Product partition latent variable model for multiple change-point detection in multivariate data

Gift Nyamundanda*, Avril Hegarty and Kevin Hayes

Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

(Received 22 April 2014; accepted 11 March 2015)

The product partition model (PPM) is a well-established efficient statistical method for detecting multiple change points in time-evolving univariate data. In this article, we refine the PPM for the purpose of detecting multiple change points in correlated multivariate time-evolving data. Our model detects distributional changes in both the mean and covariance structures of multivariate Gaussian data by exploiting a smaller dimensional representation of correlated multiple time series. The utility of the proposed method is demonstrated through experiments on simulated and real datasets.

Keywords: PPM; PPLVM; dimensionality reduction; multivariate; Gaussian

1. Introduction

Multiple change-point analysis is a statistical methodology for identifying and locating positions or times at which an underlying process generating sequenced data undergoes abrupt local changes. These models have been deployed in a variety of scientific and industrial contexts. In economics and finance, change-point models have been used to detect changes in the volatilities of returns [6,23]. In bioinformatics, Erdman and Emerson [7] used Bayesian change-point analysis to detect regions of genetic alteration identifying genes involved in cancer progression. These models are particularly useful in industrial control settings where the underlying process generating the observed data changes, or parameters measuring certain quality aspects shift over time [22]. Other recent areas of application include climate change detection [28], solar flare imagery [14], and network traffic patterns [18].

Multiple change-point analysis presents many estimation challenges: simultaneously determining both the number of change points and their locations, as well as selecting an appropriate model for data between change points, is not straightforward. Another level of complexity is added when the change points occur on multiple correlated time series, that is, in an industrial

*Corresponding author. Email: gnyamundanda@gmail.com
setting where it is required to determine endpoints of different stages of a time-evolving pharmaceutical manufacturing process [4]. This is particularly difficult, when the number of variables or time series is considerably larger than the number of time points measured, as in a microarray setting or social network interactions [27]. In these situations, the application of standard parametric statistical methods is confounded by insufficient data for parameter estimation. Hence, there is a genuine need for dimension reduction techniques which account for the covariance between several time series in order to efficiently identify and estimate change points in multivariate time series data.

Although considerable attention has been devoted to change-point models in the recent literature [7,16,17,22,29,33], there is no standard method for modeling multivariate time series data with multiple change points. Markov chain Monte Carlo (MCMC) techniques were used by Erdman and Emerson [7] for Bayesian multiple change-point analysis of univariate to multivariate time series. This approach ignores dependencies between multiple time series such that any changes on the full joint distribution of the multivariate time series will not be identified. In similar vein, the R packages change-point [15–17] and cpm [29] provide suites of parametric and non-parametric methods for multiple change-point analysis of univariate series. Bayesian change-point detection techniques of Fearnhead [8,9] were used by Xuan and Murphy [33] to perform multiple change-point analysis on multivariate time series data. Although this approach appropriately models the covariance structure of the multivariate time series, it breaks down, due to singularity problems, when the dimension of the series is increased, whilst the length of time series measured is constant. Recently, Maboudou-Tchaoa and Hawkins [22] employed a maximum-likelihood estimation approach coupled with a dynamic programming algorithm to detect multiple change points in multivariate time series. However, this approach cannot be easily extended to detect multiple change points in a high-dimensional setting.

In this article, we present a full Bayesian approach to the problem of multiple change-point detection in multivariate data known as product partition latent variable model (PPLVM). The PPLVM can be used to detect distributional changes in the mean and covariance, even in high-dimensional settings, whilst exposing underlying relationships in the series. The PPLVM draws on two different types of models: the product partition model (PPM) [2], which identifies change points by imposing a partition structure on parameters of interest, and the probabilistic principal component analysis (PPCA) model [24,25,31], which reduces data dimension whilst modeling the covariance structure. Perhaps the key benefit of the PPM over other change-point models is that it assumes that both the number of change points and their locations are random variables rather than considering them as fixed and known in advance. This model was initially used by Barry and Hartigan [3] to detect changes in the means from a Gaussian model. However, their approach does not fully explore the posterior distribution as the algorithm only collects posterior means. Loschi and Cruz [20] modified the Barry and Hartigan algorithm to extend the PPM to apply a Gibbs sampling scheme to estimate the posterior distribution for the change points, number of change points, and the posterior probability that each position is a change point. The PPLVM described here exploits these results by Losch and extends the PPM to detect changes in the full joint distribution of a multivariate time series whilst providing a smaller dimensional representation of the series.

