Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Detection of multiple change-points in multivariate data

Edgard M. Maboudou-Tchao & Douglas M. Hawkins

To cite this article: Edgard M. Maboudou-Tchao & Douglas M. Hawkins (2013) Detection of multiple change-points in multivariate data, Journal of Applied Statistics, 40:9, 1979-1995, DOI: 10.1080/02664763.2013.800471

To link to this article: http://dx.doi.org/10.1080/02664763.2013.800471

Published online: 24 May 2013.

Submit your article to this journal

Article views: 317

View related articles

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=cjas20

Download by: [University of Central Florida] **Date:** 27 February 2017, At: 12:17

Journal of Applied Statistics, 2013
Vol. 40, No. 9, 1979–1995, http://dx.doi.org/10.1080/02664763.2013.800471

Taylor & Francis
Taylor & Francis Group

Detection of multiple change-points in multivariate data

Edgard M. Maboudou-Tchao[a]* and Douglas M. Hawkins[b]

[a]*Department of Statistics, University of Central Florida, Orlando, FL, USA;* [b]*School of Statistics, University of Minnesota, Minneapolis, MN, USA*

(*Received 23 April 2012; final version received 25 April 2013*)

The statistical analysis of change-point detection and estimation has received much attention recently. A time point such that observations follow a certain statistical distribution up to that point and a different distribution – commonly of the same functional form but different parameters after that point – is called a change-point. Multiple change-point problems arise when we have more than one change-point. This paper develops a method for multivariate normally distributed data to detect change-points and estimate within-segment parameters using maximum likelihood estimation.

Keywords: regression trees; binary splitting; principle of optimality; separability; dynamic programming

1. Introduction

Segmentation and change-point models are increasingly important in numerous research fields and practical applications, including economics, finance, medicine, and multimedia processing. Harchaoui *et al.* [18] gave an example in multimedia processing where one might want to detect a change from human speech to some other sound in a movie automatically based on a data representation of features coming from both audio and visual tracks. Zhang *et al.* [43] described an application in medicine where a change-point procedure was used to detect copy-number variation along the genome and how it could be useful for microarray and genetic linkage. Othus *et al.* [31] developed a cure survival model that allowed for change-point effects in covariates to investigate a potential change-point in the age of diagnosis of prostate cancer. Terrera *et al.* [36] suggested a change-point model to fit data from a population-based longitudinal study of aging to model longitudinal processes. Balestrassi *et al.* [2] proposed a novel method to detect dynamic changes in nonlinear time series. Modisett and Maboudou-Tchao [30] used change-point models in economics and finance to detect regimes of constant drift and volatility. These are examples of the vast literature on change-point detection problem: see Basseville and Nikiforov [3] or Brodsky and Darkhovsky [8].

*Corresponding author. Email: edgard.maboudou@ucf.edu

© 2013 Taylor & Francis

Multivariate change-point detection is useful for the development, planning and modeling of multivariate control charts and time series. In control chart settings, the proposed method is ideal for phase I analysis to verify that the Phase I data set consists of a single homogeneous sequence. When process parameters are unknown, as they most commonly are, a common approach is the 'plug-in' method, in which process parameters are estimated from an 'in-control' sample and used to calibrate control charts. Current methodology, however, lacks good tools for deciding whether the phase I data set was indeed homogeneous, so the multivariate change-point detection method proposed here may be an important tool in industrial quality control where scrubbing historical data to detect shifts in distribution is necessary to establish which observations represent normal (in control) operation.

The study of change-point problems in time-ordered data started in the 1950s. Page [32,33] introduced a simple process to detect a single change. Since then, change-point problems have received extensive attention. For instance, Chernoff and Zacks [11], Gardner [14], and Hawkins [20] studied testing and estimation of a change in the mean of a normal model. Most work has assumed that the data are from some particular family of distributions, with focus on the normal family, though some has used non-parametric methods. Numerous authors have proposed methods for detection and testing of change-points. These methods include likelihood ratio, nonparametric and Bayesian approaches. Ghosh *et al.* [15] proposed a generalized likelihood ratio procedure and a Bayesian procedure for change-point problems for mean direction of the von Mises distribution, both when the concentration parameter was known and when it was unknown. Sengupta and Laha [34] proposed a simple fully Bayesian analysis of the change-point problem for directional data in the parametric framework with the von Mises or circular normal distribution as the underlying distribution. Itoh and Kurths [24] discussed change-point detection in climate time series by a nonparametric method. Darkhovski [13] proposed nonparametric methods for change-point problems using the nonparametric family of Kolmogorov–Smirnov type statistics. Lai and Xing [26] discussed a simple Bayesian approach to multiple change-points. Some authors suggested using total variation to find change-points. Harchaoui and Levy-Leduc [17] proposed to detect change-points using the least absolute shrinkage and selection operator (LASSO). Vert and Bleakley [39] discussed a fast detection of multiple change-points using group least angle regression. Bleakley and Vert [5] proposed a group fused LASSO for multiple change-point detections.

The multiple change-point problem has several issues – choosing suitable parametric forms for within-segment models, deciding whether there is any change (hypothesis testing problem), locating segment boundaries (estimation problem), and deciding the appropriate number of change-points (model selection problem).

Working with two-segment models is relatively straightforward. Computation comes down to a line search for the putative change-point, along with parameter estimation within each of the segments so defined. Likelihood ratio tests for several common models are similarly easy to define, though less so to use for formal inference. Theoretical and asymptotic results are given in Csörgö and Horvath [12]. Bonferroni tests are simple but conservative, though the improved Bonferroni approach of Worsley [41] is effective in removing much of the conservatism.

Going from two to three or more segments complicates matters considerably, from both computational and inferential viewpoints. One approach is the hierarchic binary splitting algorithm proposed by Vostrikova [40]. In this, the data are tested for a single change-point and then split into two subsegments if there is a significant change. The same binary splitting algorithm is applied to each of the resulting segments and if any gives a significance, it is split. This process is repeated on each new segment identified until none of the segments yields a significant change-point. Vostrikova's model was of normally distributed data in which the mean changed from one segment to the next and she used the Bonferroni approximation for the test. Srivastava and Worsley [35] refined the hierarchical segmentation algorithm using an improved Bonferroni multiplier.

The hierarchic solution is a 'greedy' algorithm, making choices that are optimal at each step, but not necessarily optimal in terms of optimizing the overall criterion. The global optimum, however, cannot be found using hierarchical methods and therefore, the hierarchical binary splitting algorithm does not guarantee optimum splits if there are more than two segments.

The computational aspect of multi-segment fits to data falling within the exponential family was addressed by Hawkins [21], where it was shown that, provided the overall model likelihood is separable, maximum likelihood fits can be made using a dynamic programming (DP) formulation whose computation increases only linearly with the number of segments. Similar DP formulations are given in Hawkins and Ten Krooden [23], Hawkins [19], Venter and Steele [38], and Braun and Muller [6,7].

