Technometrics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

Bayesian Hierarchical Model for Change Point Detection in Multivariate Sequences

Huaqing Jin, Guosheng Yin, Binhang Yuan & Fei Jiang

To cite this article: Huaqing Jin, Guosheng Yin, Binhang Yuan & Fei Jiang (2022) Bayesian Hierarchical Model for Change Point Detection in Multivariate Sequences, Technometrics, 64:2, 177-186, DOI: 10.1080/00401706.2021.1927848

To link to this article: https://doi.org/10.1080/00401706.2021.1927848

TECHNOMETRICS
2022, VOL. 64, NO. 2, 177–186
https://doi.org/10.1080/00401706.2021.1927848

Taylor & Francis
Taylor & Francis Group

Bayesian Hierarchical Model for Change Point Detection in Multivariate Sequences

Huaqing Jin[a], Guosheng Yin[a], Binhang Yuan[b], and Fei Jiang[c]

[a]Department of Statistics and Actuarial Science, the University of Hong Kong, Hong Kong, Hong Kong; [b]Computer Science Department, Rice University, Houston, TX; [c]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA

ABSTRACT

Motivated by the wind turbine anomaly detection, we propose a Bayesian hierarchical model (BHM) for the mean-change detection in multivariate sequences. By combining the exchange random order distribution induced from the Poisson–Dirichlet process and nonlocal priors, BHM exhibits satisfactory performance for mean-shift detection with multivariate sequences under different error distributions. In particular, BHM yields the smallest detection error compared with other competitive methods considered in the article. We use a local scan procedure to accelerate the computation, while the anomaly locations are determined by maximizing the posterior probability through dynamic programming. We establish consistency of the estimated number and locations of the change points and conduct extensive simulations to evaluate the BHM approach. Among the popular change point detection algorithms, BHM yields the best performance for most of the datasets in terms of the F1 score for the wind turbine anomaly detection.

ARTICLE HISTORY
Received August 2020
Accepted May 2021

KEYWORDS
Change points; Multivariate data; Nonlocal prior; Nonmaximum suppression; Poisson–Dirichlet process

1. Introduction

The wind turbine blade is the core component of the wind power generation system, and hence the failure of the wind turbine blade directly leads to the dysfunction of the whole system. One leading cause of the wind turbine failure is the accumulated ice that covers the turbine as shown in the left panel of Figure 1. Typically, the supervisory control and data acquisition (SCADA) system is used to examine the ice level in real time by monitoring several signal sequences simultaneously, such as wind speed, environment temperature and accelerated speed. Once the ice on the wind turbine reaches a certain level that may lead to the failure of the system, some of the signals would exhibit mean shifts as shown in the right panel of Figure 1.

To timely capture the anomalies on the wind turbines, we apply several existing mean detection algorithms to the signals generated by the SCADA system, including E-division (ECP) (Matteson and James 2014), the dynamic programming-based maximum likelihood estimation (DPMLE) (Maboudou-Tchao and Hawkins 2013), and the popular structure change detection algorithm called AutoPlait (Matsubara, Sakurai, and Faloutsos 2014). The vertical red shadows in Figure 2 are the manually labeled change point locations based on the observed wind turbine blade failure times. Although these labels may not include all change points, they may serve as the basis for the comparison among different methods. A method would be considered the best if the detected change points constitute the smallest set that contains all the labels. As shown in Figure 2, none of the existing methods under consideration provides satisfactory results: ECP tends to select a large number of spurious change points, while DPMLE and AutoPlait miss almost all the change points.

This phenomenon motivates us to propose a novel algorithm based on a Bayesian hierarchical model (BHM) to detect mean shifts among multiple sequences. The BHM naturally borrows information across multiple sequences by using a prior induced from the Poisson–Dirichlet process, and hence it can identify true change points more effectively. In addition, BHM uses the nonlocal priors to control the false discoveries, which is shown to yield the smallest detection error in comparison with the existing methods under consideration. To reduce the computational burden, we introduce an initial local scan procedure in our algorithm, and then use the dynamic programming (Bellman and Roth 1969; Du, Kao, and Kou 2016) to identify the change point locations by optimizing the posterior probability.

The contributions of our work are three-fold: (i) We develop a novel Bayesian method to estimate both the number and locations of the change points in an integrative manner. Our method is shown to outperform the competitive ones in both simulation studies and real application to the wind turbine data. (ii) We explore the advantages of using nonlocal priors in BHM for reducing the detection errors of the change points with multivariate sequences. (iii) We establish the consistency of our BHM method, in that it can identify the correct number and locations of change points asymptotically, providing the sample size is sufficiently large.

The rest of the article is organized as follows. We discuss the related work in Section 2, and provide the BHM method and its theoretical properties in Sections 3 and 4, respectively. In Section 5, we conduct extensive simulation studies to compare our proposal with some existing methods. Finally, we apply the proposed method to the wind turbine data in Section 6.

CONTACT Fei Jiang Fei.jiang@ucsf.edu Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.