The remainder of the article is structured as follows. In Section 2, the PPM, first introduced by Barry and Hartigan [2], is briefly reviewed. This is followed by a presentation of the PPLVM and how it can be used to identify distributional changes in the mean and covariance. The PPLVM is estimated within the Bayesian framework. Section 2 is a discussion of the necessary prior distributions and a brief description of the use of MCMC techniques [10] to fit the PPLVM. In Section 3, we assess the performance of the PPLVM, compared to other approaches to change-point problems for multivariate data, on several simulated and real-world examples. Finally, Section 4 is a discussion of the proposed model and further avenues of research.
2. Change-point models

Let \( x_1, \ldots, x_n \) be a sequence of observations measured over time which are assumed to be independent conditional on the parameter \( \theta_i \), with a conditional marginal density \( f(x_i | \theta_i) \), where \( i \in l = \{1, \ldots, n\} \) is a set of ordered time points. The PPM, due to \[2\], assumes that there exist a random partition \( \rho = \{c_1, \ldots, c_b\} \) which partitions a set \( l \) into temporal blocks \( c_k = \{i_{k-1} + 1, \ldots, i_k\} \), where \( k = 1, \ldots, b \) and \( i_k \in l \). Instances or positions at which changes in the distribution of \( x = (x_1, \ldots, x_n)^T \) occur identify change points and are constrained to be \( 0 = i_0 < i_1 < \ldots < i_b = n \).

The prior probability of partition \( \rho \), which partitions data \( x \) into \( b \) blocks, can be expressed as

\[
P(\rho = \{c_1, \ldots, c_b\}) = K \prod_{k=1}^{b} E_{c_k},
\]

where \( E_{c_k} \) is the prior cohesion associated with block \( k \), representing the degree of similarity amongst observations within block \( x_{c_k} = (x_{i_{k-1}+1}, \ldots, x_{i_k}) \). The constant \( K \) is chosen so that the sum of the probabilities over all partitions is one. The PPM treats the number of partitions \( b \) as a random variable \( B \) with values 1 to \( n \). Given the partition \( \rho = \{c_1, \ldots, c_b\} \), the sequence of parameters \( \theta_1, \ldots, \theta_b \) can be partitioned into independent common parameters \( \theta_{c_1}, \ldots, \theta_{c_b} \), where \( \theta_{c_k} = \theta_{i_k-1+1} = \ldots = \theta_{i_k} \).

2.1 Product partition latent variable model

Suppose \( x_i = (x_{i1}, \ldots, x_{im})^T \) is a set of measurements taken on \( m \) correlated variables on time point \( i \). A generative statistical model, the PPLVM, closely related to the PPCA model \[24,25,31\], can be used to model a high-dimensional observed data point \( x_i \) as a linear function of a corresponding low-dimensional latent variable \( u_i = (u_{i1}, \ldots, u_{iq})^T \), \( q \ll m \). Under the PPLVM, the observed data point \( x_i \) can be generated through the generative model

\[
x_i = Wu_i + \mu_i + \epsilon_i,
\]

where \( W \) is a \( m \times q \) transformation matrix, which relates the latent variables to the observed variables, and \( \mu_i \) is the mean of the data at time point \( i \). The noise \( \epsilon_i \) and the latent variables \( u_i \) are assumed to be independently Gaussian-distributed random vectors, \( \text{MVN}_m[0, \sigma_i^2I] \) and \( \text{MVN}_q[0, I] \), respectively, where \( I \) is an identity matrix.

Unlike the PPCA model, the PPLVM presented here allows the mean \( \mu_i \) and the noise covariance parameter \( \sigma_i^2 \) to vary slowly with time. The PPM introduced in Section 2 is incorporated into the PPLVM to model time dependencies within the data by imposing a random partition structure \( \rho \) on the means and the noise covariance parameters, \( \theta_1 = (\mu_1, \sigma_1^2), \ldots, \theta_n = (\mu_n, \sigma_n^2) \). This partition structure \( \rho \) identifies blocks of data \( x_{c_k} \) which share the same values of parameter \( \theta_k \) within the same block, but heterogeneous between. Hence, the PPLVM is a clustering approach which assumes that the process generating data can be split into temporal blocks in which the process is constant. Under this model, conditional on the partition \( \rho = \{c_1, \ldots, c_b\} \) and the latent variables \( u = (u_1, \ldots, u_n) \), the data \( x = (x_1, \ldots, x_n) \) have a joint distribution given as

\[
P(x|u, W, \theta, \rho, b) = \prod_{k=1}^{b} \prod_{i \in c_k} \text{MVN}_m[x_i | Wu_i + \mu_k, \sigma_k^2I],
\]

where \( \theta = (\theta_1, \ldots, \theta_b) \) and \( b \) is the number partitions.
2.2 Estimation of the PPLVM

Parameters of the PPLVM can be estimated by conditioning on the partition and then averaging over all possible partitions. However, as the size of the series \( n \) increases, the number of possible partitions increases exponentially such that it becomes difficult to compute all partitions. Hence, Bayesian approximation methods are required which can search through the space of all possible partitions for values of \( \rho \) with high posterior probability. This can be carried out by replacing the random partitions \( \rho \) with a binary latent variable \( z = (z_1, \ldots, z_{n-1}) \) suggested by Barry and Hartigan [3], where \( z_i = 1 \) if \( i \) is a change point, \( \theta_i \neq \theta_{i+1} \), and zero otherwise. Estimating the random vector \( z \) is equivalent to estimating the random partition \( \rho \). MCMC techniques such as Gibbs sampling [12] can be used to sample each \( z_i \) from its full conditional distribution. A Bayesian approach requires the specification of prior distributions for all the model parameters. Detailed derivations of the full conditional distributions for \( z \) and all other parameters are given in the additional file.

