $\boldsymbol{RS}_i$ values. The similarity value ranges from 0 to 1, with 0 implying extreme dissimilarity and 1 indicating complete similarity.

Since the goal is to analyze all variables in the time series simultaneously (cfr. comprehensive analysis), our approach is multivariate such that $\boldsymbol{RS}_i$ is a vector of the running statistic for all the *V* variables in the system. Specifically, for univariate statistics such as the mean, variance and autocorrelation, the corresponding $\boldsymbol{RS}_i$ will consist of *V* elements as each variable yields one running statistic. For correlations however, the corresponding $\boldsymbol{RS}_i$ will have more elements since there will be $\frac{V(V-1)}{2}$ pairwise correlations to monitor. Regarding the window size in extracting the running statistic, $\boldsymbol{RS_i}$, we recommend to use *w*=25 as our previous studies [1-2] showed that in comparison to larger window sizes, this leads to more power in detecting a change, as well as less biased change point estimates. We emphasize, though, that prior information can help in making this choice. For instance, one can set the window size equal to the length of the shortest expected event. If no prior information is available as is often the case, employing different window sizes to check the stability of the obtained change points can be helpful.

The variance criterion to optimize change point locations

To locate the change points, KCP optimizes a variance criterion based on the intra-phase scatter,

$$\hat{V}_{\tau_1,\tau_2,\tau_3\dots,\tau_K,m} = (\tau_m - \tau_{m-1}) - \frac{1}{\tau_m - \tau_{m-1}} \sum_{i=\tau_{m-1}+1}^{\tau_m} \sum_{j=\tau_{m-1}+1}^{\tau_m} Gk(\boldsymbol{RS}_i, \boldsymbol{RS}_j), \qquad (2)$$

where the indices, $\tau_1, \tau_2, \dots, \tau_K$, are the phase boundaries, *m* is the current phase number, and τ_{m-1} and τ_m are the first and last time points of this phase. The more similar the values of the running statistic are inside a phase, the lower is the intra-phase scatter, as $\hat{V}_m$'s rightmost term will become more negative due to the high similarities. If the user requests to locate *K* change points, the variance criterion,

$$\hat{R}(\tau_1, \tau_2, \dots, \tau_K) = \frac{1}{w} \sum_{m=1}^{K+1} \hat{V}_{\tau_1,\tau_2,\tau_3\dots,\tau_K,m} . \qquad (3)$$

which is simply the sum of all *K*+1 intra-phase scatters divided by the total number of windows, is minimized to ensure optimal homogeneity (i.e., high similarity) of the running statistic within a phase.

Let's consider the simplest case where only one change point (*K*=1) has to be estimated. Two intra-phase scatters will be optimized: $\hat{V}_{\tau_1,1}$ and $\hat{V}_{\tau_1,2}$. The phase containing the first observations until τ_1 will determine $\hat{V}_{\tau_1,1}$, and the phase comprising all remaining observations will determine $\hat{V}_{\tau_1,2}$. The goal therefore is to search among all $\tau_1 \in \{2, 3, \dots, w-1\}$ the location $\hat{\tau}_1$ that yields the lowest value of the variance criterion, $\hat{R}(\tau_1) = \frac{1}{w}\left(\hat{V}_{\tau_1,1} + \hat{V}_{\tau_1,2}\right)$, and to optimally segment the time series into phases within which the values of the running statistic are very similar. The change point is then set at $\hat{\tau}_1 + 1$. Generalizing the procedure, the optimal *K* change points, $\hat{\tau}_1 + 1$, $\hat{\tau}_2 + 1$, …$\hat{\tau}_K + 1$, are obtained by finding

$$\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_K = \arg\min \hat{R}(\tau_1, \tau_2, \dots, \tau_K). \tag{4}$$

For ease of notation, in the following steps, we will denote the optimized variance criterion for a given *K* as $\hat{R}_{min,K}$.

Step 2. Selecting the number of change points.

In practice, users often do not know the expected number of change points and should infer this from the data itself. We therefore employ a two-step procedure to choose the optimal number of change points.

Step 2.1 Testing if there is at least one significant change point.

First, a significance test is performed to decide on the presence of at least one change point in the running statistic. The test compares the variance criterion $\hat{R}_{min,K}$ of the running statistic computed from the original data to that of the running statistic from the permuted data (in which time points are reshuffled). Two subtests are implemented: the variance and the variance drop test.

Variance test. The idea of the variance test is that when the running statistic contains at least one change point, their overall variance will be large. This is not expected if the running statistic is obtained from permuted data, as possible changes across time are effectively removed via reshuffling. The variance test therefore compares the overall variance (i.e., $\hat{R}_{min,K=0}$, the variance criterion value when no change point is extracted from the running statistic) to the distribution of $\hat{R}_{min,K=0,perm}$, which is obtained by computing the same overall variance from a large number of permuted versions of the data. The p-value of the variance test is therefore given by,

$$p_{variance\ test} = \frac{\#(\hat{R}_{min,K=0,perm} > \hat{R}_{min,K=0})}{B} \tag{5}$$

where *B* denotes the number of permutations carried out. In this paper, *B* was always set to 1000.

We note that this test is not applied to the running autocorrelation as it leads to an inflated Type 1 error rate.

*Variance drop test***.** The variance drop test looks at the extent to which the variance criterion improves due to extracting change points. Recall that the variance criterion, $\hat{R}_{min,,K}$, improves (i.e., decreases) when more change points are allowed for. However, if there are no true change points present in the data, this improvement will not be substantial and will be comparable to that of permuted data. Thus, in the variance drop test, we compare the drop in the variance criterion, $\hat{R}_{min,K} - \hat{R}_{min,K-1}$ to a distribution of variance drops obtained after reshuffling the data. We only use the strongest evidence for change, by retaining the maximum $\hat{R}_{min,K} - \hat{R}_{min,K-1}$ among all considered *K*'s. The variance drop test therefore yields a p-value equal to

$$p_{variance\ drop\ test} = \frac{\#(\max variance\ drop_{perm} > \max variance\ drop)}{B} \quad (6)$$

where, *B*, again indicates the number of reshuffled data sets on which the permuted distribution is based.

