RESEARCH ARTICLE

On optimal segmentation and parameter tuning for multiple change-point detection and inference

Christina Parpoula\textsuperscript{a} and Alex Karagrigoriou\textsuperscript{b}

\textsuperscript{a}Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece; \textsuperscript{b}Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovasi, Samos, Greece

ABSTRACT
Change-point analysis is the task of finding abrupt (and significant) changes in the underlying model of a signal or time series. Change-point detection methods typically involve specifying the maximum number of segments to search for and the minimum segment length. However, there is no objective way to pre-specify these two parameters, and it mostly depends upon the particular application. Within this framework, a recursive optimization algorithm is developed that is capable of exploring and fine tuning these two input parameters, and optimally segmenting a time series. This multiple change-point detection technique therefore addresses a wide class of real-life contexts and problems where the identification of optimal level shifts in a time series is the main goal. Extensive simulation results are presented and a real-life example is given to illustrate the implementation of the developed scheme in practice and to unfold its capabilities. Concluding remarks and suggestions for future research are also provided.

1. Introduction
Change-point analysis has played an important role in modern time series study. One type of change one may seek for, is a change-point that is a moment in time where there is an abrupt change in one or more of the properties/distributional characteristics of the time series such as the mean, the median, or the standard deviation (see, for example, Pettitt [1], Aue and Horváth [2]). Change-points can also occur in variances/volatilities (see, for example, Inclan and Tiao [3], Gao et al. [4]), in the series’ correlation structures (see, for example, Davis et al. [5]), or even in the marginal distribution of the series (see, for example, Gallagher et al. [6]). Detection of change-points helps modeling and prediction of time series and is found in applications of many scientific fields. Change-point detection was studied first in product quality in manufacturing operation (see, for example, Dudding and Jennett [7]), however, over the last two decades, it has begun to spread to nonmanufacturing industries and found many applications in areas such as healthcare and...
medical condition monitoring, infectious disease surveillance, environment monitoring, meteorology, climatology, hydrology, speech recognition, image analysis, human activity analysis, bioinformatics, neuroscience, signal processing, social sciences and econometrics, among others.

Change-point analysis was initially restricted to univariate time series of independent variables under the assumption of ‘At-Most-One-Change’ (AMOC). Two classic AMOC techniques are cumulative sum (CUSUM) statistics and likelihood ratio (LR) tests (see, for example, Jandhyala et al. [8], Wang et al. [9], Shi [10]). Academic change-point research commenced with the single change-point case for independent and identically distributed (i.i.d.) data (see, for example, Page [11]). Worth citing is the work of MacNeill [12], where asymptotic theory for a single change-point is developed. The change-point subject is now vast, with thousands of papers devoted to the topic. Change-point analysis was extended to series with multiple change-points and to multivariate time series (see, for example, Lebarbier [13], Lavielle and Teyssiere [14], Harchaoui et al. [15], Killick et al. [16], Matteson and James [17], Fryzlewicz et al. [18], Niu et al. [19], Haynes et al. [20], Baranowski et al. [21], Li et al. [22], among others). New methods were developed to take into account the dependence structure or trend and periodicity within a series, or high-dimensional multivariate time series (see, for example, Ray and Tsay [23], Berkes et al. [24], Lu and Lund [25], Lund et al. [26], Gombay [27], Shao and Zhang [28], Robbins et al. [29], Xie et al. [30], Shen et al. [31], Shao [32], Cho and Fryzlewicz [33], Robbins et al. [34], Chakar et al. [35], Zhang and Lavitas [36], among others).

It is also worthwhile to note that as in other areas of statistics, there are parametric and nonparametric methods for change-point detection. A number of classical parametric methods have been developed; however, in practice, there is often not enough information on the type of distribution to make an informed choice for the distribution family and subsequently perform a parametric change-point detection analysis (see, for example, Gurevich and Vexler [37]). Recently the change-point model framework has been developed for change-point detection in situations where the available information about the distributions is very limited. Note that previous studies have analysed Pettitt’s nonparametric method in terms of its ability to detect the correct time of change for different distributions (see, for example, Xie et al. [38]). Ross [39], among others, studied several implementations of the change-point model framework (incorporating different test statistics, using both parametric and nonparametric techniques) for detecting changes in the distribution of discrete-time sequences of univariate random variables. In particular, statistics for detecting changes in either a batch of data containing at most one change-point or a sequential context with data which may contain multiple change-points were implemented. Moreover, note here that most existing change-point methods are relevant to the Phase I Statistical Process Control problems, since the offline change-point detection problem is essentially similar to the Phase I control problem in that in both cases, the sample size is fixed, and one wishes to detect any changes in the underlying process distribution.

In this section, our goal is not to present the vast fast-growing change-point detection literature in an exhaustive fashion, but rather to present the main ideas that shaped the field, and to relate them with the developed methodology – focussing on individual time series data – since recent advances have led to more and more instances where individual measurements are collected over time. Toward this end, this paper addresses the problem of ‘a posteriori’ optimal multiple change-point detection and inference, performing a global
segmentation. Change-point detection methods typically involve specifying the maximum number of segments to search for and the minimum segment length; however, there is no objective way to pre-specify these two parameters. In this context, a recursive optimization algorithm is developed for exploring and fine tuning these two input parameters, and optimally segmenting a time series. Within this framework, an empirical reasonable compromise is provided for pre-specifying the aforementioned input parameters. This issue is of great importance since in practice, up to now, pre-specifying these two parameters (initializing step of a typical change-point detection algorithm) is ‘somewhat arbitrary’, lacking statistical justification, and it mostly depends upon the particular application. In this sense, this paper, through optimal exploration and fine tuning of the input parameters is filling a gap in the relevant literature, whereas the developed scheme also provides optimal time series segmentation. The empirical and simulation-based findings indicate that the proposed scheme is a viable alternative to widely used multiple change-point detection algorithms already existing in the literature. The proposed scheme therefore addresses a wide class of real-life contexts and problems where the identification of optimal level shifts in a time series is the main goal.

The rest of the paper is organized as follows. In Section 2, the statistical framework is introduced and technical details are provided. In Section 3, an extensive simulation study is performed for specifying the tuning parameters (i.e. maximum number of segments and minimum segment length). Two empirical rules are also provided. In Section 4, a comparative study is performed for benchmarking purposes. A simulation-based performance comparison between the proposed scheme and widely used multiple change-point detection algorithms is conducted. Further, an application to real epidemiological data is presented. Within the latter context, the developed technique is compared to its competitors, as regards its ability to detect successfully the signalled start and end weeks of epidemics. Finally, in Section 5, some concluding remarks and suggestions for future research are given.

2. Research methodology

The a posteriori (or off-line) change-point problem arises when the series of observations is complete at the time to process it [40]. The interested reader may refer to Tartakovsky et al. [41] for related work on online change-point detection. We only consider in this paper the a posteriori problem; hence, our goal consists in recovering the configuration of change-points using the whole observed series. In this paper, we deal with piecewise stationarity, arguably the simplest type of deviation from stationarity. This implies a time-varying process where its parameters evolve through time but remain constant for a specific period of time. The complex problem of detecting multiple change-points in a time series is described explicitly below.

2.1. The Segmentation model under a Gaussian framework

In a segmentation framework, two types of changes can be considered: changes in the mean and the variance of the signal, or changes in the mean only. Let us define a model where the mean is segment-specific and the variance is common between segments.
Suppose \( x_t \) are realizations of independent random variables \( \{X_t\}_{t=1}^{m} \), with Gaussian distributions \( N(\mu_t, \sigma^2) \). Let us assume that this stream of process data fluctuates around some underlying signal. Suppose now that the underlying signal is piecewise constant (that is, properties of the underlying phenomenon do not change within each segment), and there exist \( K-1 \) changes (discontinuity instants) that affect the mean parameter of the distribution of the \( X_s \), at unknown coordinates \( (t_0, t_1, t_2, \ldots, t_{K-1}, t_K) \) with convention \( t_0 = 1 \) and \( t_K = m \), and that the parameters of the \( X_s \) distributions are constant between two changes

\[
\forall t \in [t_{k-1}, t_k], \quad X_t = \mu_k + e_t, \quad e_t \sim N(0, \sigma^2),
\]

where \( \mu_k \) is the mean of the \( k^{th} \) segment. The change-points \( (t_1, t_2, \ldots, t_{K-1}) \) define a partition of the individual observations into \( K \) segments denoted as

\[
I_k = \{t, t \in [t_{K-1}, t_K]\}.
\]

Due to the independence assumption, the log-likelihood can be decomposed into a sum of 'local' likelihoods, calculated on each segment as

\[
\mathcal{L}_k = \sum_{k=1}^{K} \ell_k
\]

with

\[
\ell_k = -\frac{1}{2} \sum_{t=t_{k-1}+1}^{t_k} \left[ \log(2\pi \sigma^2) + \left( \frac{x_t - \mu_k}{\sigma} \right)^2 \right].
\]

Note that because the data are supposed to be independent and Gaussian, the log-likelihood is a sum over all the segments (additive contrast) and maximum likelihood estimation is equivalent to least squares fitting, respectively.

Given the number of segments \( K \) and the segments’ coordinates \( (t_0, t_1, t_2, \ldots, t_{K-1}, t_K) \), the mean for each segment is estimated using maximum likelihood as

\[
\hat{\mu}_k = \frac{1}{t_k - t_{k-1}} \sum_{t=t_{k-1}+1}^{t_k} x_t,
\]

and since the variance of the segments is homogeneous, its estimator is given by

\[
\hat{\sigma}^2 = \frac{1}{m} \sum_{k=1}^{K} \sum_{t=t_{k-1}+1}^{t_k} (x_t - \hat{\mu}_k)^2.
\]

For a given sequence of change-points, the segment coordinates are known, and the estimation of the mean for each segment is straightforward. Then, the key problem is to estimate \( K \) and the positions of change-points. We will further proceed in two steps as follows:

Step 1: estimate the \( t_k \)s that is, find the best partition of a set of \( m \) individuals into \( K \) segments assuming that the number of segments is known,

Step 2: estimate the number of segments by adopting elbow curve method.