Under the PPLVM, the parameters of interest include \( W, \mu_1, \ldots, \mu_b \) and \( \sigma_1^2, \ldots, \sigma_b^2, \rho, \) and \( b \). Assuming independent priors on each parameter, the posterior of interest can be expressed as

\[
P(W, \theta, \nu, b, p|x, u) = \prod_{k=1}^b \left[ \prod_{i \in c_k} P(x_i|u_i, \mu_k, \sigma_k^2, W)P(u_i) \right] \\
\times P(\nu P(\beta_k^2)^2 P(\nu P(\beta_k^2)^2 P(W|\nu)P(\nu|p)P(B|p)P(p),
\]

where \( \nu = \text{diag}(\nu_1, \ldots, \nu_n) \) is a hyperparameter associated with prior covariance of \( W \), \( \beta \) is the probability of a change point, and \( P(x_i|u_i, \mu_k, \sigma_k^2, W) \) is the likelihood for block \( c_k \in \rho = \{c_1, \ldots, c_b\} \) in Equation (2).

2.3 Prior distributions

A fully Bayesian approach is taken by specifying independent priors on each of the parameters. A prior distribution on the probability of a change point \( p \) is necessary in order to control the number of change points. A prior distribution with very small values of \( p \) will result in a model that ignores small changes. By contrast, large values of \( p \) will result in oversensitive models that respond to spurious changes through time [32]. A beta distribution, \( P(p) = \text{Beta}[\alpha_p, \beta_p] \), is the most popular prior distribution over \( p \) used in most approaches to multiple change-point problems [19–21]. In case the number of change points is known to be small a priori, small values of \( \alpha \) and large values of \( \beta \) can be specified. This restricts the distribution of \( p \) to be concentrated around small values.

Assuming the binary random variables \( z_i \) are independent Bernoulli distributed and \( P(z_0 = i_0|p) = 1 \), the prior distributions on partition \( \rho \) and on the number of partitions \( B \) given \( p \) are specified, respectively, as

\[
P(\rho = i_0, \ldots, i_n|p) = p^{b-1}(1-p)^{n-b},
\]

\[
P(B = b|p) = C_{b-1}^n p^{b-1}(1-p)^{n-b},
\]

where \( C_{b-1}^n \) is the number of distinct partitions of \( n \) into \( b \) contiguous blocks.

A \( q \)-dimensional multivariate Gaussian prior distribution centered at the origin zero, with a hyperparameter \( \nu \) which controls the prior covariance of the loadings, was considered on the
rows of the loadings matrix, that is,

\[ P(W | v) = \prod_{j=1}^{m} \text{MVN}_q[w_j | \theta, v]. \]

The elements \( v_r, r = 1, \ldots, q \), of the hyperparameter \( v \) have a Gamma hyperprior distribution with parameters \( a_o \) and \( b_o \), that is, \( P(v_r) = G[v_r | a_o/2, b_o/2] \).

We impose the partition structure \( \rho \) on the space of the mean and the noise covariance parameters by assigning prior distributions that identify the clustering structure in the data. Hence, given partition \( \rho \) with \( b \) blocks, \((\mu_1, \sigma^2_1), \ldots, (\mu_b, \sigma^2_b)\) are independent with \((\mu_k, \sigma^2_k)\) having a block prior distribution \( P(\mu_k, \sigma^2_k) \). The choice of the block prior distribution also allows the mean to be related to the noise variance \( \sigma_k^2 \) as

\[ P(\mu_k | \sigma^2_k) = \text{MVN}_m \left[ \mu_k | \mu_o, \frac{\sigma^2_k}{\beta_\mu} I \right], \]

\[ P(\sigma^2_k) = \text{IG} \left[ \sigma^2_k | \frac{\alpha_o^2}{2}, \frac{\beta_o^2}{2} \right], \]

where \( \mu_o \) is the prior mean, \( \beta_\mu \) is a parameter controlling the prior covariance of the mean, and \( \alpha_o^2 \) and \( \beta_o^2 \) control the prior distribution of the noise covariance.