In this paper, we propose a method to perform a multiway split on multivariate Gaussian data for the case of change in mean vector and/or covariance matrix. We derive the likelihood model for the multiway splitting and show that a DP algorithm can be used to find globally optimal splits. Algorithms to speed up necessary computations are outlined. Finally, we illustrate our method on actual data sets.

2. Multivariate normal change-point model

In this section, we study the problem of simultaneous changes in the mean vector and/or covariance matrix of a multivariate Gaussian model. We assume that

(1) The p component vector $\boldsymbol{x}_i$, $i = 1, 2, \ldots, n$ is a sequence of independently distributed p-dimensional normal random vectors.
(2) We have k segments defined by $k-1$ change-points, $\tau = (\tau_1, \tau_2, \ldots, \tau_{k-1})$. For notational convenience, we will also bracket the whole sequence with notional change-points $\tau_0 = 0$, $\tau_k = n$.
(3) The data vectors within segment j are independently and identically distributed multivariate normal with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, in other words for

$$\tau_{j-1} < i \leq \tau_j, \quad \boldsymbol{x}_i \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

(all $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown).

Inferentially, we are interested in testing the hypothesis:

H_0: No change-point, that is $k = 1$, vs.
H_a: k segments with $k-1$ change-points $\tau = (\tau_1, \tau_2, \ldots, \tau_{k-1})$ and $\boldsymbol{\mu}_j \neq \boldsymbol{\mu}_{j-1}$ and/or $\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}_{j-1}, j = 2, \ldots, k$.

2.1 *Maximum likelihood estimation of the parameters*

Under H_0, the log likelihood is

$$\log L_0(\boldsymbol{\theta}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}), \tag{1}$$

where the overall parameter vector $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $|\boldsymbol{\Sigma}|$ represents the determinant of $\boldsymbol{\Sigma}$.

Under H_a, the log likelihood is

$$\log L_1(\boldsymbol{\theta}) = -\frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{k} r_j \log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{k}\sum_{i=\tau_{j-1}+1}^{\tau_j} (\boldsymbol{x}_i - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_j), \qquad (2)$$

where $r_j = \tau_j - \tau_{j-1}$ is the number of vectors in the jth segment, $\boldsymbol{\mu}_j$ is the unknown mean vector of the jth segment, $\boldsymbol{\Sigma}_j$ is the unknown covariance matrix of the jth segment, and $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \ldots, (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.

Maximization of the likelihood is a two-step procedure involving an 'outer' and an 'inner' problem. The outer problem consists of estimating the change-points τ_j. Once trial values for the change-points are specified, this leads to the 'inner' problem of estimating the within-segment mean vectors and covariance matrices of the segments.

Now if the change-points τ_j are given, then the inner problem of finding MLEs of the mean vectors and covariance matrices is solved by the standard one-way Manova layout.

Consider a segment $\boldsymbol{x}_{h+1}, \ldots, \boldsymbol{x}_m$. Write the sample mean vector and sum of squares and cross products matrix of this subsequence as

$$\bar{\boldsymbol{x}}_{h,m} = \sum_{j=h+1}^{m} \frac{\boldsymbol{x}_j}{(m-h)},$$

$$\mathbf{A}_{h,m} = \sum_{j=h+1}^{m} (\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h,m})(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h,m})'.$$

Then, given the boundaries τ_{j-1}, τ_j, the MLEs of the jth segment are

$$\hat{\boldsymbol{\mu}}_j = \bar{\boldsymbol{x}}_{\tau_{j-1},\tau_j}, \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_j = \frac{\mathbf{A}_{\tau_{j-1},\tau_j}}{r_j}, \qquad (3)$$

where r_j is the segment length $\tau_j - \tau_{j-1}$.

It is also convenient to define the 'stage return'

$$Q(h,m) = (m-h)\log\left|\frac{\mathbf{A}_{h,m}}{(m-h)}\right|. \qquad (4)$$

THEOREM 1 *Under the assumptions* 1–3, -2 *times the log-likelihood, from which the multiple change-point for the multivariate Gaussian model of* $\boldsymbol{x}$ *under* H_a *can be deduced, is*

$$-2\log L_1(\hat{\boldsymbol{\theta}}) = np(\log(2\pi)+1) + \sum_{j=1}^{k} Q(\tau_{j-1}, \tau_j). \qquad (5)$$

Proof Substituting the estimated $\boldsymbol{\theta}$ from Equation (3) in Equation (2)

$$\begin{aligned}
-2\log L_1(\hat{\boldsymbol{\theta}}) &= np\log(2\pi) + \sum_{j=1}^{k+1} r_j \log|\hat{\boldsymbol{\Sigma}}_j| + \sum_{j=1}^{k}\sum_{i=\tau_{j-1}+1}^{\tau_j} (\mathbf{X}_i - \boldsymbol{\mu}_j)'\hat{\boldsymbol{\Sigma}}_j^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_j) \\
&= np\log(2\pi) + \sum_{j=1}^{k} r_j \log|\hat{\boldsymbol{\Sigma}}_j| + \sum_{j=1}^{k+1} tr\Big(\hat{\boldsymbol{\Sigma}}_j^{-1}\sum_{i=\tau_{j-1}+1}^{\tau_j} (\mathbf{X}_i - \boldsymbol{\mu}_j)(\mathbf{X}_i - \boldsymbol{\mu}_j)'\Big) \\
&= np\log(2\pi) + \sum_{j=1}^{k} r_j \log|\hat{\boldsymbol{\Sigma}}_j| + \sum_{j=1}^{k+1} tr(\hat{\boldsymbol{\Sigma}}_j^{-1} r_j \hat{\boldsymbol{\Sigma}}_j)
\end{aligned}$$

$$= np\log(2\pi) + \sum_{j=1}^{k} r_j \log|\hat{\boldsymbol{\Sigma}}_j| + \sum_{j=1}^{k+1} pr_j$$

$$= np\log(2\pi) + \sum_{j=1}^{k} r_j \log|\hat{\boldsymbol{\Sigma}}_j| + np$$

$$= np(\log(2\pi) + 1) + \sum_{j=1}^{k} r_j \log|\hat{\boldsymbol{\Sigma}}_j|$$

$$= np(\log(2\pi) + 1) + \sum_{j=1}^{k} Q(\tau_{j-1}, \tau_j),$$

recalling the definition of $Q(h, m)$. ■

In the optimization calculation and the subsequent testing, the irrelevant constant $np(\log(2\pi) + 1)$ is omitted, and attention focused on the term $\sum_{j=1}^{k} r_j \log|\hat{\boldsymbol{\Sigma}}_j|$ or $\sum_{j=1}^{k} Q(\tau_{j-1}, \tau_j)$.