When the number of segments \( K \) is known, the problem is to find the best partition of \( \{1, \ldots, m\} \) into \( K \) segments, according to the likelihood, where \( m \) is the sample size.
However, in this context, an exhaustive search becomes unfeasible for large \( m \) and \( K \), since there are \( \binom{m-1}{K-1} \) possible choices for the positions of the change-points \( (t_1, t_2, \ldots, t_{K-1}) \).

An efficient dynamic programming approach to compute an exact solution for the multiple change-point problem was presented in Auger and Lawrence [42], with \( O(Qm^2) \) time complexity for a given upper bound \( Q \) on the number of change-points. Improvements to this algorithm is made by requiring the cost function to be additive, i.e.

\[
C(x_{a:b}) = C(x_{a:t-1}) + C(x_{t:b})
\]

where \( C(\cdot) \) is a cost function for a segment. Using this assumption, Jackson et al. [43] presented a dynamic programming algorithm with \( O(m^2) \) complexity. Killick et al. [16] subsequently extended this method with a pruning step on potential change-point locations, which reduces the time complexity to be approximately linear under certain conditions on the data generating process and the distribution of change-points. A different pruning approach was proposed by Rigaill [44] using further assumptions on the cost function, which was extended by Maidstone et al. [45] achieving a significant reduction in computation time \((m^* \log(m))\), and made robust against outliers by Fearnhead and Rigaill [46].

In this work, we also use a dynamic programming approach to reduce the computational load. Programs are coded in R statistical environment and are available in Supplementary File1. Under the assumption of additivity for the cost function, we adopt a recursive optimization (RO) scheme allowing us to solve the multiple change-point detection problem with a time complexity \( O(m^2) \). The recursions involved in the RO algorithm are described below.

### 2.1.1. Step 1

Let us assume that \( K \) is known, we then have to minimize

\[
J_k(1, m) = \sum_{k=1}^{K} \sum_{t \in I_k} (x_t - \hat{\mu}_k)^2.
\]

The quantity \( \sum_{t \in I_k} (x_t - \hat{\mu}_k)^2 \) can be viewed as the ‘cost’ of segment \( I_k \), in other words the cost of putting data \( x_{t_{k-1}+1} \) to \( x_{t_{k+1}} \) in a single segment. In these settings, \( J_k \) denotes the cost matrix and the global optimum is given by \( J_k(1, m) \). This optimization problem is actually a shortest path problem that can be solved thanks to dynamic programming based on Bellmann’s [47] optimality principle ‘any optimal path is composed of sub-paths which are themselves optimal’, following the least-squares approach for a change in the mean of a series of Lavielle and Moulines [48].

Let \( \hat{\mathcal{L}}_{k+1}(i,j) \) be the maximum log-likelihood obtained by the best partition of the data \( \{X(i), X(i+1), \ldots, X(j)\} \) into \( k+1 \) segments, with \( k \) change-points, and let denote \( \hat{J}_{k+1}(i,j) = -2\hat{\mathcal{L}}_{k+1}(i,j) \). The RO algorithm is as follows:

**Initialization**: \( k = 0, \forall 0 \leq i < j \leq m, \quad J_1(i,j) = \sum_{t=i+1}^{j} \left[ \log(2\pi\sigma^2) + \left( \frac{x_t - \mu_1}{\sigma} \right)^2 \right] \);

**Step \( k \)**: \( \forall k \in [1, K_{\text{max}}], \quad J_{k+1}(1,j) = \min_{i \leq h \leq j} \{J_k(1,h) + J_1(h+1,j)\}, \)

where \( K_{\text{max}} \) denotes a given maximum number of segments. Details for (pre)specifying \( K_{\text{max}} \) are provided later. Note that this problem of partitioning is analogous to the search
for the shortest path to do a trip from one point to another, where $\hat{J}_{k+1}(1,m)$ represents the total length of a $(k + 1)$-step path connecting the point with coordinate 1 to the point with coordinate $m$.

At this point, it is important to highlight some of the inherent statistical properties of the RO scheme as above described. Because of the normality assumption, the maximum likelihood estimator of the change-points coincides with the least-square estimator, and the cost function values coincide with the residual sum of squares (RSS). In this segmentation model developed under a Gaussian likelihood-based framework, we assumed that the sequence of residual errors are i.i.d Gaussian variables. Lavielle [40] pointed out that a Gaussian likelihood can be effectively used as a contrast function for detecting changes in the mean (and/or in the variance) for any sequence of random (not necessarily Gaussian) variables. Further, the adopted dynamic programming approach takes advantage of the additivity of the log-likelihood described above, considering that ‘a partition of the data into $k + 1$ segments is a union of a partition into $k$ segments and a set containing 1 segment’. This approach is advantageous for two main reasons; first, it yields an exact solution for the global optimum of the likelihood [42]; second, for a given $K$, the computational load is reduced from $O(m^K)$ to $O(m^2)$, since the algorithm only requires the storage of an upper $m \times m$ triangular matrix. Note that the recursions presented in the RO algorithm are more efficient than the Segment Neighbourhood method [42] as the location and number of change-points are decided in one pass of the data. This results in an $O(m^2)$ method compared to $O(Qm^2)$, if the number of change-points $Q$ were known. At the end of this procedure, the quantities $\hat{J}_1(1,m), \ldots, \hat{J}_{K_{\text{max}}}(1,m)$ are stored and will be used in the following step.

### 2.1.2. Step 2

Let us assume that the likelihood $\hat{L}_k$ measures the adjustment of the segmentation model with $K$ segments to the process data. We therefore aim at determining the optimal dimension $K$ given a stream of process data. A method for selecting the optimal $K$ consists in looking at the $\hat{L}_k$ value obtained with different number of segments. A classical and natural graphical method for selecting the dimension $K$ can be summarized as follows. First, we examine how the $\hat{L}_k$ value decreases when $K$ increases; then we select the dimension $K$ for which the $\hat{L}_k$ value ceases to decrease significantly. In other words, this heuristic approach (similar rationale with Elbow method/Elbow curve) looks for maximum curvature in the plot $(K, \hat{L}_k), \forall k \in [1,K_{\text{max}}]$. Thus, to look for where $\hat{L}_k$ ceases to decrease, means to look for and locate the break in the slope of this curve. Conclusively, the value of $K$ at which the slope abruptly changes provides an estimate $\hat{K}$ of the optimal number of segments, and we could then derive the $\hat{K} - 1$ change-points, i.e. $(\hat{t}_{\hat{K},1}, \ldots, \hat{t}_{\hat{K},\hat{K}-1})$.

Note here that selecting the number of segments only based on the likelihood criterion could probably lead to over-fitting [49]. Furthermore, the number of parameters to estimate is proportional to the number of segments, and a too large number of segments would lead to a large estimation error. Tuning input parameters, that is setting a reasonable upper limit on the size of the segmentation space subject to the constraint of a minimum segment length is here used in order to achieve a trade-off between a good adjustment and a reasonable number of parameters to estimate. However, despite this trading-off, one
still needs to implement elbow method and visually inspect the optimal number of segments. The interested reader may refer to Picard et al. [49] for an alternative method to automatically locate the break in the slope of the likelihood and estimate the number of segments via penalized likelihood. This procedure estimates a penalty constant and depends on a threshold which is chosen arbitrarily. Nevertheless, despite this thresholding, the procedure remains adaptive since the penalty constant is estimated according to the data, and it is also easy to implement. Further theoretical considerations regarding this adaptive method can be found in Lavielle [50].

### 3. Tuning parameters

Change-point detection methods typically involve setting an upper limit on the size of the segmentation space subject to the constraint that distance between change-points should be above a minimum length. Since there is no objective way to pre-specify these two parameters, one can fine tune them by ‘trial and error’ before arriving at the chosen one. Besides, the choice for the gap between two successive change-points mostly depends upon the particular application. The presented RO scheme also depends upon the pre-specification of a maximum number of segments to search for subject to the constraint of a minimum segment length. In this section, we summarize the results of an extensive simulation study conducted for specifying the minimum segment length and the maximum number of segments, as well as providing two practical rules as a reasonable compromise for both tuning parameters.

#### 3.1. Specification of the minimum segment length

The choice of the minimum segment length (denoted as $L_{\text{min}}$) should be chosen balancing the importance of detecting step (and even transient) shifts, with the cost of investigating a larger number of partitions. It is therefore necessary to constrain the minimum length between the successive change-points in front of a real data sequence. For this purpose a minimum length between two successive change-points is imposed: instead of minimizing over all possible partitions, only the partitions such that $t_{r} - t_{r-1} \geq L_{\text{min}}$, $r = 1, \ldots, K$ are considered. $L_{\text{min}}$ is a positive integer, and the (default) minimum segment length allowed by theory is a single observation for a change in mean, whereas for a change in variance (with unknown mean), the minimum segment length is two observations [16]. However, assuming that the minimum segment length is a single observation $t_{r} - t_{r-1} \geq 1$ (as in our case for detection of changes in the mean) is likely to yield an overfitting model with redundant change-points. In order to prevent overfitting, some sufficient minimum segment length can be chosen [51]. Thus, a simulation study was conducted here in order to choose a minimum segment length that seems satisfactory for practical prescriptions.