### 2.4 Sampling from the PPLVM: the Gibbs sampler

Given the specified prior distributions, the resulting posterior distribution \( P(W, \theta, v, \rho, b, p|x, u) \) is intricate and MCMC methods are required to produce realizations of the model parameters. The full conditional distributions of all the parameters exist in standard forms and a straightforward Gibbs sampler can be employed to draw samples. Detailed derivations of the full conditional distributions for the PPLVM parameters and latent variables are given in the additional file. Samples can be drawn from the posterior distribution of the PPLVM using the following steps. For \( s = 1, \ldots, S \), the number of iterations or cycles, and starting from an initial value \( z^0 = (z^0_1, \ldots, z^0_{n-1}) \),

1. generate a random partition \( \rho^s \) by sampling elements \( z^s_r, r = 1, \ldots, n - 1 \), of vector \( z^s \) from \( P(z^s_r | x, u^{s-1}, W^{s-1}, z^{s-1}) \),
2. for \( k = 1, \ldots, b \), find the block \( c^s_k \) in \( \rho^s = \{c^s_1, \ldots, c^s_b\} \) and
   - generate the mean \( \mu^s_k \) from the MVN full conditional distribution,
   - generate the covariance parameter \( \sigma^2_{k(s)} \) from the inverse-Gamma full conditional distribution,
3. generate the latent variables \( u^s_i \), for \( i = 1, \ldots, n \), from the MVN full conditional distribution,
4. generate transformation matrix \( W^s \) by sampling \( w^s_j \), for \( j = 1, \ldots, m \), from the MVN full conditional distribution,
5. generate hyperparameter \( v^s \) from the inverse Gamma full conditional distribution,

where \( z_{-r} \) refers to the values of \( z \) except the one corresponding to time point \( r \).

### 2.5 The PPLVM identification

If the transformation matrix and latent variables are subjected to an orthogonal rotation, this gives rise to the same distribution for the observed data. Hence, the PPLVM suffers from identification issues as it is difficult to identify the model parameters unless restrictions are imposed.
There are several approaches in the literature available for dealing with non-identifiability of the related factor analytic models which can also be adopted for the PPLVM. A unique PPLVM is defined by constraining the transformation matrix such that the first \( q \) rows are lower triangular with positive diagonal elements \([11]\). Although this approach is known to impose structure on the ordering of the variables \([1]\), this does not affect how the PPLVM identifies the change points.

3. Simulations and case studies

In this section, we present results of applying the proposed and current existing approaches to the problem of multiple change points in a multivariate data setting on simulated and real datasets. A Gibbs sampling scheme of 20,000 iterations, thinning every 20th iteration, was set out for all analyses. The first 5000 iterations were discarded as burn-in.

3.1 Simulation study

Simulation experiments were carried out to assess the performance of the PPLVM in comparison to other current existing methods for multiple change-point analysis of multivariate time series such as \( \text{bcp} \) \([7]\), \( \text{ecp} \) \([13]\), and dynamic programming (\( \text{DP} \)) by Maboudou-Tchaoa and Hawkins \([22]\). Different simulation scenarios presented in this section illustrate different aspects of the proposed approach, the PPLVM, in detecting distributional changes in the mean and covariance structure. Three scenarios were considered and data were generated from a Gaussian distribution, \( \text{MVN}_m(\mu, \Sigma) \), where the mean \( \mu = (\mu_1, \ldots, \mu_m) \) and covariance \( \Sigma \) were allowed to change at certain time points. In the first scenario, the distributional changes were imposed on the mean structure; in the second on the covariance structure; and in the third on both the mean and the covariance structure.

**Scenario 1: Changes in the mean structure.** In order to compare our approach to the other three methods, data in this scenario were simulated following the simulation strategy in \( \text{DP} \) \([22]\). Three time series \( m = 3 \) of length \( n = 125 \) with change points in the mean at time points \( \tau_1 = 30, \tau_2 = 65, \) and \( \tau_3 = 100 \) were generated according to the \( \text{DP} \) scheme as follows:

- \( \mu_1 = (0, 0, 0) \) for observations 1:30,
- \( \mu_2 = (0, \mu_o, 0) \) for observations 31:65,
- \( \mu_3 = (0, \mu_o, \mu_o) \) for observations 66:100, and
- \( \mu_4 = (-\mu_o, \mu_o, \mu_o) \) for observations 101:125.

The covariance is fixed at \( \Sigma = I_m \) (identity covariance) and the mean parameters were set at \( \mu_o = \mu + k\sigma \), where \( \mu = 0, k = 3, \) and \( \sigma = 1 \). The PPLVM, \( \text{bcp} \), \( \text{ecp} \), and \( \text{DP} \) were all applied to the generated data to identify the three change points and results are presented in Figure 1 and Table 1.