This solution to the 'inner' problem of estimating the within-segment parameters given the segment boundaries shifts the focus to the 'outer' problem of finding the maximum likelihood estimators of the change-points τ_j. As the likelihood function is not continuous in τ_j, conventional function optimization procedures are not helpful for this. The objective function

$$\sum_{j=1}^{k} Q(\tau_{j-1}, \tau_j)$$

is however separable, and satisfies Bellman's principle of optimality, which leads to a DP algorithm to find the optimal k segment fit.

Omitting the irrelevant constant, define $L(r, m)$ to be the minimized -2 log-likelihood function for fitting $r - 1$ change-points to the first m observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$.

$$L(r, m) = \min_{t} \sum_{j=1}^{r} Q(t_{j-1}, t_j), \tag{6}$$

where the optimization is over the change-points $t_0 = 0 < t_1 < \cdots < t_r = m$. The solution to our actual problem of fitting $k - 1$ segments to all n data vectors is provided by the t values leading to $L(k, n)$ and the resultant within-segment mean vectors and covariance matrices.

This solution is provided by DP, paralleling that of Hawkins [21] for the mean-only univariate setting. Following Hawkins [21], it is easily shown by contradiction that

$$L(1, m) = Q(0, m), \quad L(r, m) = \min_{t_r} L(r - 1, t_r) + Q(t_r, m). \tag{7}$$

The DP exploits this recursive relation for $L(r, m)$ by the initialization $L(0, m) = Q(0, m), m = 1, 2, \ldots, n$. Thereafter, each $L(r, m)$ can be computed using a table of the previously computed $L(r - 1, t)$ and $Q(t, m)$ to find the t minimizing the sum (7). While performing the minimization leading to each $L(r, m)$, we also keep track of $H(r, m)$, the t value leading to the minimum. This table of $H(r, m)$ is used for the final stage of 'back-tracing' to find the optimal segment boundaries. The rightmost segment boundary is given by $\hat{\tau}_{k-1} = H(k, n)$, and each subsequent boundary calculated as $\hat{\tau}_j = H(j + 1, \hat{\tau}_{j+1})$.

Note the inherent degeneracy of the formulation as discussed so far. The SSCP matrix $\mathbf{A}_{h,m}$ is necessarily singular, and so the likelihood is infinite, if $(m - h) \leq p$, and so there is an unhelpful minimum of the overall objective function obtained by setting any pair of putative change-points

closer than $p+1$ apart. To avoid this degenerate solution, the range of values of t_r searched in the calculation of $L(r,m)$ is not allowed to extends above $m-p$. This degeneracy also implies that $L(r,m)$ is not usefully defined for $m < r(p+1)$.

The maximized log likelihood of the k segments fitted to the full data set is (apart from an irrelevant additive constant) $-\frac{1}{2}L(k,n)$. The individual segment means and covariance matrices are given by these segment boundaries and the $\bar{\boldsymbol{x}}_{h,m}$ and $\mathbf{A}_{h,m}/(m-h)$ defined earlier.

2.2 *'Bartlett-style' correction for the LR test*

The model formulation used is a heteroscedastic multivariate analysis of variance layout, and the criterion parallels the multivariate generalized likelihood ration (GLR) test of identity of normal samples. This test statistic for identity gives an asymptotic chi-squared distribution, but this asymptotic approximation is well known to be poor, and specifically liberal, if any of the populations involved in the test is small. This suggests that the methodology, as set out so far, will have a propensity to fit very short segments, whereas the objective is to treat all segment lengths even-handedly. The standard remedy for small-sample departure of the GLR test for identity from the asymptotic chi-squared distribution is a Bartlett correction, in which the test statistic is divided by its null expectation.

Rather than applying the Bartlett correction to the final test of the number of segments, we prefer to incorporate a sample-size correction into the optimization step itself to reduce the tendency to form very small groups in the first place. This correction consists of redefining the stage return $Q(h,m)$ by subtracting its null expectation. Recall that

$$\begin{aligned} Q(h,m) &= (m-h)\log\left|\frac{\mathbf{A}_{h,m}}{(m-h)}\right| \\ &= (m-h)\log|\mathbf{A}_{h,m}| - (m-h)p\log(m-h). \end{aligned}$$

If the true covariance matrix of the segment is $\mathbf{I}$, then by Johnson and Kotz [25, p. 198], $E\log|\mathbf{A}_{h,m}| = p\log 2 + \sum_{j=1}^{p}\psi(m-h-j)/2$, where $\psi(z)$ is the digamma function

$$\psi(z) = \frac{\mathrm{d}}{\mathrm{d}z}\log\Gamma(z).$$

Abramowitz and Stegun [1, p. 258] give basic results on the digamma function, notably: $\psi(\frac{1}{2}) = -\gamma - 2\log 2$, $\psi(1) = -\gamma$, where γ is the Euler constant 0.5772, and $\psi(z+1) = \psi(z) + 1/z$. As all required values of ψ are for integer or half-integer arguments, this allows us to calculate the digamma function exactly for any arguments that we need.

This leads to the modified definition of the stage returns

$$Q^*(h,m) = (m-h)[-p\log(m-h) + \log|\mathbf{A}_{h,m}|] - g(m-h,p), \tag{8}$$

where $g(m,p) = pm\log(2/m) + m\sum_{j=1}^{p}\psi((m-j)/2)$, which we use in place of the $Q(h,m)$ of the DP algorithm.

3. Computation

3.1 *Stage return computation*

The main computational burden of the DP optimization is the stage returns $Q^*(h,m)$. These may be computed once and for all and stored in one half of an $n \times n$ array, or they may be computed 'on the fly', as they are used. As all Q values are re-used with each additional segment fitted, precomputation is by far the more efficient.

A reasonably fast, numerically stable approach is given by the use of Cholesky factorizations. Append 1 to each data vector $\boldsymbol{x}_i$ by writing

$$\boldsymbol{z}_i = (1, \boldsymbol{x}_i')'$$

and defining

$$\mathbf{B}_{h,m} = \sum_{j=h+1}^{m} \boldsymbol{z}_j \boldsymbol{z}_j',$$

form the lower triangular Cholesky factorization

$$\mathbf{B}_{h,m} = \mathbf{C}_{h,m} \mathbf{C}_{h,m}',$$

where the Cholesky factor matrix $\mathbf{C}_{h,m}$ is of dimension $p+1$. Standard results then show that (writing $c_{j,j}$ for the jth diagonal element of $\mathbf{C}_{h,m}$)

$$|\mathbf{A}_{h,m}| = \prod_{j=2}^{p+1} c_{j,j}^2.$$

The attraction of the Cholesky factorization is that adding a new data vector $\boldsymbol{x}_{m+1}$, corresponding $\mathbf{C}_{h,m+1}$ can be computed from $\mathbf{C}_{h,m}$ and $\boldsymbol{z}_{m+1}$ with a fast, stable update (see, for example, [9]).