**Remark 3.1:** Note that it is possible, instead of imposing a minimum length between two successive change-points, for example, to constrain the estimates of the mean to lie within a compact set or imposing either stronger conditions on the process [48].

**Remark 3.2:** Lavielle [40] obtained some asymptotical results when the length of each segment tends to infinity (at the same rate as the total number of observations) which is
though beyond the scope of this paper. Indeed, in this paper, the choice of the minimum segment length value is justified by practical considerations.

Following similar simulation settings for the detection of changes in the mean as in Lavielle [50], the simulated series are sequences of $m$ independent Gaussian random variables of variance $\sigma^2 = 1$. The means in the five segments considered are $(0, \alpha, 0, 2\alpha, 0)$ respectively. We simulated 10,000 series with $\alpha = 0.5$ (representing the smallest jump size) and 10,000 series with $\alpha = 1$ (representing the largest jump size). For each of these series and several combinations of $L_{\text{min}}$ and $K$ values, we computed the values of the cost function. Based on these 10,000 simulations, we then computed the average cost function values for each combination of $L_{\text{min}}$ and $1 \leq K \leq K_{\text{max}}$. Results will be presented here only for $m = 50$ for space reasons. Several different sample sizes ($m = 100, 200, 500, 1000$) were examined and additional simulation results can be found in Supplementary File 2 ($m = 100, 200, 500, 1000$). Note here that in practical applications, we do not know in advance when the transition between a stable and unstable period occurs. Here, the stable period stands for the case in which there is no change and the mean remains constant, whereas the unstable period stands for the case in which the mean changes abruptly at some unknown instant(s). Thus, the positions of the change times were assumed to be random, and were chosen according to the three following patterns:

Pattern I: The unstable period is longer than the stable period overall (e.g. 30 observations that shift in position and 20 observations for which no change is assumed for $m = 50$, 60 observations that shift in position and 40 observations for which no change is assumed for $m = 100$, etc.);

Pattern II: The unstable period is as long as the stable period overall (e.g. 25 observations that shift in position and 25 observations for which no change is assumed for $m = 50$, 50 observations that shift in position and 50 observations for which no change is assumed for $m = 100$, etc.);

Pattern III: The stable period is longer than the unstable period (e.g. 20 observations that shift in position and 30 observations for which no change is assumed for $m = 50$, 40 observations that shift in position and 60 observations for which no change is assumed for $m = 100$, etc.).

For $m = 50$, the change-point locations considered are 10, 25, 30, 45 for Pattern I; 10, 20, 30, 45 for Pattern II and 10, 20, 30, 40 for Pattern III, respectively; for $m = 100$, the change-point locations considered are 20, 50, 60, 90 for Pattern I; 20, 40, 60, 90 for Pattern II and 20, 40, 60, 80 for Pattern III, respectively, and so on.

Tables 1–3 summarize the derived results for $m = 50$ for Patterns I-III, respectively. Tables 1–3 present the average cost function values obtained for several combinations of $L_{\text{min}}$ and $K$ values (both ranging from 1 to 10), for $\alpha = 0.5$ and $\alpha = 1$ (representing the smallest and largest jump size, respectively). In each table, the up (↑) and down (↓) arrows symbols in the column cells indicate the RSS values’ increase or decrease, respectively. Note here that ‘Inf’ corresponds to cases for which minimization of the RSS terminated due to proximity of estimation to a value at which the cost function is infinite. These correspond to cases for which the RO scheme could not run successfully (abnormal termination) due to the infinite cost function, since the number of observations required to run the algorithm
Table 1. RSS values for $L_{\min}$ vs. $K$ for Pattern I ($m = 50$).

| $L_{\min}$ vs. $K$ | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1                   | 57.75 | 57.75 | 57.75 | 57.75 | 57.75 | 57.75 | 57.75 | 57.75 | 57.75 | 57.75 |
| 2                   | 50.09 | 50.17 | 50.25 | 50.33 | 50.48 | 50.74 | 50.99 | 51.07 | 51.19 | 51.34 |
| 3                   | 43.39 | 43.51 | 43.81 | 44.12 | 44.53 | 45.42 | 46.21 | 46.88 | 47.55 | 48.29 |
| 4                   | 38.72 | 39.39 | 39.99 | 40.56 | 41.19 | 42.39 | 43.47 | 44.45 | 45.46 | 46.72 |
| 5                   | 34.59 | 35.62 | 36.55 | 37.43 | 38.38 | 39.98 | 41.47 | 42.93 | 44.82 | 48.22 |
| 6                   | 31.28 | 32.79 | 34.17 | 35.46 | 36.83 | 38.91 | 41.25 | Inf   | Inf   | Inf   |
| 7                   | 28.31 | 30.32 | 32.17 | 33.94 | 35.85 | 38.68 | 44.89 | Inf   | Inf   | Inf   |
| 8                   | 25.69 | 28.25 | 30.63 | 32.90 | 35.39 | 40.93 | Inf   | Inf   | Inf   | Inf   |
| 9                   | 23.33 | 26.45 | 29.38 | 32.20 | 35.82 | Inf   | Inf   | Inf   | Inf   | Inf   |
| 10                  | 21.19 | 24.91 | 28.42 | 31.86 | 40.09 | Inf   | Inf   | Inf   | Inf   | Inf   |

Table 2. RSS values for $L_{\min}$ vs. $K$ for Pattern I ($\alpha = 1.0$ ($m = 50$)).

| $L_{\min}$ vs. $K$ | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1                   | 83.62 | 83.62 | 83.62 | 83.62 | 83.62 | 83.62 | 83.62 | 83.62 | 83.62 | 83.62 |
| 2                   | 66.80 | 66.81 | 66.83 | 66.88 | 66.96 | 67.33 | 67.49 | 67.58 | 67.68 | 67.83 |
| 3                   | 51.51 | 51.55 | 51.61 | 51.73 | 52.02 | 55.04 | 57.25 | 58.88 | 60.30 | 61.86 |
| 4                   | 44.47 | 44.68 | 44.89 | 45.16 | 45.64 | 49.06 | 51.69 | 53.77 | 55.72 | 58.10 |
| 5                   | 38.12 | 38.56 | 39.01 | 39.51 | 40.28 | 44.34 | 47.64 | 50.37 | 53.22 | 57.61 |
| 6                   | 34.09 | 35.02 | 35.96 | 36.87 | 38.08 | 42.85 | 47.76 | 56.11 | Inf   | Inf   |
| 7                   | 30.54 | 32.03 | 33.52 | 34.99 | 36.88 | 42.46 | 54.19 | Inf   | Inf   | Inf   |
| 8                   | 27.56 | 29.65 | 31.73 | 33.77 | 36.31 | 45.38 | Inf   | Inf   | Inf   | Inf   |
| 9                   | 24.92 | 27.62 | 30.32 | 32.98 | 36.68 | Inf   | Inf   | Inf   | Inf   | Inf   |
| 10                  | 22.58 | 25.90 | 29.26 | 32.58 | 40.09 | Inf   | Inf   | Inf   | Inf   | Inf   |

in accordance with the specified $L_{\min}$ and $K$, exceeds the number of available observations (e.g. for $L_{\min} = 6$ and $K = 9$ at least 54 observations are required, thus for $m = 50$ the scheme returns the cost function value as infinite). Tables 1–3 show that the choice of $L_{\min} = 1$ leads to overfitting, as expected. This is also the case for $L_{\min} = 2, 3, 4$; in particular, the sequence $(J_K, 1 \leq K \leq K_{\max})$ keeps decreasing significantly, and a significant improvement of the fit continues to be obtained when we start adding change-points. An improvement of the fit (decrease in the RSS values) is generally obtained when we start adding change-points for any $L_{\min}$ value considered. Nevertheless, for $L_{\min} = 5$, these improvements become less obvious after detecting the main 5 jumps. Indeed, adding more change-points only allows to detect very small changes in the empirical mean and reduces the RSS very slightly. In particular, when the number of segments is too close to the maximum bound ($m/L_{\min}$) the algorithm tends to oversegment. For identifying changes in mean, the minimum segment length in our trial and error, for various sample sizes examined, ranged between 5 to 10. This particular choice was found to avoid spurious and excessive number of change-points and at the same time does not unduly lower the number of identified changes.

We then implemented the elbow method to help selecting the optimal number of segments by fitting the model with a range of values for $L_{\min}$ and $K$. Figures 1–6 display the RSS values obtained for $m = 50$, for Patterns I-III, for the smallest ($\alpha = 0.5$) and the largest ($\alpha = 1$) jump size as well. Note here that Figures 1–6 show the RSS values on the $y$-axis ($J_K, 1 \leq K \leq K_{\max}$) obtained only for $L_{\min}$ values varying from 1 to 5 versus different number of segments $K$ ($1 \leq K \leq 10$) on the $x$-axis, since for all cases examined the best-case scenario is obtained for the first time for $L_{\min} = 5$, that is the RSS value shows
Table 2. RSS values for $L_{\text{min}}$ vs. $K$ for Pattern II ($m = 50$).