Small posterior probabilities \( pp \) is the model conveying the absence of a change point, \( pp \) close to 1 is the model identifying a change point with much resolution, whilst \( pp \) close to 0.5 is the model giving a less clear-cut signal. Table 1 shows that the PPLVM, as well as \( \text{bcp} \) and \( \text{DP} \), performed very well in predicting the three change points, whilst \( \text{ecp} \) failed to correctly locate the last change point. Although \( \text{bcp} \) estimates for the change points match those of the PPLVM, the posterior probabilities corresponding to the identification of these change points were consistently higher for the PPLVM method (0.85, 0.81, 0.97) than the \( \text{bcp} \) method (0.44, 0.51, 0.81) as shown in Figure 1(d), implying that specificity provided by the former method is greater. The posterior means estimated by the PPLVM presented in Figure 1(a)–(c) show that the PPLVM predicted the true average time profile.
Figure 1. Results of scenario 1. Results of applying the PPLVM to scenario 1 data: Plots (a)–(c) show the PPLVM predicted time profiles for the three time series denoted by solid lines, whilst the true observed profiles are denoted by dotted lines. Plot (d) shows the posterior probabilities of change $P(z_i = 1 | x, u, W)$ estimated by the PPLVM shown in red solid line and by bcp shown in black dashed line. The PPLVM estimated change points at $\tau_1 = 30$, $\tau_2 = 65$, and $\tau_3 = 100$ with probabilities 0.85, 0.81, and 0.97 respectively. Both methods identify the three change points, but bcp was less certain of the change points since its estimated posterior probabilities were lower.

Table 1. Scenario 1: Estimated change points by the bcp, PPLVM, DP, and ecp.

| Methods | $\tau_1 = 30$ | $\tau_2 = 65$ | $\tau_3 = 100$ |
|---------|--------------|--------------|----------------|
| bcp     | 30 (pp = 0.44) | 65 (pp = 0.51) | 100 (pp = 0.81) |
| PPLVM   | 30 (pp = 0.85) | 65 (pp = 0.81) | 100 (pp = 0.97) |
| DP      | 30 (se = 1.24) | 65 (se = 1.27) | 100 (se = 1.00) |
| ecp     | 31 (se = 1.21) | 66 (se = 1.18) | 98 (se = 1.20)  |

Note: Given in parentheses are posterior probabilities (pp) for bcp and PPLVM, and standard errors (se) for DP and ecp.

Scenario 2: Changes in the covariance structure. In this scenario, we examine the performance of the PPLVM when the joint distribution of the multivariate time series changes whilst the marginal distribution is constant. Simulation approach taken in ecp [13] was adopted in this
scenario such that \( m = 5 \) time series of length \( n = 125 \) were generated with change points in the covariance structure \( \Sigma \) at the same positions as in the first scenario, \( \tau_1 = 30 \), \( \tau_2 = 65 \), and \( \tau_3 = 100 \), as follows:

- \( \Sigma_1 = I_m \) for observations 1:30,
- \( \Sigma_2 = \begin{pmatrix} 1 & 0.9 & 0.9 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 & 0.9 & 0.9 \\ 0.9 & 0.9 & 1 & 0.9 & 0.9 \\ 0.9 & 0.9 & 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 0.9 & 0.9 & 1 \end{pmatrix} \) for observations 31:65,
- \( \Sigma_3 = \Sigma_1 \) for observations 66:100, and
- \( \Sigma_4 = \Sigma_2 \) for observations 101:125,

whilst the mean is set at \( \mu = (0, 0, 0, 0, 0) \). Only the PPLVM, \( bcp \), and \( ecp \) were applied to investigate change points in this scenario and results are shown in Table 2.

The performance of the methods depends on the type of distributional changes. Since the change in this second scenario is on the full joint distribution of the multivariate time series data, Table 2 shows that the \( bcp \) method failed (pp \( \ll 0.5 \)) to identify any change points. The \( bcp \) approach cannot detect changes in the covariance structure as it assumes that the multivariate time series are independent. Figure 2 shows that the posterior probabilities for \( bcp \) are flat throughout, whilst the proposed model accurately estimated change points at \( \tau_1 = 30 \), \( \tau_2 = 65 \), and \( \tau_3 = 100 \) with probabilities 0.91, 0.77, and 0.75, respectively. Although the simulation approach in this second scenario was adopted from the \( ecp \) method, it estimated change-point locations which were slightly different from the true change points. Table 2 shows that the PPLVM outperforms the two methods as any changes in the correlation structure is exposed in its noise model. This highlights that, in many cases where the parametric distribution of the data is known, non-parametric approaches are less powerful in detecting changes when compared to parametric approaches.