This leads to the outline algorithm: For each $h = 1$ to $n - p$, initialize $\mathbf{C}_{h,h-1} = 0$. Then, for $m = h$ to n, calculate $\mathbf{C}_{h,m}$ from $\mathbf{C}_{h,m-1}$ and $\boldsymbol{z}_m$ using the update. Once enough observations are included to achieve nonsingularity, use the diagonal elements of $\mathbf{C}_{h,m}$ to compute $\log |A_{h,m}|$ and from this $Q^*(h, m)$.

Fortran 95 code implementing this update in a different context can be found in Hawkins [22], and an R code is given in Maboudou-Tchao [28].

3.2 *DP solution*

The DP recursion is

(1) $L(1, m) = Q^*(0, m), \quad m = p+1, p+2, \ldots, n$
(2) $L(k, m) = \min_h L(k-1, h) + Q^*(h, m), \quad m = p(k-1)+1, \ldots, n.$

Each segment must contain at least $p+1$ observations to ensure that the covariance matrix is full rank. Failure of this requirement will lead to indeterminacy in that Q^* is of the form $\infty - \infty$. To indicate the infeasibility of such solutions, we set $L(k, h) = \infty$ wherever a feasible segmentation is impossible.

One attraction of DP solutions is the wealth of additional information they provide, just for the effort of harvesting it. The algorithm aims at the problem of fitting k change-points to the sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, but a byproduct of this calculation is the optimal change-points for fitting any smaller number of segments to any leading subsequence of the data. Recall that $L(r, m)$ gives the optimum r-segment split of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$, and that $H(r, m)$ gives the rightmost optimal change-point for this subsequence. Working backward from this, the rest of the optimal change-points for an r-segment segmentation to these first m observations is obtained by back-tracing.

4. Number of change-points

To be useful, the methodology also needs a rule for deciding how many segments are needed to model the data. While the objective is a maximized likelihood, and is in principle amenable

to likelihood ratio testing, the problem does not satisfy Cramer regularity, and so conventional asymptotic approaches are inapplicable. Several methods to select the number of change-points are available in the literature. Yao [42] proposed a method using the Schwarz information criterion (SIC) where the number of change-points k is chosen to minimize a penalized likelihood function. Other penalized likelihood methods are available. Several proposals [4,16,37] use a penalty function that depends on a shrinkage parameter that must be chosen by the user through cross-validation. Lavielle [27] gave a data-adaptive way of choosing the penalty term.

We propose using the SIC with penalty term $p(p+3)(k-1)\log(n)/2$ suggested by Chen and Gupta [10]. This approach gives a consistent estimate of the order of the true model. Applying SIC to our problem yields the criterion

$$\mathrm{SIC}(k) = np(\log(2\pi)+1) + \sum_{j=1}^{k}(Q(\tau_{j-1},\tau_j) - g(r_j,p)) + \frac{p(p+3)(k-1)}{2}\log(n). \quad (9)$$

For each fixed k up to some maximal value K, we compute the SIC using Equation (9) and select the number of change points as the value k^* that minimizes $\mathrm{SIC}(k)$.

5. Diagnosis after the detection of the change-points

Once the change-points are located, a follow-up diagnosis is needed to understand the origins of the change observed. A method to explain the changes detected is proposed here. Assume that we have decided to use $k-1$ change-points and so k segments and consider segment i. It is clear that once change-points have been identified, changes may come from the mean vector or the covariance matrix or from a combination of both. A first step is to use a multivariate approach to check with a broad brush-stroke whether there are differences in the mean vectors, and whether there are differences in the covariance matrices between neighboring segments. To do this:

(1) Compare the mean vectors of segments i and $i+1$, testing

$$H_0 : \boldsymbol{\mu}_i = \boldsymbol{\mu}_{i+1} \quad vs \quad H_a : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i+1}.$$

There is no guarantee that $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i+1}$, so an appropriate test statistic is

$$T_i^2 = (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i+1})' \left(\frac{1}{r_i}\mathbf{S}_i + \frac{1}{r_{i+1}}\mathbf{S}_{i+1}\right)^{-1} (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i+1}), \quad (10)$$

where $\bar{\boldsymbol{x}}_i$ and $\mathbf{S}_i$ are the estimated mean vector and covariance matrix of segment i, respectively. T_i^2 has an approximate χ_p^2 when the null hypothesis is true. A large value for T_i^2 gives us the evidence of a change in the mean vectors between segments i and $i+1$.

(2) Also, compare the covariance matrices of segments i and $i+1$ by performing the test

$$H_0 : \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i+1} \quad \text{vs.} \quad H_a : \boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_{i+1}.$$

For this test, we suggest Box's M test, defined by

$$\log M_i = \tfrac{1}{2}\left[\nu_i \log|\mathbf{S}_i| + \nu_{i+1}\log|\mathbf{S}_{i+1}|\right] - \tfrac{1}{2}(\nu_i+\nu_{i+1})\log|\mathbf{S}_p|, \quad (11)$$

where $\nu_i = r_i - 1$ and $\mathbf{S}_p = (1/(r_i + r_{i+1} - 2))(\mathbf{S}_i + \mathbf{S}_{i+1})$ is the estimated pooled covariance matrix between segments i and $i+1$.

Calculate a Bartlett correction factor c_1 by

$$c_1 = \left[\frac{1}{\nu_i} + \frac{1}{\nu_{i+1}} - \frac{1}{\nu_i + \nu_{i+1}}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)}\right].$$

Then,

$$u_i = -2(1 - c_1)\log M_i \tag{12}$$

has an approximate $\chi^2_{(1/2)p(p+1)}$.

Sometimes one of these test statistics is large, but not the other, and this then leads one to diagnose a shift in the mean vector or in the covariance matrix. Sometimes both test statistics are large, leading one to diagnose a shift in both the mean vector and the covariance matrix. And sometimes, neither test statistic is particularly large, leading to the disappointing conclusion that there is no clear diagnosis of the separation between those neighboring segments to be made.

Turning to the more satisfying situation that there is a clear indication of a specific type of shift, the next step is to highlight what variables are involved and in what way.

The simplest tests are marginal tests on the individual means, and on the individual variances. For the marginal means, write $\mu_{i,j}$ for the mean of the ith segment of variable j, then to compare the means, we compute a two-sample t-test between the segments i and $i + 1$ using Satterthwaite's approximation. This is defined as

$$H_0 : \mu_{i,j} = \mu_{i+1,j} \quad \text{vs.} \quad H_a : \mu_{i,j} \neq \mu_{i+1,j}.$$

The test is achieved using the statistic $T_{i,j}$ defined as:

$$T_{i,j} = \frac{\hat{\mu}_{i+1,j} - \hat{\mu}_{i,j}}{\sqrt{s^2_{i+1,j}/r_{i+1} + s^2_{i,j}/r_i}}, \tag{13}$$

where $\hat{\mu}_{i,j}$ represents the estimated mean of segment i of variable j and $s^2_{i,j}$ is the estimated variance of the ith segment of variable j.