![RSS values table](image)

a steep decrease up to the main 5 jumps (which in our case we know to be the optimal number) that are clearly visible, and a slow decrease after the optimum. If the line chart resembles an arm, then the elbow (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In the visualizer, elbow ($K = 3, 5$) is annotated with a two dashed vertical line. The RSS plots show a constantly decreasing curve with clear breakpoints for the optimal number of segments, notably when the largest jump size is considered. The value(s) of $K$ at which the slope abruptly changes provides an estimate $\hat{K}$ of the number of segments. The favourite numbers of segments for the cases examined appear to be $\hat{K} = 3$ (as expected, the smallest jump with $\alpha = 0.5$ is not always detected) and $\hat{K} = 5$. An improvement of the fit (decrease in the RSS values) is generally obtained when we start adding change-points for any $L_{\text{min}}$ value considered. When the model is fit with $L_{\text{min}} = 5$, we can see the elbow in the graph for the first time. An 'artifactual' decrease of RSS can happen when $K$ is too high (close to $m/L_{\text{min}}$) and corresponds generally to an oversegmentation. Moreover, it is observed that in some cases, for example, when the model is fit with $L_{\text{min}} = 5$, then for $K = 9$ and $K = 10$ (for which $K$ is too close to $m/L_{\text{min}} = 50/5 = 10$) the RSS values instead of decreasing start to increase. Therefore, the maximum number of segments to search for ($K_{\text{max}}$) could be by default fixed to the maximum bound ($m/L_{\text{min}}$), but as it can be seen, setting a smaller value is needed to achieve a more reasonable performance. In fact, the contrast $I_K$ should necessarily decrease when the model becomes more complex.

Taking into account the information provided from Tables 1–3 and Figures 1–6, one can conclude that:
Table 3. RSS values for $L_{\text{min}}$ vs. $K$ for Pattern III ($m = 50$).

| $L_{\text{min}}$ vs. $K$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------------------|---|---|---|---|---|---|---|---|---|----|
| Pattern III: $\alpha = 0.5$ ($m = 50$) |
| 1                        | 57.10 | 57.10 | 57.10 | 57.10 | 57.10 | 57.10 | 57.10 | 57.10 | 57.10 | 57.10 |
| 2                        | 50.99↓ ↓ | 51.09↓ ↓ | 51.19↓ ↓ | 51.28↓ ↓ | 51.37↓ ↓ | 51.46↓ ↓ | 51.56↓ ↓ | 51.66↓ ↓ | 51.78↓ ↓ | 51.99↓ ↓ |
| 3                        | 43.47↓ ↓ | 43.76↓ ↓ | 44.02↓ ↓ | 44.27↓ ↓ | 44.51↓ ↓ | 44.76↓ ↓ | 45.05↓ ↓ | 45.38↓ ↓ | 45.84↓ ↓ | 46.58↓ ↓ |
| 4                        | 38.95↓ ↓ | 39.54↓ ↓ | 40.07↓ ↓ | 40.56↓ ↓ | 41.07↓ ↓ | 41.57↓ ↓ | 42.06↓ ↓ | 42.61↓ ↓ | 43.32↓ ↓ | 44.47↓ ↓ |
| 5                        | 34.69↓ ↓ | 35.65↓ ↓ | 36.50↓ ↓ | 37.29↓ ↓ | 38.04↓ ↓ | 38.80↓ ↓ | 39.57↓ ↓ | 40.48↓ ↓ | 41.88↓ ↓ | 45.08↓ ↓ |
| 6                        | 31.32↓ ↓ | 32.74↓ ↓ | 34.04↓ ↓ | 35.25↓ ↓ | 36.47↓ ↓ | 37.29↓ ↓ | 39.63↓ ↓ | 44.63↓ Inf | Inf Inf Inf |
| 7                        | 28.32↓ ↓ | 30.27↓ ↓ | 32.08↓ ↓ | 33.83↓ ↓ | 35.58↓ ↓ | 37.91↑ ↓ | 44.73↑ Inf | Inf Inf Inf |
| 8                        | 25.68↓ ↓ | 28.21↓ ↓ | 30.56↓ ↓ | 32.86↓ ↓ | 35.20↓ ↓ | 41.16↑ Inf | Inf Inf Inf |
| 9                        | 23.31↓ ↓ | 26.42↓ ↓ | 29.36↓ ↓ | 32.24↓ ↓ | 35.72↑ Inf | Inf Inf Inf |
| 10                       | 21.18↓ ↓ | 24.89↓ ↓ | 28.42↓ ↓ | 31.94↓ ↓ | 40.09↑ Inf | Inf Inf Inf |
| Pattern III: $\alpha = 1.0$ ($m = 50$) |
| 1                        | 81.08 | 81.08 | 81.08 | 81.08 | 81.08 | 81.08 | 81.08 | 81.08 | 81.08 | 81.08 |
| 2                        | 69.99↓ ↓ | 70.01↓ ↓ | 70.04↓ ↓ | 70.08↓ ↓ | 70.13↓ ↓ | 70.18↓ ↓ | 70.26↓ ↓ | 70.35↓ ↓ | 70.49↓ ↓ | 70.76↓ ↓ |
| 3                        | 51.46↓ ↓ | 51.47↓ ↓ | 51.49↓ ↓ | 51.53↓ ↓ | 51.57↓ ↓ | 51.63↓ ↓ | 51.72↓ ↓ | 51.89↓ ↓ | 52.20↓ ↓ | 53.09↓ ↓ |
| 4                        | 45.01↓ ↓ | 45.22↓ ↓ | 45.41↓ ↓ | 45.61↓ ↓ | 45.85↓ ↓ | 46.12↓ ↓ | 46.38↓ ↓ | 46.72↓ ↓ | 47.26↓ ↓ | 48.55↓ ↓ |
| 5                        | 38.20↓ ↓ | 38.59↓ ↓ | 38.95↓ ↓ | 39.31↓ ↓ | 39.68↓ ↓ | 40.11↓ ↓ | 40.54↓ ↓ | 41.14↓ ↓ | 42.17↓ ↓ | 45.09↓ ↓ |
| 6                        | 34.04↓ ↓ | 34.88↓ ↓ | 35.69↓ ↓ | 36.51↓ ↓ | 37.54↓ ↓ | 39.13↓ ↓ | 41.98↑ ↓ | 52.12↑ Inf | Inf Inf Inf |
| 7                        | 30.50↓ ↓ | 31.93↓ ↓ | 33.35↓ ↓ | 34.81↓ ↓ | 36.49↓ ↓ | 40.37↑ ↓ | 53.89↑ Inf | Inf Inf Inf |
| 8                        | 27.51↓ ↓ | 29.55↓ ↓ | 31.59↓ ↓ | 33.72↓ ↓ | 36.07↓ ↓ | 46.72↑ Inf | Inf Inf Inf |
| 9                        | 24.88↓ ↓ | 27.55↓ ↓ | 30.24↓ ↓ | 33.05↓ ↓ | 36.47↑ Inf | Inf Inf Inf |
| 10                       | 22.54↓ ↓ | 25.85↓ ↓ | 29.22↓ ↓ | 32.75↓ ↓ | 40.09↑ Inf | Inf Inf Inf |

(a) The performance of the RO scheme is similar regardless of the Pattern (I–III) considered. In other words, the location of change-points and the overall duration of the stable/unstable period do affect the RSS values obtained, but not the overall performance observed.

(b) Statistical modeling of the simulated data allows us to get the optimal segmentation for any number of segments $K$ and proposes some possible number of segments. The value(s) of $K$ at which the slope abruptly changes provides an estimate $\hat{K}$ of the number of segments. Breakpoints for the optimal number of segments are clear when the largest jump scenario is considered, whereas the smallest jump is not always detected.

(c) The sequence $(J_K, 1 \leq K \leq K_{\text{max}})$ keeps decreasing significantly, and a significant improvement of the fit continues to be obtained when we start adding change-points for any $L_{\text{min}}$ value considered. Nevertheless, for $L_{\text{min}} = 5$, these improvements (decrease in the RSS values) suddenly become less obvious after detecting the main 5 jumps that are visible. Indeed, we observe a steep decrease up to the main 5 jumps (which in our case we know to be the optimal number), and a slow decrease after this optimum. Thus, adding more change-points only allows to detect very small changes in the empirical mean and reduces the RSS very slightly.

(d) When the number of segments is too close to the maximum bound ($m/L_{\text{min}}$) the algorithm tends to oversegment. Therefore, the maximum number of segments to search for ($K_{\text{max}}$) could be by default fixed to the maximum bound ($m/L_{\text{min}}$), but as it was seen, setting a smaller value is needed to achieve a more reasonable performance.

(e) For identifying changes in mean, the minimum segment length in our trial and error, for various sample sizes examined, ranged between 5 to 10. This particular choice was
Figure 1. RSS values for Pattern I ($\alpha = 0.5$).

Figure 2. RSS values for Pattern I ($\alpha = 1.0$).
**Figure 3.** RSS values for Pattern II ($\alpha = 0.5$).

**Figure 4.** RSS values for Pattern II ($\alpha = 1.0$).
Figure 5. RSS values for Pattern III ($\alpha = 0.5$).

Figure 6. RSS values for Pattern III ($\alpha = 1.0$).
found to avoid spurious and excessive number of change-points and at the same time does not unduly lower the number of identified changes.

In general, our results showed that

\[ L_{\text{min}} = \max \left( 5, \min \left( 10, \left\lfloor \frac{m}{10} \right\rfloor \right) \right) \]

(here \( \lfloor x \rfloor \) denotes the integer closest to \( x \)) provides a reasonable compromise.

Conclusively, in order to get a reasonable set of change-points we summarize here some restrictions that should be taken into account when proposing a new set of change-points. For any data set, it is not reasonable to propose a change-point at the first and last position in the series. When inserting a change-point between two other change-points, make sure that the difference between their positions is at least five, and check if after inserting a change-point, the difference between all consecutive break-points is at least five. This restriction on minimal segment length \( (L_{\text{min}} = 5) \) was found to help prevent overfitting without hindering the power to detect short, transient unstable phases. This practical rule accords with the minimum segment length chosen by Capizzi and Mazarotto [52]. However, for a practical purpose, it can be useful to decide automatically the minimum length of a step shift in accordance with the number of available observations \((m)\), thus the above empirical rule could be adopted.