**Scenario 3: Changes in both mean and covariance structure.** Unlike the first two scenarios, the third scenario has only two change points at \( \tau_1 = 40 \) and \( \tau_2 = 70 \) on both the mean and the covariance structure. As in the first scenario, data for this scenario were generated according to a simulation strategy by Maboudou-Tchaoa and Hawkins [22] but it was improved to allow the covariances between the three observed variables to change over time. Three correlated time series of length \( n = 100 \) with two change points on both \( \mu \) and \( \Sigma \) were generated using the following scheme:

- \( \mu_1 = (0, 0, 0) \) and \( \Sigma_1 = \begin{pmatrix} 0.80 & 0.02 & 0.05 \\ 0.02 & 1.30 & 0.10 \\ 0.05 & 0.10 & 1.10 \end{pmatrix} \) for observations 1:40,
- \( \mu_2 = (0, \mu_o, 0) \) and \( \Sigma_2 = \begin{pmatrix} 5.00 & 0.70 & 0.65 \\ 0.70 & 6.50 & 0.80 \\ 0.65 & 0.80 & 5.0 \end{pmatrix} \) for observations 41:70, and
- \( \mu_3 = (0, \mu_o, \mu_o) \) and \( \Sigma_3 = \Sigma_1 \) for observations 71:100.

| Methods | \( \tau_1 = 30 \) | \( \tau_2 = 65 \) | \( \tau_3 = 100 \) |
|---------|----------------|----------------|----------------|
| \( bcp \) | 30 (pp = 0.09) | 65 (pp = 0.11) | 100 (pp = 0.01) |
| PPLVM   | 30 (pp = 0.91) | 65 (pp = 0.77) | 100 (pp = 0.75) |
| \( ecp \) | 26 (se = 1.31) | 64 (se = 1.29) | 98 (se = 1.33) |

Note: Given in parentheses are posterior probabilities (pp) for PPLVM and \( bcp \), and standard errors (se) for \( ecp \).
Figure 2. Results of scenario 2. Results of applying the PPLVM to scenario 2 data: The posterior means for the five time series are shown in solid lines, whilst the observed series are shown in dotted lines. The plot shows that the model managed to predict the mean profiles of the five time series which was fixed at zero. The lower plot compares the posterior probabilities \( P(z_i = 1 \mid x, u, W) \) estimated by the PPLVM shown in solid lines and \( \text{bcp} \) in dashed lines. The PPLVM model identified the three change points but the \( \text{bcp} \) method failed.

The results of applying the PPLVM, \( \text{bcp} \), and \( \text{ecp} \) are presented in Table 3.

The table shows that the proposed approach managed to correctly infer that the generated data have two change points at \( \tau_1 = 40 \) and \( \tau_2 = 70 \) with probabilities 0.80 and 0.82, respectively. As in the second scenario, the \( \text{bcp} \) approach found it difficult to identify any change points since the changes were on the joint distribution of the data. As shown in Table 3, in this scenario, the \( \text{ecp} \) method provided correct estimates of the change points.

### 3.2 Case study 1: Climate change data

In this section, we assess the potential utility of the PPLVM in climate change analysis by applying it to five Swiss station annual temperature data sets publicly available through the European Climate Assessment and Database (ECAD) project web page [www.eca.knmi.nl](http://www.eca.knmi.nl). The data consist of 104 average annual temperatures for the period 1900–2004 for 5 Swiss stations, Basel, Geneva, Lugano, Saentis, and Zurich. The aim is to establish timing of potential
changes in the Swiss temperatures in this period. These data were analyzed in the web address www.chartsgraphs.wordpress.com using cumulative control charts (CUSUM) first developed by Page [26] for detecting shifts in trends. In their analysis, they identified four change points in Geneva which include 1920, 1943, 1963, and 1989. Lugano and Zurich had three change points (1919, 1940, 1988) and (1943, 1952, 1988), respectively. Saentis had two change points 1929 and 1988. Basel had only one change point 1988 and this change point was common to all five Swiss stations.

Although the five stations are far apart and their temperatures are different, the PPLVM can be applied to compare and identify timing of joint changes in the trends. Figure 3 indicates that the PPLVM clearly identifies three changes in the trend of the five temperature stations on years 1929, 1963, and 1989 and assigns high posterior probabilities to each, 0.54, 0.97, and 0.97, respectively. Although the PPLVM is less certain (pp ≪ 0.5) of the change point identified by O’Day on year 1943 in Geneva and Zurich, a notable spike in the posterior probability (0.41) on this year can be seen in Figure 3. The posterior estimates of the annual temperatures for the five stations are also presented in Figure 3. The model predicts that the change point in annual temperature on year 1963 was less pronounced in Basel. Since these change points were on the joint distribution of the five cities, the bcp approach had problems identifying these change points, and the estimated problems were less than 0.5.

3.3 Case study 2: Freeze drying data

In an industrial controlled setting, this model can be used to detect changes in the underlying process generating data. In this example, the PPLVM is applied to the freeze drying dataset from pharmaceutical industry to detect changes in the process called freeze drying. Freeze drying, also known as lyophilization, is a process used in pharmaceutical industry to remove water from pharmaceutical products by sublimation at very low temperatures [4]. Freeze drying products improve their stability for increased shelf life, preserve their biological activity, and maintain the basic structure of the product. Freeze drying process is performed through three successive stages: (i) the freezing stage, water is converted to ice and solutes are crystallized; (ii) the primary drying stage, the product temperature is slowly increased whilst pressure is reduced removing ice crystals by sublimation; and (iii) the secondary drying stage, temperature is gradually increased under high vacuum to remove remaining water. In order to reduce time and costs of the process, the main task is to monitor and determine endpoints of the freeze drying process.