We can get an idea of the significance of the statistic $T_{i,j}$ using the Satterthwaite approximate t-distribution with ν degrees of freedom under H_0, where

$$\nu = \frac{(s^2_{i+1,j}/r_{i+1} + s^2_{i,j}/r_i)^2}{(s^2_{i+1,j})^2/r^2_{i+1}(r_{i+1} - 1) + (s^2_{i,j})^2/r^2_i(r_i - 1)}.$$

If the preliminary screen suggests a difference in the covariance matrices, then the corresponding marginal diagnostic for a shift in the component variances is an F ratio. Defining $\sigma_{i,j}$ as the standard deviation of the ith segment of variable j, to compare the standard deviations, we use the following test

$$H_0 : \sigma^2_{i,j} = \sigma^2_{i+1,j} \quad \text{vs.} \quad H_a : \sigma^2_{i,j} \neq \sigma^2_{i+1,j}.$$

So, the test statistic is

$$F_{i,j} = \frac{s^2_{i+1,j}}{s^2_{i,j}}, \tag{14}$$

which we can assess against the F-distribution with $r_{i+1} - 1$ numerator and $r_i - 1$ denominator degrees of freedom under the null hypothesis.

Note that as these follow-up tests are constructed only after we pass the preliminary screen, the t and F reference distributions are no more than guidelines and do not give conventional formal significance tests.

Turning to the third possibility of changes in the correlations, let $\rho_{i,j,l}$ be the correlation between variables j and l for segment i, $j \neq l$ and $\hat{\rho}_{i,j,l}$ be the estimated correlation between variables j and l for segment $i, j \neq l$. To compare the correlations of segments i and $i+1$, we use Fisher's z-transformation of the correlation $\hat{\rho}$, $z_{i,j,l} = \frac{1}{2}\log(|(1+\hat{\rho}_{i,j,l})/(1-\hat{\rho}_{i,j,l})|)$. We perform the following hypothesis testing:

$$H_0 : \rho_{i,j,l} = \rho_{i+1,j,l} \quad \text{vs.} \quad H_a : \rho_{i,j,l} \neq \rho_{i+1,j,l}.$$

The test statistic is given by

$$Z_{i,j,l} = \frac{z_{i+1,j,l} - z_{i,j,l}}{\sqrt{1/(r_{i+1}-3) + 1/(r_i - 3)}}, \tag{15}$$

which can be judged against the standard normal distribution. These marginal parameter tests are the simplest. More complex step-down tests along the lines of the Mason–Tracy–Young [29] approach can also be used, particularly if the elements of the data vectors $\boldsymbol{x}$ have some natural ordering.

6. Simulation study

6.1 *Invariance under linear transformation*

On the face of it, there are many parameters that might affect the segmentation, but in fact this is only partially true. The approach is affine invariant and gives the same results even if an arbitrary non-singular linear transformation of the data is made. To show this in general terms, let $\boldsymbol{\Gamma}$ be an arbitrary $p \times p$ nonsingular matrix and $\mathbf{a}$ an arbitrary p-component vector, and let $\boldsymbol{y}_i = \boldsymbol{\Gamma}\boldsymbol{x}_i + \mathbf{a}$. Then, the change-points produced by segmenting the sequence $\boldsymbol{y}_i$ are identical to those produced by segmenting the sequence $\boldsymbol{x}_i$, and the maximized likelihood is identical apart from an irrelevant constant multiplier.

THEOREM 2 *Under the assumptions 1–3 on* $\boldsymbol{y}_i$, $\boldsymbol{y}_i = \boldsymbol{\Gamma}\boldsymbol{x}_i + \boldsymbol{a}$ *and* $\boldsymbol{x}_i$ *have the same change-point locations.*

Proof Under $H_0, \boldsymbol{y} \sim N(\boldsymbol{\Gamma}\boldsymbol{\mu} + \mathbf{a}, \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}')$, so the log likelihood of the sequence $\boldsymbol{y}_i, i = 1, \ldots, n$ is

$$\begin{aligned}\log L_{y0}(\boldsymbol{\theta}) &= -\frac{np}{2}\log(2\pi) - n\log|\boldsymbol{\Gamma}| - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) \\ &= \log L_0(\boldsymbol{\theta}) - n\log|\boldsymbol{\Gamma}|,\end{aligned}$$

where $\log L_0(\boldsymbol{\theta})$ is the log-likelihood of $\boldsymbol{x}$ under the null hypothesis.

Under H_a, assumption 3 means

$$\tau_{j-1} < i \leq \tau_j, \quad \boldsymbol{y}_i \sim N(\boldsymbol{\Gamma}\boldsymbol{\mu}_j + \mathbf{a}, \boldsymbol{\Gamma}\boldsymbol{\Sigma}_j\boldsymbol{\Gamma}')$$

(all $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown). Therefore, the log likelihood is

$$\begin{aligned}\log L_{y1}(\boldsymbol{\theta}) &= -\frac{np}{2}\log(2\pi) - n\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{j=1}^{k} r_j \log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{k}\sum_{i=\tau_{j-1}+1}^{\tau_j}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_j) \\ &= \log L_1(\boldsymbol{\theta}) - n\log|\boldsymbol{\Gamma}|,\end{aligned}$$

where $\log L_1(\boldsymbol{\theta})$ is the log-likelihood of $\boldsymbol{x}$ under H_a.

As the location of the change-points does not depend on the constant term, then $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ have the same change-point locations. ■

Therefore, neither the overall mean vector, nor the variances, nor the correlations are relevant per se. What is relevant is the relative eigenvalues between successive segments in the data, and the differences in the mean vectors expressed in multistandardized units.

6.2 *Numerical experiments with simulated data*

Since the means, variances, and correlations are not directly relevant, we explore the proposed methodology using the standard multivariate normal $N(\mathbf{0}, \mathbf{I}_p)$ with $p = 3$. Therefore, under the null hypothesis, $\boldsymbol{x} \sim N(\mathbf{0}, \mathbf{I}_3)$.

Three main changes were investigated. In the first scenario ('mean change'), we generate $n = 125$ observations with change-points in the mean vector at time $\tau_1 = 30$, $\tau_2 = 65$, and $\tau_3 = 100$ (4 segments) according to the following scheme.

$\boldsymbol{\mu}_1' = (0, 0, 0)$ from observations 1 to 30,
$\boldsymbol{\mu}_2' = (0, \mu_0, 0)$ from observations 31 to 65,
$\boldsymbol{\mu}_3' = (0, \mu_0, \mu_0)$ from observations 66 to 100,
and $\boldsymbol{\mu}_4' = (-\mu_0, \mu_0, \mu_0)$ from observations 101 to 125.

The covariance matrix was unchanged at $\boldsymbol{\Sigma} = \mathbf{I}_p$.

In the second scenario ('variance change'), we again generate $n = 125$ observations with change-points in the covariance matrix at time $\tau_1 = 30$, $\tau_2 = 65$, and $\tau_3 = 100$ using the following scheme.