### 3.2. Setting an upper limit on the size of the segmentation space

As aforementioned, when the number of segments to search for is set too close to the maximum bound \((m/L_{\text{min}})\) the algorithm may oversegment, and it is thus advisable to look carefully at the RSS curve in such a case. Moreover, this maximum bound can considerably slow the calculations for larger sample sizes and needs to be reduced to a more reasonable value. Details are given below.

Since the main goal here is to specify the maximum number of segments to search for in accordance with the sample size, we propose a procedure to ‘fine tune’ this parameter choosing the penalty constant adaptively to the data sample size, implementing the sample-size-adjusted BIC (SABIC) [53] by replacing sample size \( m \) with \( m^* \left(= \frac{m+2}{24} \right) \). The SABIC has the appealing properties of usually being stricter than AIC (depending on \( m \)) but not as strict as BIC, and it ‘adjusts’ for the sample size (except for very small sample sizes, i.e. \( m \leq 21 \), having a negative penalty encouraging needless complexity). This adjustment \((m^*)\) on the penalty component originates from Rissanen’s work on model selection for autoregressive time series models from a minimum description length perspective [54], and in practice it is usually used in latent class modelling working fairly well [55,56].

In this paper, we implement SABIC in terms of the RSS values, defined as

\[
S\text{ABIC}_{\text{RSS}} = m \ln \left( \frac{\text{RSS}}{m} \right) + k \ln \left( \frac{m+2}{24} \right),
\]

where \( k \) is the number of estimated parameters.

We followed similar simulation settings as previously described in 3.1. In all cases, we adopted the previously described empirical rule for (pre)specifying \( L_{\text{min}} \). The greatest value of \( K \) for which the \( S\text{ABIC}_{\text{RSS}} \) is minimized – for at least a 5% of the examined
Table 4. Changes in the mean for Patterns I-III ($m = 50$).

| Pattern (α = 0.5) | Estimated no. of segments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------------|--------------------------|---|---|---|---|---|---|---|---|---|----|
| No. of times      |                          |   |   |   |   |   |   |   |   |   |    |
| 6                 | 40                       | 272 | 804 | 2568 | 3212 | 2234 | 771 | 92 | 1 |
| Pattern (α = 1.0) | Estimated no. of segments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of times      |                          |   |   |   |   |   |   |   |   |   |    |
| 0                 | 0                        | 28 | 180 | 1814 | 3636 | 3009 | 1159 | 168 | 6 |
| Pattern (α = 0.5) | Estimated no. of segments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of times      |                          |   |   |   |   |   |   |   |   |   |    |
| 4                 | 25                       | 249 | 770 | 2555 | 3345 | 2262 | 719 | 71 | 0 |
| Pattern (α = 1.0) | Estimated no. of segments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of times      |                          |   |   |   |   |   |   |   |   |   |    |
| 0                 | 0                        | 24 | 113 | 1821 | 3832 | 3080 | 1028 | 101 | 1 |
| Pattern (α = 0.5) | Estimated no. of segments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of times      |                          |   |   |   |   |   |   |   |   |   |    |
| 1                 | 41                       | 255 | 739 | 2762 | 3503 | 2116 | 536 | 47 | 0 |
| Pattern (α = 1.0) | Estimated no. of segments | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of times      |                          |   |   |   |   |   |   |   |   |   |    |
| 0                 | 0                        | 35 | 157 | 2310 | 3888 | 2737 | 798 | 74 | 1 |

Cases (500/10,000 series) approximately – is chosen as the maximum number of segments to search for (instead of adopting the default maximum bound $m/L_{\text{min}}$). In other words, we retain the greatest value of $K$ for which

$$\min_{1 \leq K \leq K_{\text{max}}} \text{SABIC}_{\text{RSS}(K)}$$

is obtained.

The results are summarized in Tables 4 and 5, for $m = 50$ and $m = 100$ respectively. Tables 4 and 5 show that:

- for $m = 50$, 8 segments are detected 771 times ($\alpha = 0.5$) and 1159 times ($\alpha = 1.0$) for Pattern I, 719 times ($\alpha = 0.5$) and 1028 times ($\alpha = 1.0$) for Pattern II, 536 times ($\alpha = 0.5$) and 798 times ($\alpha = 1.0$) for Pattern III;
- for $m = 100$, 11 segments are detected 467 times ($\alpha = 0.5$) and 693 times ($\alpha = 1.0$) for Pattern I, 458 times ($\alpha = 0.5$) and 670 times ($\alpha = 1.0$) for Pattern II, 433 times ($\alpha = 0.5$) and 630 times ($\alpha = 1.0$) for Pattern III.

Based on the derived results from different sample sizes examined, we suggest using

$$K_{\text{max}} = \min(50, [\sqrt{m}] + 1)$$

(here $[x]$ denotes the largest integer closest to $x$) for (pre)specifying $K_{\text{max}}$. This upper limit on the size of the segmentation space provides a reasonable compromise (smaller than the maximum bound $m/L_{\text{min}}$) to avoid i. oversegmenting and ii. slowing the calculations for larger sample sizes. Only if the optimal number of segments identified ($\hat{K}$) is too close to $K_{\text{max}}$, one should think about increasing $K_{\text{max}}$. 


Table 5. Changes in the mean for Patterns I-III ($m = 100$).

| Estimated no. of segments | Pattern I ($\alpha = 0.5$) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10   |
|---------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| No. of times              | 3                           | 16  | 204 | 428 | 1218| 1477| 1904| 1713| 1332| 900 |
| Estimated no. of segments | 11                          | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| No. of times              | 467                         | 240 | 69  | 22  | 7   | 0   | 0   | 0   | 0   | 0   |

| Estimated no. of segments | Pattern I ($\alpha = 1.0$) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10   |
|---------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| No. of times              | 0                           | 0   | 4   | 47  | 744 | 1239| 1824| 1968| 1717| 1269|
| Estimated no. of segments | 11                          | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| No. of times              | 693                         | 306 | 139 | 240 | 69  | 22  | 7   | 0   | 0   | 0   |

| Estimated no. of segments | Pattern II ($\alpha = 0.5$) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10   |
|---------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| No. of times              | 2                           | 6   | 205 | 407 | 1223| 1589| 1864| 1764| 1299| 840 |
| Estimated no. of segments | 11                          | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| No. of times              | 458                         | 209 | 75  | 20  | 8   | 1   | 0   | 0   | 0   | 0   |

| Estimated no. of segments | Pattern II ($\alpha = 1.0$) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10   |
|---------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| No. of times              | 0                           | 0   | 9   | 33  | 745 | 1333| 1878| 1988| 1690| 1179|
| Estimated no. of segments | 11                          | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| No. of times              | 670                         | 299 | 130 | 39  | 4   | 3   | 0   | 0   | 0   | 0   |

| Estimated no. of segments | Pattern III ($\alpha = 0.5$) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10   |
|---------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| No. of times              | 4                           | 11  | 195 | 364 | 1196| 1983| 1736| 1292| 802 |
| Estimated no. of segments | 11                          | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| No. of times              | 433                         | 200 | 77  | 12  | 5   | 0   | 0   | 0   | 0   | 0   |

| Estimated no. of segments | Pattern III ($\alpha = 1.0$) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10   |
|---------------------------|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| No. of times              | 0                           | 0   | 6   | 26  | 724 | 1352| 1998| 2000| 1715| 1160|
| Estimated no. of segments | 11                          | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
| No. of times              | 630                         | 256 | 100 | 23  | 9   | 1   | 0   | 0   | 0   | 0   |

Remark 3.3: The results of the simulation study in Section 3 allow the reader to comprehend the reasoning behind the process followed for deriving the proposed empirical rules for the input parameters $L_{\text{min}}$ and $K_{\text{max}}$. We have conducted additional extensive simulations for various sample sizes and change-point patterns (considering different number of change-points), in order to demonstrate the robustness of the proposed empirical rules in various cases that are often encountered in practice. The results are given in Supplementary File 3 for space reasons.

4. Comparative study

In this section a comparative study is performed. A simulation-based performance comparison between the proposed scheme and widely used multiple change-point detection algorithms is conducted. Further, an application to real epidemiological data is presented; in particular, the developed technique is compared to its competitors, as regards its ability to detect successfully the signalled start and end weeks of epidemics.
4.1. Simulation-based benchmarking

In this subsection, the proposed RO scheme is compared to widely used multiple change-point detection algorithms, under a Gaussian framework. In particular, multiple changes in mean are found using approximate BinSeg [57] or exact (RO, SegNeigh [42] and PELT [16]) methods. Following similar settings as in Killick and Eckley [58] for identifying changes in mean, the simulated series are sequences of \( m = 200 \) independent Gaussian random variables of variance \( \sigma^2 = 1 \). The means in the four segments considered are \((0, 5, 10, 3)\) respectively. Multiple changes in mean are considered at locations 50, 100, 150 in simulated normal data. We conducted 1000 simulations using the \( R \) statistical environment and calling ‘cpt.mean’ function of the \( R \) package changepoint [58]. This function is capable of calculating the optimal positioning and number of change-points for data using the user specified method. As for the proposed RO scheme, the previously described empirical rules for (pre)specifying \( L_{\min} \) and \( K_{\max} \) were adopted. Details regarding the (hyper)parameters selected for the rest of the employed methods (BinSeg, SegNeigh and PELT) are given as follows:

**BinSeg:** for BinSeg algorithm (‘method = BinSeg’) the default choices were selected, that is ‘penalty = Manual’ and ‘pen.value = 2*\( \log(n) \)’, the minimum segment length was set equal to 2 (‘minseglen = 2’), and the maximum number of change-points to search for was set equal to 5 (‘Q = 5’);

**SegNeigh:** for SegNeigh algorithm (‘method = SegNeigh’) the default choices were selected, that is ‘penalty = Asymptotic’ and ‘pen.value = 0.01’, the minimum segment length was set equal to 1 (‘minseglen = 1’), and the maximum number of segments to search for was set equal to 5 (‘Q = 5’);

**PELT:** for PELT algorithm (‘method = PELT”) the default choices were selected, that is ‘penalty = Asymptotic’ and ‘pen.value = 0.01’. Here, the minimum segment length equals 1 and the maximum number of change-points to search for is infinite.