Spectroscopic process analyzers in combination with chemometrics tools are used to monitor different stages of freeze drying [4,5,30]. The main limitations with chemometrics methods is that they are not based on any probability density model. They assume observations are independent, ignoring correlation due to repeated time measurements of the freeze drying cycle. They cannot be used to efficiently predict the endpoints of freeze drying cycle with certain level of assurance.
The problem of determining endpoints of different stages of freeze drying can be treated as a multiple change-point problem. In this example, the PPLVM is employed to identify the end and start of primary and secondary drying, respectively, using information measured from the product temperature, condenser temperature, and chamber pressure variables only. The freeze drying data used here consist of 200 time points of primary to secondary drying stages measured on 17 variables (2 shelf temperature, 9 product temperature, 5 condenser temperature and 1 chamber pressure). Figure 4 shows that the PPLVM managed to correctly identify the end or start of primary or secondary drying, respectively (which was after 99.6 h shown by the first green arrow) with probability 1, whilst bcp identifies this change point at 98.6 with probability 0.49 and a notable spike at 99.6. The PPLVM identified the end of secondary drying to be around 108 h (which is 1 h later than the correct change point shown by the second green arrow) with probability 1, whilst bcp identified it at around 109 h (2 h later than the correct change point) with probability 0.81.
Figure 4. Results of applying the PPLVM to the Freeze drying data. The upper plot shows the product and condenser temperature profiles denoted by black and grey dashed and dotted lines, respectively. The pressure profile is denoted by a blue dash-dotted line. The shelf temperature profiles in red show the true endpoints for primary and secondary drying (set by the manufacturer) denoted by two green arrows, respectively. The lower plot compares the posterior probabilities $P(z_i = 1 | \mathbf{x}, \mathbf{u}, \mathbf{W})$ estimated from the data without the shelf temperature profile by the PPLVM in solid red line and bcp in black dashed lines. The PPLVM identifies the end of both primary and secondary drying at $\tau_1 = 99.6$ and $\tau_2 = 108$ h with probability 1, respectively.

4. Discussion

Simulated and real examples presented in this article demonstrate that the PPLVM is an important tool for detecting multiple change points in multivariate time series data. Perhaps the key contribution of this article is the extension of the Barry and Hartigan [3] PPM by providing a full Bayesian approach to the problem of detecting distributional changes in the mean and covariance of a multivariate time series Gaussian data. Compared to the currently existing methods for multiple change-point analysis for multivariate data, the dimensionality reduction approach, the PPLVM, was shown to perform well. One important advantage of this approach is that it can also be used to model data from a high-dimensional setting, providing a smaller dimensional representation of the observed series. The PPLVM is very flexible and it can be applied to a wide variety of settings which include finance, bioinformatics, genetics, and engineering.
Currently, compared to other approaches, the PPLVM has computational complexity quadratic relative to the length of the time series. Further work is required in the effort to improve further the computational efficiency of the PPLVM.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

We acknowledge the support of the Mathematics Applications Consortium for Science and Industry (www.macsi.ul.ie) funded by the Science Foundation Ireland Investigator Award 12/IA/1683 and the support of Enterprise Ireland, within the framework of the Pharmaceutical Manufacturing Technology Centre (PMTC) Initial Research Programme, Grant TC/2012/5105.

Supplemental data

Supplemental data for this article can be accessed at doi:10.1080/02664763.2015.1029444.