We changed the covariance matrix from its initial value of $\mathbf{I}_p$ to a matrix having δ in the (1,1) position with the other elements unchanged from observations 1 to 30, δ in the (1,1) and (2,2) positions with the other elements unchanged from observations 31 to 65, δ in the (1,1), (2,2), and (3,3) positions with the other elements unchanged from observations 66 to 100, and 0 in the (1,1) position and δ in the (2,2) and (3,3) positions with the other elements unchanged from observations 101 to 125.

In this scenario, we left the mean vector at $\mathbf{0}$.

The third scenario ('simultaneous mean and variance change') combine the first and the second scenarios.

For the 'mean change', we set $\mu_0 = \mu + k\sigma$. In our study, $\mu = 0$, $\sigma = 1$. We investigated the different scenarios:

(1) S1: Mean change with $k = 2$,
(2) S2: Mean change with $k = 3$,
(3) S3: Variance change with $\delta = 5$,
(4) S4: Variance change with $\delta = 10$,
(5) S5: Mean and variance changes with $k = 3$ and $\delta = 5$.

We then applied the multiple change-point detection procedure to 1000 independent replicates and the estimated locations of the change-points are obtained by averaging over the 1000 replicates. The results are given in Table 1.

From the results from Table 1, the proposed method is doing a very good job in detecting the location of the change-points. Also, we can see that the performance of the methodology proposed depends on the type of change but the estimated change-point locations are close to the true change-point locations. Note that the estimated change-point locations for the mean change are slightly

Table 1. Estimated change-point location.

Case	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$
S1	30.30(4.44)	64.96(5.05)	99.31(5.39)
S2	30.04(1.24)	65.05(1.27)	99.98(1.00)
S3	35.54(15.37)	66.62(15.36)	93.36(14.23)
S4	33.49(10.63)	65.61(12.96)	96.20(11.37)
S5	30.38(2.46)	65.19(3.32)	99.51(3.69)

Note: Standard errors are given within parentheses.

different from the one obtained with the covariance change. This result should be expected. For the mean change, the covariance matrix is the identity matrix and it remains unchanged through all the segments. For the variance change, the covariance matrix changes through all the segments. Note that the correlations are unchanged. However, because of the different change structures in the covariance matrix, the eigenvalues also change through the different segments. Consequently, the estimated change-point locations for the mean change will only differ from the covariance change.

6.3 *Comparison with the binary segmentation procedure*

It is of interest to compare our method with the hierarchic binary segmentation (BS) procedure of Chen and Gupta [10], which we did using two scenarios – 'mean change' and 'simultaneous mean and variance change' – both with $n = 100$ and 3 true segments. For the mean change, we generate $n = 100$ points with double change-points in the mean vector at time $\tau_1 = 40$ and $\tau_2 = 70$. We set $\boldsymbol{\mu}_1' = (0, 0, 0)$ from observations 1 to 40, $\boldsymbol{\mu}_2' = (0, \mu_0, 0)$ from observations 41 to 70, and $\boldsymbol{\mu}_3' = (0, \mu_0, \mu_0)$ from observations 71 to 100, leaving the covariance matrix unchanged at $\boldsymbol{\Sigma} = \mathbf{I}_p$. For the 'simultaneous mean and variance change', we combined a variance change with the mean change. For the variance change, we generate $n = 100$ observations with double change-points in the covariance matrix at time $\tau_1 = 40$ and $\tau_2 = 70$ using the following scheme. We changed the covariance matrix from its initial value of $\mathbf{I}_p$ to a matrix having δ in the (1,1) position with the other elements unchanged from observations 1 to 40, δ in the (1,1) and (2,2) positions with the other elements unchanged from observations 41 to 70, and δ in the (1,1), (2,2), and (3,3) positions with the other elements unchanged from observations 71 to 100, and that leaves the mean vector at $\mathbf{0}$.

For the 'mean change', we set $\mu_0 = \mu + k\sigma$. We investigated the different scenarios:

(1) S1: Mean change with $k = 2$,
(2) S2: Mean change with $k = 3$,
(3) S3: Mean and variance changes with $k = 3$ and $\delta = 5$.

We apply our methodology and the binary splitting algorithm of Chen and Gupta [10] to 1000 independent replicates and the estimated locations of the change-points are obtained by just averaging over the 1000 replicates. The results are available in Table 2. 'DP' represents our procedure while 'BS' is used for binary segmentation procedure.

In all our simulated samples, our method identified the correct number of segments, but this was not true of repeated BS, which on occasion terminated with fewer, or with more segments than were actually present. These cases in which BS failed to identify the correct number of components have been omitted from the calculations of Table 2 and so this table paints a somewhat rosier picture of BS than it should. Despite this, however, the table shows that our method produces considerably more accurate estimates of the true change-points.

Table 2. Comparison between DP and BS – estimated change-point location.

	Case	$\hat{\tau}_1$	$\hat{\tau}_2$
DP	S1	40.06(3.57)	69.86(1.98)
	S2	40.00(0.56)	70.03(0.65)
	S3	40.29(2.07)	70.29(2.01)
BS	S1	38.97(5.62)	68.95(6.25)
	S2	40.10(0.87)	70.06(1.03)
	S3	41.01(2.45)	69.11(4.08)

Note: Standard errors are given within parentheses.

7. Example

7.1 *Presentation of the ambulatory monitoring data set*

The data set used is from a long-standing research project in ambulatory monitoring (see http://www.msi.umn.edu/halberg for deeper background). In this work, subjects are equipped with instruments that measure and record physiological variables. The wearer's blood pressure and heart rate were measured and recorded every 15 minutes for 6 years. Each week's raw data are condensed into weekly summary numbers which include:

- SBP: A mean systolic blood pressure,
- DBP: A mean diastolic blood pressure,
- HR: A mean of heart rate,
- MAP: A mean arterial pressure.

We want to investigate if all the observations follow the same statistical distribution or if there are some change-points where the data follow different statistical distributions. Our methodology is well suited to answer this question.

7.2 *Results*

We execute the segmentation fitting $K = 10$ segments, obtaining as by-products the optimal change-points for segmentations into fewer than 10 segments. The results of our fit are shown in Table 3.

In this table, column 1 represents the number of segments k. For example, $k = 1$ means one segment for no change-point in the data. $k = 2$ means two segments for a single change-point. Column 2 represents SIC(k) and column 3 shows the estimated change-points. For example, the 4 segment model has SIC = 5313.45, and has optimal change-points $\hat{\tau} = (51, 189, 323)$. Subsegment 1 consists of observations 1 to 51, segment 2 from 52 to 189, segment 3 from 190 to 323, and segment 4 from 324 to 360. The optimal change-points are not always hierarchical, pointing to situations in which sequential BS method would fail. The minimum SIC, 5293.82, occurs for $k = 3$, indicating that the series contains three natural segments – 1 to 51, 52 to 189, and 190 to 360.