Table 6 presents all combinations of change-points detected after executing simultaneously all methods employed in the simulation study. The findings reveal that exact (RO, SegNeigh and PELT) methods outperform the approximate BinSeg method, in the sense that they return ‘3’ as the optimal number of change-points and ‘50, 100, 150’ as their exact locations, for 97% of the cases, whereas BinSeg for 86% of the cases, respectively. It can therefore be seen that the proposed RO scheme is a viable alternative to widely used exact methods for multiple change-point detection and inference, under a Gaussian framework.

4.2. A real-case example

A real-life example is given here to illustrate the implementation of the RO scheme in practice and to unfold its capabilities. In particular, the performance of the developed scheme is studied in the context of detecting outbreaks of an infectious disease. This paper focuses on the study of weekly influenza-like illness (ILI) rate data (provided from the Hellenic Centre for Disease Control and Prevention) for Greece, between September 29, 2014 (week40/2014) and October 2, 2016 (week39/2016), which were used for analysis purposes. Within this framework, it is of interest to compare the developed change-point
Table 6. Estimated change-points returned from each employed method.

| BinSeg     | SegNeigh    | PELT        | RO          | No. of times |
|------------|-------------|-------------|-------------|--------------|
| 50 100 150 | 50 100 150  | 50 100 150  | 50 100 150  | 863          |
| 51 100 150 | 50 100 150  | 50 100 150  | 50 100 150  | 15           |
| 50 100 150 | 50 100 150  | 50 100 150  | 50 100 150  | 12           |
| 50 99 150  | 50 99 150   | 50 99 150   | 50 99 150   | 11           |
| 50 71 100 150 | 50 100 150 | 50 100 150  | 50 100 150  | 10           |
| 45 50 100 150 | 50 100 150 | 50 100 150  | 50 100 150  | 10           |
| 50 100 138 150 | 50 100 150 | 50 100 150  | 50 100 150  | 10           |
| 50 100 150 170 | 50 100 150 | 50 100 150  | 50 100 150  | 8            |
| 50 52 150  | 50 100 150  | 50 100 150  | 50 100 150  | 8            |
| 51 100 150  | 50 100 150  | 50 100 150  | 50 100 150  | 7            |
| 50 100 110 150 | 50 100 150 | 50 100 150  | 50 100 150  | 7            |
| 50 100 150 156 | 50 100 150 | 50 100 150  | 50 100 150  | 7            |
| 5 50 100 150 | 50 100 150  | 50 100 150  | 50 100 150  | 6            |
| 50 95 100 150 | 50 100 150 | 50 100 150  | 50 100 150  | 6            |
| 50 100 150 183 | 50 100 150 | 50 100 150  | 50 100 150  | 5            |
| 50 101 150  | 50 101 150  | 50 101 150  | 50 101 150  | 4            |
| 49 50 100 150 | 49 100 150 | 49 100 150  | 49 100 150  | 4            |
| 50 99 117 150 | 50 99 150  | 50 99 150   | 50 99 150   | 2            |
| 50 81 87 100 150 | 50 100 150 | 50 100 150  | 50 100 150  | 1            |
| 49 50 100 150 155 | 50 100 150 | 50 100 150  | 50 100 150  | 1            |
| 50 100 118 144 150 | 50 100 150 | 50 100 150  | 50 100 150  | 1            |

detection technique to some competitors existing in the literature which are especially designed for outbreak detection purposes. The efficacy of the RO scheme is examined in terms of successful identification of the signalled start and end weeks of the detected epidemics.

Here, we perform the RO scheme for two periods under study (1st period: week 40/2014-week 39/2015, 2nd period: week 40/2015-week 39/2016), and then compare it with its competitors. Note here that from a statistical perspective, our problem consists in recovering the configuration of change-points (epidemic outbreaks) using the whole observed series instead of treating this problem as having two datasets. However, from an epidemiological perspective, it is of interest the observations to be processed in one batch for each period under study (Phase I change-point detection) and information to be returned regarding whether each sequence contains change-points. In this way, epidemiologically speaking, it is of interest to reveal the diverse disease patterns changing from year to year, since it is well known that the epidemiology and the burden of a disease also vary from year to year. Further, as previous authors working on the use of change-point analysis in the context of surveillance data of infectious disease counts (see for example Texier et al. [59], among others) we assume that the endemic or epidemic component of the process is driven by parameter (as mean incidence) which is considered as piecewise constant. Identifying an epidemic regime requires to identify the endemic states that come before and after this epidemic segment.

Figures 7 and 8 implement the elbow method to help selecting the optimal number of segments by fitting the model for the first and second period under study, respectively. The RSS plot (see Figure 7) shows a decreasing curve with a clear breakpoint for the optimal number of segments, that is \( \hat{K} = 3 \), for the first period under study, whereas the favourite numbers of segments appear to be \( \hat{K} = 4 \) and \( \hat{K} = 3 \) for the second period (see Figure 8).
Figures 9 and 10 illustrate the application of the RO scheme to each influenza period under study, respectively. The vertical lines in each panel show the estimated change-points whereas the horizontal lines represent the estimated values of the time-varying process mean within each segment. The behaviour of the vertical lines suggests an increase in ILI rate after the first 13 weeks and a decrease after the 25th week, for the first period under study (a case of ‘two-steps’ level change); it also suggests an increase in ILI rate after the first 14 weeks, a decrease after the 22nd week as well a further decrease after the 29th week, for the second period under study (a case of ‘three-steps’ level change). Note here that in the
latter case, from an epidemiological perspective, the RO scheme except of the sustained step shifts also identified short, transient shifts ($\hat{\tau}_{23}$: week09/2016 to $\hat{\tau}_{29}$: week15/2016) since the number of successive observations between these change-points is very close to the minimum number of successive observations allowed between two change-points ($L_{\text{min}} = 5$). It is important to clarify that in such a case only a practitioner may decide which change-points are ‘physically’ significant. It is not a statistical criteria that can decide, for instance, if the change-point located at a particular time-point should be associated to a significant change of the inherent properties of the characteristic(s) under study (e.g. change of the epidemiology of a disease). For this reason, we also fitted the model with $\hat{K} = 3$ segments (see Figure 11). Figure 11 suggests an increase in ILI rate after the first 14 weeks and a decrease after the 22nd week, for the second period.

We therefore concluded that the extracted signalled start (sw) and end weeks (ew) of the detected epidemics are sw01-ew12/2015 and sw01-ew08/2016, for the first and second period under study, respectively. We proceed below in examining the ability of the RO scheme for ‘successful’ change-point detection through benchmarking. For this purpose, the derived change-points (sw01-ew12/2015 and sw01-ew08/2016) are compared with those derived after executing some competing modelling techniques existing in the literature which are especially designed for outbreak detection purposes, in particular:
Figure 11. Change-point analysis of ILI rate data (week 40/2015 to week 39/2016) with $\hat{K} = 3$.

- $X(t) = \alpha_0 + \alpha_1 t + \gamma_1 \cos\left(\frac{2\pi t}{m}\right) + \delta_1 \sin\left(\frac{2\pi t}{m}\right) + \varepsilon(t)$, where $X(t)$ are the observed time series values (weekly ILI rate), $\varepsilon(t)$ are centred zero-mean random variables with variance $\sigma^2$, and $m$ denotes the number of observations within one year; M11 represents the current approach to ILI rate surveillance, that is the standard flu detection algorithm based on Serfling’s cyclic regression model Serfling [60], a routine method adapted in epidemiological surveillance systems in European countries (European Centre for Disease Prevention and Control - ECDC) and in the U.S.A. (Centers for Disease Control and Prevention - CDC);

- $X(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \gamma_1 \cos\left(\frac{2\pi t}{m}\right) + \delta_1 \sin\left(\frac{2\pi t}{m}\right) + \gamma_2 \cos\left(\frac{4\pi t}{m}\right) + \delta_2 \sin\left(\frac{4\pi t}{m}\right) + \gamma_3 \cos\left(\frac{8\pi t}{m}\right) + \delta_3 \sin\left(\frac{8\pi t}{m}\right) + \varepsilon(t)$; M23 represents an extended Serfling-type periodic regression model selected as the best fitting one by Parpoula et al. [61];

- $X(t) = \alpha_0 + \alpha_1 t + \gamma_1 \cos\left(\frac{2\pi t}{m}\right) + \delta_1 \sin\left(\frac{2\pi t}{m}\right) + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t + \lambda_1 \epsilon_{t-1} + \omega_1 \text{mintemp}$; MXM11 represents a model incorporating Auto-Regressive Moving Average (ARMA) terms and random meteorological covariates in the model structure as described by Kalligeris et al. [62];

- Segment neighbourhood algorithm [42] in conjunction with periodic-type ARMA time series modelling (SegNeigh-ARMA); an alternative methodology capable not only of identifying the beginning and end of the extreme periods that occur, but also of modelling the typical (non extreme) parts of the time series as described by Kalligeris et al. [63]; as for the (hyper)parameters selected for SegNeigh algorithm the default choices in ‘cpt.mean’ function [58] were selected, that is the minimum segment length was set equal to 1 (`minseglen = 1`) and the maximum number of segments to search for was set equal to 5 (`Q = 5`).