*Combining the sub-tests***.** Since two sub-tests are carried out, we correct for multiple testing by adopting a corrected significance level of $\frac{\alpha}{2}$ for each test. If at least one of the sub-tests yield a significant result, we declare that there is at least one change point in the running statistics. For the running autocorrelation, we only employ the variance drop test, and thus, in this case, its significance level is set to α.

Step 2.2 If there is at least one change point (K>0), choose the optimal number of K via penalization.

If the permutation test above yields a significant result, we employ the penalty term $pen_K = C\frac{V_{max}(K+1)}{\mathrm{w}}\left[1 + \log\left(\frac{\mathrm{w}}{K+1}\right)\right]$ proposed by Arlot et al. [3], to aid in choosing a *K* that balances fit, as evidenced by a reduction in the variance criterion, $\hat{R}_{min,K}$, and complexity, by penalizing too large *K*-values that would lead to unnecessary change points. The idea is to pick the number of change points that minimizes the sum of the variance criterion and the penalty term:

$$\hat{K} = \arg\min \hat{R}_{min,K} + pen_K. \quad (7)$$

The penalty coefficient *C* can be tuned by the user to adjust the impact of the penalty term. Small *C*-values imply a weak penalty for extra change points in the solution and will therefore tend to favor higher *K*-values (many change points). Large *C*-values, on the other hand, more strongly penalize adding more change points and therefore would favor small *K*-values (less change points). The remaining term, V_{max}, is obtained by computing the trace of the empirical covariance matrix of the first and last 5% elements of the running statistic, retaining whichever is larger among the two.

Since the choice of the penalty coefficient, C, highly influences the result of the penalization scheme, Cabrieto et al.[4] proposed to tune this coefficient through a grid search so that the choice will not solely depend on a single choice for C. The procedure starts by first setting C equal to 1, which will always yield $Kmax$ (the largest K) as a solution because the penalty is not strong. Moving on, we increase C by incrementing it linearly. This step increases the influence of the penalty term, and thus lower K-values are eventually favored. The grid search terminates once C becomes too large such that the K-value returned is zero. The whole procedure, at the end, outputs the most stable K, that is, the most frequently returned K-value across all considered C-values.

The final KCP change point solution then consists of the change point locations that correspond to the most stable K.

II. Multiple testing correction for monitoring several statistics

When multiple statistics are tracked, as is done in this paper, a multiple testing correction can be used in Step 2.1. Since we monitored three running statistics (i.e., running variance, running correlation, and running autocorrelation) in three separate KCP-analyses, we maintained the overall type 1 error rate of .05 by allotting a corrected α-level of .017 (i.e., .05 divided by 3) to each analysis and thus each running statistic. This implies that for the running autocorrelation, where only the variance drop test is employed, the corrected α-level is .017. For the running variance and running correlation where two sub-tests are employed, the α-level for each sub-test is corrected such that the per statistic α-level (.017) is maintained. This means that each subtest will have an alpha level of .008 (i.e., per statistic α-level divided by 2). In Table 1, we present this multiple testing plan and the corresponding KCP permutation test results for the depression data.

We remark that using such a Bonferroni correction is a conservative way of controlling for multiple testing. Less conservative methods have been proposed, for instance a non-parametric combination of dependent permutation tests [5]. However, to the best of our knowledge, this combined test is an Omnibus test, postulating the global null hypothesis that none of the running statistics contains a change point and the corresponding alternative hypothesis that at least one running statistic shows at least one change point. In contrast, we are interested in local effects, in that we aim to test for each running statistic separately whether it contains at least one change point. Nevertheless, exploring the potential of a non-parametric combined test for increasing the power of KCP-analyses, might be a useful direction for future research.

Table 1. Corrected α-levels and observed p-values for the KCP permutation tests employed to determine whether the 3 running statistics (i.e., running autocorrelation, running variance and running correlation) contain at least one change point.

Data	Running Statistic	Tests	Corrected α-level	p-value
Full	Running autocorrelation*	Variance	.017	.01
	Running variance*	Variance	.008	.016
		Variance drop	.008	0
	Running correlation*	Variance	.008	0
		Variance drop	.008	0
Before Relapse	Running autocorrelation	Variance	.017	.221
	Running variance*	Variance	.008	0
		Variance drop	.008	0
	Running correlation*	Variance	.008	0
		Variance drop	.008	.148

* Contains at least one significant change point

References

1. Cabrieto J, Tuerlinckx F, Kuppens P, Borbála H, Ceulemans E. Testing for the presence of correlation changes in a multivariate time series: A permutation based approach. Sci Rep 2018 Jan;8(1):769.
2. Cabrieto J, Tuerlinckx F, Kuppens P, Wilhelm FH, Liedlgruber M, Ceulemans E. Capturing correlation changes by applying kernel change point detection on the running correlations. Information Sciences 2018 Jun;447:117-139.
3. Arlot S, Celisse A, Harchaoui Z. Kernel change-point detection. https://arxiv.org/pdf/1202.3878.pdf, 2012.
4. Cabrieto J, Tuerlinckx F, Kuppens P, Grassmann M, Ceulemans E. Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. Behav Res Methods 2017 Jun;49(2):988-1005.
5. Pesarin F, Salmaso S. Permutation Tests for Complex Data: Theory, Applications and Software. Chichester, UK: Wiley & Sons; 2010. 450 p.