From the results described earlier, some possible comments can be made:

The SIC suggests a two change-point model for this data (SIC(k) attains its minimum for $k = 3$, cf. Table 3) and the change-points are located at cases 51 and 189. The estimation portion is summarized below. The means and standard deviations of the segments are displayed in Table 4,

Table 3. Estimated change-points with all data.

k	SIC(k)	$\hat{\tau}$
1	6454.75	No change
2	5418.95	51
3	5293.82	51, 189
4	5313.45	51, 189, 323
5	5340.96	51, 189, 303, 323
6	5374.41	51, 171, 235, 289, 304
7	5405.31	51, 149, 209, 214, 289, 304
8	5446.05	51, 149, 209, 214, 301, 323, 346
9	5488.01	51, 149, 209, 214, 299, 306, 324, 346
10	5530.24	51, 149, 209, 214, 284, 289, 306, 324, 346

Table 4. Estimated means and standard deviations.

	Segment	SBP	DBP	HR	MAP
Mean	1	125.55	79.04	82.35	94.21
	2	128.17	77.40	83.58	98.36
	3	126.40	77.01	80.10	97.34
SD	1	3.26	3.13	2.86	2.74
	2	2.84	2.05	2.60	2.44
	3	4.02	2.59	2.98	3.47

Table 5. Correlation matrices for the three segments.

$$\hat{\rho}_1 = \begin{pmatrix} 1.00 & 0.43 & 0.45 & 0.73 \\ 0.43 & 1.00 & 0.55 & 0.93 \\ 0.45 & 0.55 & 1.00 & 0.59 \\ 0.73 & 0.93 & 0.59 & 1.00 \end{pmatrix} \quad \hat{\rho}_2 = \begin{pmatrix} 1.00 & 0.92 & 0.49 & 0.95 \\ 0.92 & 1.00 & 0.48 & 0.95 \\ 0.49 & 0.48 & 1.00 & 0.52 \\ 0.95 & 0.95 & 0.52 & 1.00 \end{pmatrix}$$

$$\hat{\rho}_3 = \begin{pmatrix} 1.00 & 0.92 & 0.57 & 0.94 \\ 0.92 & 1.00 & 0.56 & 0.94 \\ 0.57 & 0.56 & 1.00 & 0.64 \\ 0.94 & 0.94 & 0.64 & 1.00 \end{pmatrix}$$

while the correlations for the segments are given in Table 5. Next, we run a diagnosis after the detection of the change-points. We applied the method outlined earlier for the diagnosis step. For this data set, $k = 3$ so $i = 1, 2$. As we have four variables, then $j = 1, \ldots, 4$.

First, we apply multivariate follow-up testing (shown in Table 6) to see whether each neighboring pair of segments differ perceptibly in mean vector, in covariance matrix, or both. T^2 is calculated using Equation (10) and detects change in the mean vectors while Box's M statistic u (Equation (12)) tests for change in the covariance matrices. The results suggest that there are changes in both the mean vectors and covariance matrices between segments 1 and 2 and also between segments 2 and 3. This justifies our claim that there were two change-points in this data and our proposed methodology locate these change-points.

As with any multivariate methods, T^2 and Box's M tests do not provide information on the variables responsible for the changes observed. Then, we use the second part of the diagnosis analysis proposed. The results are reported in Tables 7 and 8.

Table 6. Segment mean vectors and covariance matrices comparison.

	T^2	Box's M
1 vs. 2	7336.50 (0.000)	363.15 (0.000)
2 vs. 3	193.52 (0.000)	69.33 (0.000)

Note: The p-values are given within parentheses.

Table 7. Segment means and standard deviations comparison.

		SBP	DBP	HR	MAP
Mean	1 vs. 2	5.07 (0.000)	−3.49 (0.008)	2.69 (0.008)	9.49 (0.000)
	2 vs. 3	−4.54 (0.000)	−1.45 (0.146)	−10.95 (0.000)	−3.01 (0.002)
SD	1 vs. 2	1.32 (0.208)	2.32 (0.000)	1.21 (0.389)	1.26 (0.304)
	2 vs. 3	2.01 (0.000)	1.59 (0.005)	1.32 (0.093)	2.02 (0.000)

Note: The p-values are given within parentheses.

Table 8. Segment correlation matrices comparison.

$$\boldsymbol{Z}_1 = \begin{pmatrix} 6.77(0.000) & 0.342(0.731) & 5.218(0.000) \\ & -0.573(0.566) & 1.267(0.204) \\ & & -0.617(0.537) \end{pmatrix} \qquad \boldsymbol{Z}_2 = \begin{pmatrix} -0.033(0.973) & 0.912(0.361) & -0.447(0.654) \\ & 1.022(0.306) & -1.352(0.176) \\ & & 1.537(0.124) \end{pmatrix}$$

Note: The p-values are given within parentheses.

The values listed in the table are the Satterthwaite t-statistics to assess the significance of differences in each component of the mean vector and the F ratio to assess that of differences in the marginal variances. Also shown are the nominal P values (in parentheses) of these test statistics.

Comparing segments 1 and 2

- The means of all four variables appear to have risen significantly. The dimensionless t values show that MAP has the greatest difference between the two segments.
- The standard deviation of DBP shows a significant change, but that of the other variables does not.
- The correlation between SBP and both DBP and MAP is much higher in the second segment than in the first.

Turning to segments 2 and 3.

- The means of all measures except DPB have dropped significantly.
- The standard deviations of all except HR have risen significantly.
- The correlations do not appear to have changed.

In short, the first change-point is caused by a combination of changes in the mean vector, in the variability, and in the relationship between some variables. The second change-point is due primarily to changes in the mean vector.

8. Conclusion

Multiple change-point problems are of interest in different areas of data analysis. This paper presents an effective and fast algorithm to solve the problem when the data can be represented by a Gaussian model. It solves the algorithmic problem of finding the optimal heteroscedastic segmentation, and is helpful for retrospective segmentation of multivariate sequences.

Diagnosis is harder when the number of changes is unknown, as is commonly the case, so we present a criterion for selecting the number of segments. Unlike the repeated hierarchic algorithm, our method finds the global optimum of the likelihood function. The DP algorithm used has a computational complexity linear in the number of segments (though quadratic in the number of points).

In our algorithm, we include a 'Bartlett-style' sample-size correction into the optimization step to reduce the tendency to form unduly short segments. Computational efficiency is achieved using fast numerically stable update of the Cholesky factorization.

References

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1970.

[2] P.P. Balestrassi, A.P. Paiva, A.C. Zambroni de Souza, J.B. Turrioni, and E. Popova, *A multivariate descriptor method for change-point detection in nonlinear time series*, J. Appl. Stat. 38(2) (2011), pp. 327–342.

[3] M. Basseville and N. Nikiforov, *Detection of Abrupt Changes: Theory and Application. Information and System Sciences Series*, Prentice Hall Information, Englewood Cliffs, NJ, 1993.