References

[1] O. Aguilar and M. West, Bayesian dynamic factor models and portfolio allocation, J. Bus. Econom. Stat. 18 (2000), pp. 338–357.
[2] D. Barry and J.A. Hartigan, Product partition models for change point problems, Ann. Stat. 20 (1992), pp. 260–279.
[3] D. Barry and J.A. Hartigan, A Bayesian analysis for change point problems, J. Am. Statist. Assoc. 88 (1993), pp. 309–319.
[4] T.D. Beer, M.G.T. Alleso, F. Goethals, A. Coppens, Y. Vander Heyden, H. Lopez De Diego, J. Rantanen, F. Verpoort, C. Vervaet, J.P. Remon, and W.R.G. Baeyens, Implementation of a process analytical technology system in a freeze drying process using Raman spectroscopy for in-line process monitoring, Anal. Chem. 79 (2007), pp. 382–388.
[5] T.D. Beer, P. Vercruysse, J. Quinten, J. Ouyang, X. Zhang, C. Vervaet, J.P. Remon, and W.R.G. Baeyens, On-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (pat) tools, J. Pharm. Sci. 98 (2009), pp. 3430–3446.
[6] G. Bormetti, M.E. De Giuli, D. Delpini, and C. Tarantola, Bayesian value-at-risk with product partition models, Quant. Finance, 12(5) (2010), pp. 769–780.
[7] C. Erdman and J.W. Emerson, A fast Bayesian change point analysis for the segmentation of microarray data, Bioinformatics 24 (2008), pp. 161–166.
[8] P. Fearnhead, Exact Bayesian curve fitting and signal segmentation, IEEE Trans. Signal Process. 53 (2004), pp. 2160–2166.
[9] P. Fearnhead, Exact and efficient Bayesian inference for multiple changepoint problems, Stat. Comput. 16 (2006), pp. 203–213.
[10] A. Gelman, Bayesian Data Analysis, Chapman and Hall, London, 1995.
[11] J. Geweke and G. Zhou, Measuring the pricing error of the arbitrage pricing theory, Rev. Financ. Stud. 9 (1996), pp. 557–587.
[12] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman and Hall, London, 1996.
[13] N.A. James and D.S. Matteson, ecp: An r package for nonparametric multiple change point analysis of multivariate data, Tech. Rep., Cornell University, 2013.
[14] J. Kappenman, A perfect storm of planetary proportions, IEEE Spectrum 49 (2012), pp. 26–31.
[15] R. Killick and I.A. Eckley, Changepoint: An r Package for Changepoint Analysis, J. Stat. Softw. 58(3) (2014). Available at http://www.jstatsoft.org/v58/i03.
[16] R. Killick, I.A. Eckley, P. Jonathan, and K. Ewans, Detection of changes in the characteristics of oceanographic time-series using statistical change point analysis, Ocean Eng. 37 (2010), pp. 1120–1126.
[17] R. Killick, P. Fearnhead, and I.A. Eckley, Optimal detection of changepoints with a linear computational cost, J. Am. Statist. Assoc. 107 (2012), pp. 1579–1995.
[18] L. Lakhina, M. Crovella, and C. Diot, Diagnosing network-wide traffic anomalies, ACM SIGCOMM, Comput. Commun. Rev. 34(4) (2004), pp. 219–230.
[19] R.H. Loschi and F.R.B. Cruz, *An analysis of the influence of some prior specifications in the identification of change points via product partition model*, Comput. Stat. Data Anal. 39 (2002), pp. 477–501.

[20] R.H. Loschi and F.R.B. Cruz, *Extension to the product partition model: Computing the probability of a change*, Comput. Stat. Data Anal. 48 (2005), pp. 255–268.

[21] R.H. Loschi and F.R.B. Cruz, *Multiple change-point analysis for linear regression models*, Chilean J. Stat. 1 (2010), pp. 93–112.

[22] E.M. Maboudou-Tchaoa and D.M. Hawkins, *Detection of multiple change-points in multivariate data*, J. Appl. Stat. 10 (2013), pp. 1979–1995.

[23] M. Modisett and E.M. Maboudou-Tchao, *Significantly lower estimates of volatility arise from the use of open high low close price data*, North Am. Actuar. J. 14 (2010), pp. 68–85.

[24] G. Nyamundanda, L. Brennan, and I.C. Gormley, *Probabilistic principal component analysis for metabolomic data*, Bioinformatics 11(1) (2010). http://www.biomedcentral.com/1471-2105/11/571.

[25] G. Nyamundanda, I.C. Gormley, and L. Brennan, *A dynamic probabilistic principal components model for the analysis of longitudinal metabolomics data*, J. R. Statist. Soc. Ser. C 63 (2014), pp. 763–782.

[26] E.S. Page, *Continuous inspection scheme*, Biometrika 41 (1954), pp. 100–115.

[27] M. Raginsky, R. Willett, C. Horn, J. Silva, and R. Marcia, *Sequential anomaly detection in the presence of noise and limited feedback*, IEEE Trans. Inform. Theory 58 (2012), pp. 5544–5562.

[28] J. Reeves, J. Chen, X.L. Wang, R. Lund, and Q. Lu, *A review and comparison of changepoint detection techniques for climate data*, J. Appl. Meteorol. Climatol. 46 (2007), pp. 900–915.

[29] S. Rowes, *Procrustes analysis*, Tech. Rep., Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, 2000.

[30] M.C. Sarraguca, T. De Beer, C. Vervaet, J.P. Remon, and J. Lopes, *A batch modelling approach to monitor a freeze drying process using in-line Raman spectroscopy*, Talanta 83(1) (2010), pp. 130–138.

[31] M.E. Tipping and C.M. Bishop, *Probabilistic principal component analysis*, J. R. Statist. Soc. Ser. B 61 (1999), pp. 611–622.

[32] J. Wyse and N. Friel, *Simulation-based Bayesian analysis for multiple change points*, arXiv:1011.2932v1[stat.CO], 2010.

[33] X. Xuan and K. Murphy, *Modeling changing dependency structure in multivariate time series*, Proceedings of the 24th international conference on machine learning, Corvalis, OR, 2007.