Summarizing the extracted signalled start and end weeks of the epidemics, i.e.

M11: sw01-ew13/2015 and sw01-ew08/2016,
M23: sw01-ew13/2015 and sw01-ew08/2016,
MXM11: sw01-ew12/2015 and sw05-ew08/2016,
SegNeigh-ARMA: sw01-ew12/2015 and sw01-ew08/2016,
RO: sw01-ew12/2015 nd sw01-ew08/2016,
Table 7. Estimated metrics for RO, SegNeigh-ARMA and MXM11 approaches.

| Method | RO & SegNeigh-ARMA | MXM11 |
|--------|--------------------|-------|
| AUC (95% CI) | 0.976 (0.926 to 0.996) | 0.881 (0.803 to 0.936) |
| ACC (95% CI) | 99.05% (94.81% to 99.98%) | 95.24% (89.24% to 98.44%) |
| SENS (95% CI) | 95.24% (76.18% to 99.88%) | 76.19% (52.83% to 91.78%) |
| SPEC (95% CI) | 100.0% (95.71% to 100.0%) | 100.0% (95.71% to 100.0%) |

all methods considered gave rise to similar results, producing almost identical signalled start and end weeks of the epidemics (except for MXM11 for the second period under study). This is the main reason we were forced to summarize in this way the start and end weeks for the epidemics (instead of showing them in a figure). Note that by placing all methods in one figure for comparison purposes, the vertical lines (breakpoints) of all methods almost coincide, thus differences are not visible.

We further examined the ability of RO, SegNeigh-ARMA and MXM11 approaches for detection of epidemics, using Receiver Operating Characteristic (ROC) curve analysis and its related metrics (Accuracy-ACC, Sensitivity-SENS, Specificity-SPEC, Area Under the ROC curve-AUC). Here, the derived change-points from RO, SegNeigh-ARMA and MXM11 approaches are compared with those derived after executing the ‘gold standard’ approach to influenza surveillance (standard and extended ECDC and CDC flu detection algorithm, models M11 & M23, respectively). Recall here that the extracted signalled start and end weeks of epidemics were found to be identical considering either M11 or M23 (i.e. sw01-ew13/2015 and sw01-ew08/2016). The same holds for SegNeigh and RO scheme, both based on change-point analysis, i.e. both yielded sw01-ew12/2015 and sw01-ew08/2016.

ROC curves have been a constructive measure for determining the trade-off between hit rates and false alarm rates of classifiers, allowing the determination and the comparison of the performances of several classification models, as discussed in Parpoula [64]. For a more detailed discussion on the theory and practice of ROC curves, the interested reader may refer to Pepe [65] among others. In our study, we performed pairwise comparison of ROC curves for the approaches examined in our study (RO, SegNeigh-ARMA and MXM11). More specifically, we estimated the aforementioned metrics along with their 95% Confidence Interval (CI) (exact Clopper-Pearson CIs for ACC, SENS and SPEC, and exact binomial CI for each derived AUC) following similar rationale as in Parpoula [64]. Table 7 presents the estimated metrics along with their 95% CIs. Table 8 presents the pairwise comparison of ROC curves. A two-sided statistical test was performed in a paired design, that is comparing paired data AUCs based on the empirical (nonparametric) ROC curves estimation implementing the empirical (nonparametric) methods of DeLong et al. [66] and of Hanley and McNeil [67]. In Table 8, SE denotes the standard error of the difference (diff.) between the AUCs (AUC\(_1\) – AUC\(_2\)), z-value corresponds to the calculated z-statistic (Zhou et al. [68]) for testing \(H_0 : AUC_1 = AUC_2\), and significance level represents the \(p\)-value (P), that is the probability that the true AUC\(_1\) equals AUC\(_2\), given the sample data. Note here that the aforementioned calculations were obtained for each employed method as well as for both periods under study.

Table 7 indicates that RO and SegNeigh-ARMA approaches outperform MXM11 in terms of higher AUC, ACC and SENS values, and seem to detect successfully the
Table 8. Pairwise comparison of ROC curves for RO, SegNeigh-ARMA and MXM11 approaches.

| RO & SegNeigh-ARMA ~ MXM11 | DeLong et al. [66] | Hanley & McNeil [67] |
|----------------------------|---------------------|-----------------------|
| Diff. between AUCs         | 0.0952              | 0.0952                |
| SE                        | 0.0439              | 0.0591                |
| 95% CI                    | 0.00919 to 0.181    | −0.0207 to 0.211      |
| z-value                   | 2.169               | 1.611                 |
| significance level        | $P = 0.0301$        | $P = 0.1073$          |

change-points (epidemic outbreaks) compared to the ‘gold standard’ approach to influenza surveillance. Further, the RO scheme does indeed compare favourably with its competitors since it may suffice for epidemiological surveillance purposes without need for recourse to an unfamiliar and more complicated model. Moreover, Table 8 indicates that statistically significant differences between AUCs were identified with the RO and MXM11, as well as the SegNeigh-ARMA and MXM11 approaches, but not statistically significant differences between AUCs were identified with the RO and SegNeigh-ARMA approaches (both based on change-point analysis). If the prevalent concern is that the process mean may incur a sustained jump in the mean, the results indicate that the RO scheme is a viable alternative to existing outbreak detection methods.

5. Concluding remarks

This paper addressed the problem of ‘a posteriori’ optimal multiple change-point detection and inference, performing a global segmentation. Our main goal was to recover the configuration of change-points using the whole observed series, aiming at detecting any change in the mean of a process in historical individual time series data. Within this framework, a recursive optimization algorithm was developed for optimally segmenting a time series and fine tuning the input parameters (that is, maximum number of segments to search for and the minimum segment length) which are usually unknown and there is no objective way to pre-specify them. Therefore, the proposed scheme addresses a wide class of real-life contexts and problems where the identification of optimal level shifts in a time series is the main goal.

Extensive simulation results showed that $L_{\text{min}} = \max(5, \min(10, \left\lfloor \frac{m}{10} \right\rfloor))$ and $K_{\text{max}} = \min(50, \left\lfloor \sqrt{m} \right\rfloor + 1)$ provide a reasonable compromise for practical prescriptions, useful to pre-specify automatically these two tuning parameters in accordance with the number of available observations. This restriction on minimum segment length was found to help prevent overfitting without hindering the power to detect short and transient OC phases, whereas this upper limit on the size of the segmentation space was found to help avoid oversegmenting and slowing the calculations for larger sample sizes. Further, the simulation-based findings reveal that exact (RO, SegNeigh and PELT) methods outperform the approximate BinSeg method, in the sense that they return correctly the optimal number of change-points and their exact locations for 97% of the examined cases, whereas BinSeg for 86% of the cases, respectively. It can therefore be seen that the proposed RO scheme is a viable alternative to widely used exact methods for multiple change-point detection and inference, under a Gaussian framework.
A real-life example was also given to illustrate the implementation of the proposed scheme in practice and to unfold its capabilities in the context of detecting outbreaks of an infectious disease. The developed scheme was found to compare favourably with its competitors since it detects successfully the change-points (epidemic outbreaks) compared to some typical modelling approaches existing in the literature (designed for outbreak detection purposes) and suffices for epidemiological surveillance purposes without need for recourse to an unfamiliar and more complicated model. Note here that in the presented Gaussian likelihood-based framework, we defined the probability distribution of the observed data assuming that the sequence of residual errors are i.i.d Gaussian variables, and in these settings the model was parametric. Lavielle [40] pointed out that a Gaussian likelihood can be effectively used as a contrast function for detecting changes in the mean (and/or in the variance) for any sequence of random (not necessarily Gaussian) variables, and shows through an example, that significative changes are well recovered without any assumptions on the dependence structure of the data (even for a discrete distribution considered). This is also the case for the real-life example presented here, where the change-points are well recovered without the need to take into account for autocorrelation in the epidemiological historical process data.

It is worth noting here that it was beyond the scope of the paper to develop a distribution-free change-point detection scheme, an alternative one to those already existing in the literature. The twofold goal of this paper was to develop a change-point detection scheme capable not only for optimal time series segmentation under a Gaussian framework, but also for input parameter fine-tuning in the context of multiple change-point analysis. Nevertheless, the derived practical rules (for pre-specifying the maximum number of hypothetical change-points to search for subject to the constraint of a minimum segment length, in accordance with a given sample size) could also be adopted for change-point detection methods monitoring process locations that make no distributional assumptions. As a future work, it would be interesting to adapt such an optimization scheme based on a multivariate data concept, explore and fine-tune the input parameters in order to provide accordingly reasonable compromises to be implemented either on a fully parametric or a distribution-free change-point analysis framework.