[4] L. Birge and P. Massart, *Gaussian model selection*, J. Euro. Math. Soc. 3 (2001), pp. 203–268.

[5] K. Bleakley and J.P. Vert, *The group fused Lasso for multiple change-point detection*, Technical report HAL-00602121 (2011).

[6] J.V. Braun and H.-G. Hans-Georg Muller, *Statistical methods for DNA sequence segmentation*, Stat. Sci. 13 (1998), pp. 142–162.

[7] J.V. Braun, R.K. Braun, and H.-G. Muller, *Multiple changepoint fitting via quasi likelihood, with application to DNA sequence segmentation*, Biometrika 87 (2000), pp. 301–314.

[8] B. Brodsky and B. Darkhovsky, *Nonparametric Methods in Change-Point Problems, Mathematics and Its Applications*, Vol 243, Kluwer Academic Publishers, Dordrecht, 1993.

[9] J.M. Chambers, *Regression updating*, J. Amer. Statist. Assoc. 66 (1971), pp. 744–748.

[10] J. Chen and A.K. Gupta, *Parametric Statistical Change-point Analysis*, Birkhauser Verlag, Basel, 2000.

[11] H. Chernoff and S. Zacks, *Estimating the current mean of a normal distribution which is subject to changes in time*, Ann. Math. Stat. 35 (1964), pp. 999–1018.

[12] M. Csörgö, and L. Horvath, *Limit Theorems in Change-Point Analysis*, Wiley, New York, 1997.

[13] B.S. Darkhovski, *Nonparametric methods in change-point problems: A general approach and some concrete algorithms*, IMS Lect. Notes Monogr. Ser. 23 (1994) pp. 99–107.

[14] L.A. Gardner, *On detecting change in the mean of normal variates*, Ann. Math. Stat. 40 (1969), pp. 116–126.

[15] K. Ghosh, S.R. Jammalamadaka, and M. Vasudaven, *Change-point problems for the von Mises distribution*, J. Appl. Stat. 26(4) (1999), pp. 423–434.

[16] C. Gu and J. Wang, 2003. *Penalized likelihood density estimation: Direct cross-validation and scalable approximation*, Statist. Sinica 13 (2003), pp. 811–826.

[17] Z. Harchaoui and C. Levy-Leduc, *Catching change-points with LASSO*, in *Advances in Neural Information Processing Systems 20*, J.C. Platt, D. Koller, Y. Singer and S. Roweis, eds., The MIT Press, Cambridge, MA, 2008, pp. 617–624.

[18] Z. Harchaoui, F. Vallet, A. Lung-Yut-Fong, and O. Cappe, *A Regularized Kernel-based Approach to Unsupervised Audio Segmentation*, in ICASSP '09': Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Washington, DC, IEEE Computer Society, 2009, pp. 1665–1668.

[19] D.M. Hawkins, *Point estimation of the parameters of piecewise regression models*, Appl. Stat. 25 (1976), pp. 51–58.

[20] D.M. Hawkins, *Detecting shifts in functions of multivariate location and covariance parameters*, J. Statist. Plann. Inference 33 (1992), pp. 233–244.

[21] D.M. Hawkins, *Fitting multiple change-point models to data*, Comput. Statist. Data Anal. 37 (2001), pp. 323–341.

[22] D.M. Hawkins and E.M. Maboudou-Tchao, *Self-starting multivariate exponentially weighted moving average control charting for location*, Technometrics 49 (2007), pp. 199–209.

[23] D.M. Hawkins and J.A. Ten Krooden, *Zonation of sequences of heteroscedastic multivariate data*, Comput. Geosci. 5 (1979), pp. 189–194.
[24] N. Itoh and J. Kurths, *Change-Point Detection of Climate Time Series by Nonparametric Method*, in Proceedings of the World Congress on Engineering and Computer Science 2010, Vol. 1, pp. 445–448, San Francisco, 2010.
[25] N.L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York, 1972.
[26] T.L. Lai and H. Xing, *A simple Bayesian approach to multiple change-points*, Statist. Sinica 21 (2011), pp. 539–569.
[27] M. Lavielle, *Using penalized contrasts for the change-point problem*, Signal Process. 85 (2005), pp. 1501–1510.
[28] E.M. Maboudou-Tchao, *Self-starting multivariate exponentially weighted moving average control charting*, Ph.D. thesis, School of Statistics, University of Minnesota, Minneapolis, MN, 2006.
[29] R.L. Mason and J.C. Young, *Multivariate Statistical Process Control with Industrial Applications*, SIAM, Philadelphia, 2002.
[30] M. Modisett and E.M. Maboudou-Tchao, *Significantly lower estimates of volatility arise from the use of open-high-low-close price data*, N. Am. Actuar. J. 14(1) (2010), pp. 68–85.
[31] M. Othus, Y. Li, and R. Tiwari, *Change-point cure models with application to estimating the change-point effect of age of diagnosis among prostate cancer patients*, J. Appl. Stat. 39(4) (2012), pp. 901–911.
[32] E.S. Page, *Continuous inspection schemes*, Biometrika 41 (1954), pp. 100–115.
[33] E.S. Page, *A test for a chance in a parameter occurring at an unknown point*, Biometrika 42 (1955), pp. 523–527.
[34] A. Sengupta and A.K. Laha, *A Bayesian analysis of the change-point problem for directional data*, J. Appl. Stat. 35(6) (2008), pp. 693–700.
[35] M.S. Srivastava and K.J. Worsley, *Likelihood ratio tests for a change in the multivariate normal mean*, J. Amer. Statist. Assoc. 81 (1986), p. 199–204.
[36] G.M. Terrera, A. van den Hout, and F.E. Matthews, *Random change point models: Investigating cognitive decline in the presence of missing data*, J. Appl. Stat. 38(4) (2011), pp. 705–716
[37] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B 58 (1986), pp. 267–288.
[38] J.H. Venter and S.J. Steel, *Finding multiple abrupt change points*, Comput. Statist. Data Anal. 22 (1996), pp. 481–504
[39] J.P. Vert and K. Bleakley, *Fast detection of multiple change-points shared by many signals using group LARS*, I, Adv. Neural Inform. Process. Syst. 22 (2010), pp. 2343–2352
[40] L.J. Vostrikova, *Detecting disorder in multidimensional random processes*, Soviet Math. Dokl 24 (1981), pp. 55–59.
[41] K.J. Worsley, *An improved Bonferroni inequality and applications*, Biometrika 69 (1982), pp. 297–302.
[42] Y.C. Yao, *Estimating the number of change-points via Schwarz's criterion*, Stat. Probab. Lett. 6 (1988), pp. 181–189.
[43] N.R. Zhang, D.O. Siegmund, H. Ji, and J. Li, *Detecting simultaneous change-points in multiple sequences*, Biometrika 97(3) (2010), pp. 631–645.