**Acknowledgments**

The authors would like to thank the Department of Epidemiological Surveillance and Intervention of the Hellenic National Public Health Organization (NPHO) for providing the influenza-like illness (ILI) rate data, collected weekly through the sentinel surveillance system. The authors would also like to thank the editor and the anonymous referees who provided valuable suggestions and comments which improved significantly the paper.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**ORCID**

Christina Parpoula [http://orcid.org/0000-0002-3585-4773](http://orcid.org/0000-0002-3585-4773)

Alex Karagrigoriou [http://orcid.org/0000-0002-4919-2133](http://orcid.org/0000-0002-4919-2133)
References

[1] Pettitt AN. A non-parametric approach to the change point problem. Appl Stat. 1979;28(2):126–135.
[2] Aue A, Horváth L. Structural breaks in time series. J Time Ser Anal. 2013;34(1):1–16.
[3] Inclan C, Tiao GC. Use of cumulative sums of squares for retrospective detection of changes of variance. J Am Stat Assoc. 1994;89(427):913–923.
[4] Gao Z, Shang Z, Du P, et al. Variance change point detection under a smoothly-changing mean trend with application to liver procurement. J Am Stat Assoc. 2019;114(526):773–781.
[5] Davis RA, Lee T, Rodriguez-Yam GA. Structural break estimation for nonstationary time series models. J Am Stat Assoc. 2006;101(473):223–239.
[6] Gallagher C, Lund R, Robbins M. Changepoint detection in daily precipitation data. Environmetrics. 2012;23(5):407–419.
[7] Dudding BP, Jennett W. Quality control charts: BS 600R. London: British Standards Institution; 1942.
[8] Jandhyala V, Fotopoulos S, MacNeill I, et al. Inference for single and multiple change-points in time series. J Time Ser Anal. 2013;34(4):423–446.
[9] Wang D, Yu Y, Rinaldo A. Univariate mean change point detection: penalization, CUSUM and optimality. Electron J Stat. 2020;14(1):1917–1961.
[10] Shi X. A survey of changepoint techniques for time series data [all dissertations]. Vol. 2697, Clemson University; 2020.
[11] Page ES. A test for a change in a parameter occurring at an unknown point. Biometrika. 1955;42(3/4):523–527.
[12] MacNeill IB. Tests for change of parameter at unknown times and distributions of some related functionals on brownian motion. Ann Stat. 1974;2(5):950–962.
[13] Lebarbier E. Detecting multiple change-points in the mean of Gaussian process by model selection. Signal Processing. 2005;85(4):717–736.
[14] Lavielle M, Teyssiere G. Detection of multiple change-points in multivariate time series. Lithuanian Math J. 2006;46(3):287–306.
[15] Harchaoui Z, Lévy-Leduc C. Multiple change-point estimation with a total variation penalty. J Am Stat Assoc. 2010;105(492):1480–1493.
[16] Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc. 2012;107(500):1590–1598.
[17] Matteson DS, James NA. A nonparametric approach for multiple change point analysis of multivariate data. J Am Stat Assoc. 2014;109(505):334–345.
[18] Fryzlewicz P. Wild binary segmentation for multiple changepoint detection. Ann Stat. 2014;42(6):2243–2281.
[19] Niu YS, Hao N, Zhang H. Multiple change-point detection: a selective overview. Stat Sci. 2016;31(4):611–623.
[20] Haynes K, Eckley IA, Fearnhead P. Computationally efficient changepoint detection for a range of penalties. J Comput Graph Stat. 2017;26(1):134–143.
[21] Baranowski RCY, Fryzlewicz P. Narrowest-over-threshold detection of multiple change points and change-point-like features. J R Stat Soc: Ser B. 2019;81(3):649–672.
[22] Li Y, Lund RB, Hewsarachchi A. Multiple changepoint detection with partial information on changepoint times. Electron J Stat. 2019;13(2):2462–2520.
[23] Ray BK, Tsay RS. Bayesian methods for change-point detection long-range dependent processes. J Time Ser Anal. 2002;23(6):687–705.
[24] Berkes I, Horváth L, Kokoszka PSQ-M. On discriminating between long-range dependence and changes in mean. Ann Stat. 2006;34(3):1140–1165.
[25] Lu Q, Lund RB. Simple linear regression with multiple level shifts. Canadian J Stat. 2007;35(3):447–458.
[26] Lund R, Wang XL, Lu Q, et al. Changepoint detection in periodic and autocorrelated time series. J Clim. 2007;20(20):5178–5190.
[27] Gombay E. Change detection in autoregressive time series. J Multivar Anal. 2008;99(3): 451–464.
[28] Shao X, Zhang X. Testing for change points in time series. J Am Stat Assoc. 2010;105(491): 1228–1240.
[29] Robbins M, Gallagher C, Lund R, et al. Mean shift testing in correlated data. J Time Ser Anal. 2011;32(5):498–511.
[30] Xie Y, Huang J, Willett R. Change-point detection for high-dimensional time series with missing data. IEEE J Sel Top Signal Process. 2012;7(1):12–27.
[31] Shen J, Gallagher CM, Lu Q. Detection of multiple undocumented change-points using adaptive lasso. J Appl Stat. 2014;41(6):1161–1173.
[32] Shao X. Self-normalization for time series: a review of recent developments. J Am Stat Assoc. 2015;110(512):1797–1817.
[33] Cho H, Fryzlewicz P. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. J R Stat Soc B. 2015;77(2):475–507.
[34] Robbins MW, Gallagher CM, Lund RB. A general regression changepoint test for time series data. J Am Stat Assoc. 2016;111(514):670–683.
[35] Chakar S, Lebarbier E, Lévy-Leduc C, et al. A robust approach for estimating change-points in the mean of an AR(1) process. Bernoulli. 2017;23(2):1408–1447.
[36] Zhang T, Lavitas L. Unsupervised self-normalized change-point testing for time series. J Am Stat Assoc. 2018;113(522):637–648.
[37] Gurevich G, Vexler A. Retrospective change point detection: from parametric to distribution free policies. Commun Stat –Simul Comput. 2010;39(5):899–920.
[38] Xie H, Li D, Xiong L. Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation. Stochastic Environ Res Risk Assess. 2014;28(7):1643–1655.
[39] Ross GJ. Parametric and nonparametric sequential change detection in R: the cpm package. J Stat Softw. 2015;66(3):1–20.
[40] Lavielle M. Detection of multiple changes in a sequence of dependent variables. Stoch Process Their Appl. 1999;83(1):79–102.
[41] Tartakovsky A, Nikiforov I, Basseville M. Sequential analysis: hypothesis testing and change-point detection. Boca Raton: Chapman and Hall/CRC; 2015.
[42] Auger IE, Lawrence CE. Algorithms for the optimal identification of segment neighborhoods. Bull Math Biol. 1989;51(1):39–54.
[43] Jackson B, Scargle JD, Barnes D, et al. An algorithm for optimal partitioning of data on an interval. IEEE Signal Process Lett. 2005;12(2):105–108.
[44] Rigail G. A pruned dynamic programming algorithm to recover the best segmentations with 1 to K max change-points. Journal de la Société Française de Statistique. 2015;156(4):180–205.
[45] Maidstone R, Hocking T, Rigail G, et al. On optimal multiple changepoint algorithms for large data. Stat Comput. 2017;27(2):519–533.
[46] Fearnhead P, Rigail G. Changepoint detection in the presence of outliers. J Am Stat Assoc. 2019;114(525):169–183.
[47] Bellman R. The theory of dynamic programming. Bulletin Amer Math Soc. 1954;60:503–515.
[48] Lavielle M, Moulines E. Least-squares estimation of an unknown number of shifts in a time series. J Time Ser Anal. 2000;21(1):33–59.
[49] Picard F, Robin S, Lavielle MVC, et al. A statistical approach for array CGH data analysis. BMC Bioinformatics. 2005;6:27.
[50] Lavielle M. Using penalized contrasts for the change-point problem. Signal Processing. 2005;85(8):1501–1510.
[51] Niekum S, Osentoski S, Barto AG. CHAMP: changepoint detection using approximate model parameters. Carnegie-Mellon University Pittsburgh PA Robotics Institute; 2014. Tech Rep. (No. CMU-RITR-14-10).
[52] Capizzi G, Masarotto G. Phase I distribution-free analysis of univariate data. J Quality Technol. 2013;45(3):273–284.
[53] Sclove SL. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika. 1987;52:333–343.
[54] Stine RA. Model selection using information theory and the MDL principle. Sociol Methods Res. 2004;33(2):230–260.
[55] Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Struct Equ Modeling. 2007;14(4):535–569.
[56] Tein JY, Coxe S, Cham H. Statistical power to detect the correct number of classes in latent profile analysis. Struct Equ Modeling. 2013;20(4):640–657.
[57] Scott AJ, Knott M. A cluster analysis method for grouping means in the analysis of variance. Biometrics. 1974;30(3):507–512.
[58] Killick R, Eckley IA. changepoint: an R package for changepoint analysis. J Stat Softw. 2014;58(3):1–19.
[59] Texier G, Farouh M, Pellegrin L, et al. Outbreak definition by change point analysis: a tool for public health decision?. BMC Med Inform Decis Mak. 2016;16:33.
[60] Serfling R. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep. 1963;78(6):494–506.
[61] Parpoula C, Karagrigoriou A, Lambrou A. Epidemic intelligence statistical modelling for biosurveillance. In: Blömer J, Kotsireas IS, Kutsia T, Simos DE, editors. Mathematical aspects of computer and information sciences. Switzerland: Springer International Publishing; 2017. p. 349–363. (Lecture Notes in Computer Science (LNCS10963)).
[62] Kalligeris EN, Karagrigoriou A, Parpoula C. On mixed PARMA modeling of epidemiological time series data. Commun Stat: Case Stud, Data Anal Appl. 2020;6(1):36–49.
[63] Kalligeris EN, Karagrigoriou A, Parpoula C. Periodic-type ARMA modeling with covariates for time-series incidence data via changepoint detection. Stat Methods Med Res. 2020;29(6):1639–1649.
[64] Parpoula C. A distribution-Free control charting technique based on change-point analysis for detection of epidemics. Stat Methods Med Res. 2022;1–18. DOI:10.1177/09622802221079347.
[65] Pepe MS. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press; 2003.
[66] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–845.
[67] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148:839–843.
[68] Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. 2nd ed. New Jersey: John Wiley and Sons; 2011.