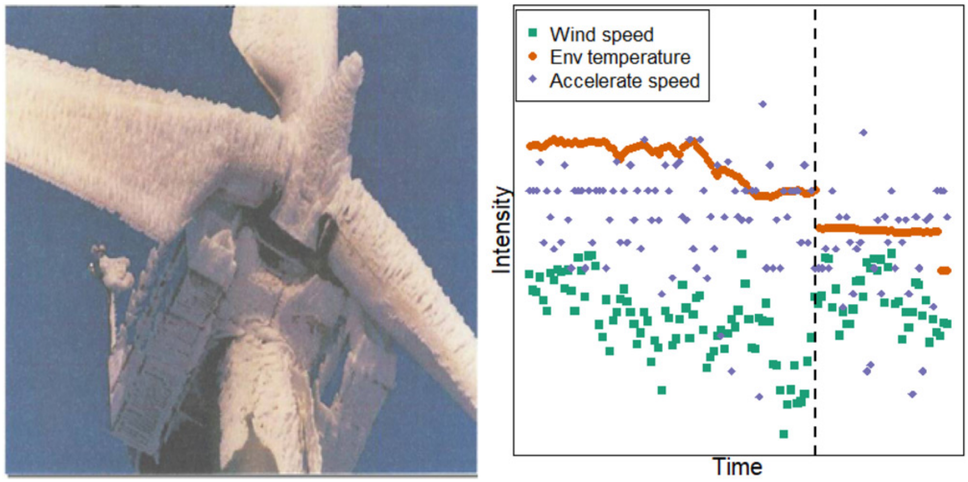

Figure 1. The wind turbine blade covered by ice (left) and the mean shifts of the wind speed, environment temperature and accelerated speed (right). The vertical dashed black line highlights the change-point location.

2. Related Work

The wind turbine anomaly detection problem has been extensively studied (Tautz-Weinert and Watson 2016) under different contexts. Other than the standard SCADA system, novel anomaly physical detectors (Muñoz, Jiménez, and Márquez 2018) with higher signal-to-noise ratios have been developed aiming at amplifying the abnormal signals of the wind turbine failure. However, the practical use of these new physical detectors is limited due to the high cost (Yang, Court, and Jiang 2013). Moreover, many anomaly detection algorithms have been developed to capture the anomalies using the widely used SCADA system (Tautz-Weinert and Watson 2016). Yang, Court, and Jiang (2013) proposed a trending method using bin averaging with output power, wind speed or generator speed, and a quantifying criterion was introduced based on a correlation model of historical and present data. Kim et al. (2011) applied an artificial neural network self-organizing map approach to the SCADA data, which detected the anomalies through clustering. Relying upon the correlation analysis and physics of the system, high-order polynomial models were developed for the anomaly detection (Wilkinson et al. 2014). Gray and Watson (2010) introduced a damage model based on a physical understanding of the particular failure mode of interest for damage calculation as well as failure probability estimation. However, all the aforementioned methods rely on additional knowledge about other wind turbine features, historical data or domain experience, which are not available in the current wind turbine dataset. This motivates us to develop a new change point detection algorithm to identify anomalies solely based on the signal patterns.

Change point detection algorithms have been widely used to detect anomalies (Muggeo and Adelfio 2010). Under the frequentist framework, the change point detection relies on optimizing certain objective functions, such as a parametric or nonparametric log-likelihood function (Hawkins 2001; Zou et al. 2014), quadratic loss (Rigaill 2015), and cumulative sums (Hinkley 1971; Manogaran and Lopez 2018). The Bayesian information criterion (BIC) (Yao 1988) and its variants (Yao and Au 1989; Zhang and Siegmund 2007) are commonly used for determining the number of change points.

On the other side, the Bayesian algorithms identify the change point locations by maximizing the posterior distribution (Martínez and Mena 2014) or the marginal likelihood (Du, Kao, and Kou 2016). The optimization routines are performed through the Markov chain Monte Carlo (MCMC) (Barry and Hartigan 1993; Martínez and Mena 2014) or dynamic programming (Du, Kao, and Kou 2016). More recently, Hopfield's network has been advocated for use to identify change points (Fuentes-García, Mena, and Walker 2019). By considering the randomness in the model parameters and incorporating prior distributions, Bayesian methods automatically add penalties on the number of change points (Lavielle 2005). As a result, the number of change points can be determined seamlessly in conjunction with their locations (Truong, Oudre, and Vayatis 2018).

Change point problems can also be considered under the context of the statistical process control (SPC) (Qiu 2013). The conventional methods for change point detection under SPC are the Shewhart and cumulative sum charts as well as the exponential weighted moving average (EWMA) chart (Hawkins, Qiu, and Kang 2003). Tsiamyrtzis and Hawkins (2005) introduced a dynamic model to handle the short-run process with the aim to detect the mean shift. Using a sequential estimation technique, Zamba and Hawkins (2006) considered the multivariate change point problem for SPC when the in-control parameters were unknown. Using an autoregression model, Tsiamyrtzis and Hawkins (2008) worked on autocorrelated processes with mean shift under a Bayesian EWMA method. The goal of these change point detection methods is to detect a single change in the sequence, while they can be adapted to identify multiple changes with an unknown number of change points.

To construct the posterior distribution or the marginal likelihood under the Bayesian paradigm, one crucial step is to specify the prior distributions. The commonly used priors for the mean differences include the local priors (Bertolino, Racugno, and Moreno 2000), such as normal prior distributions (Du,

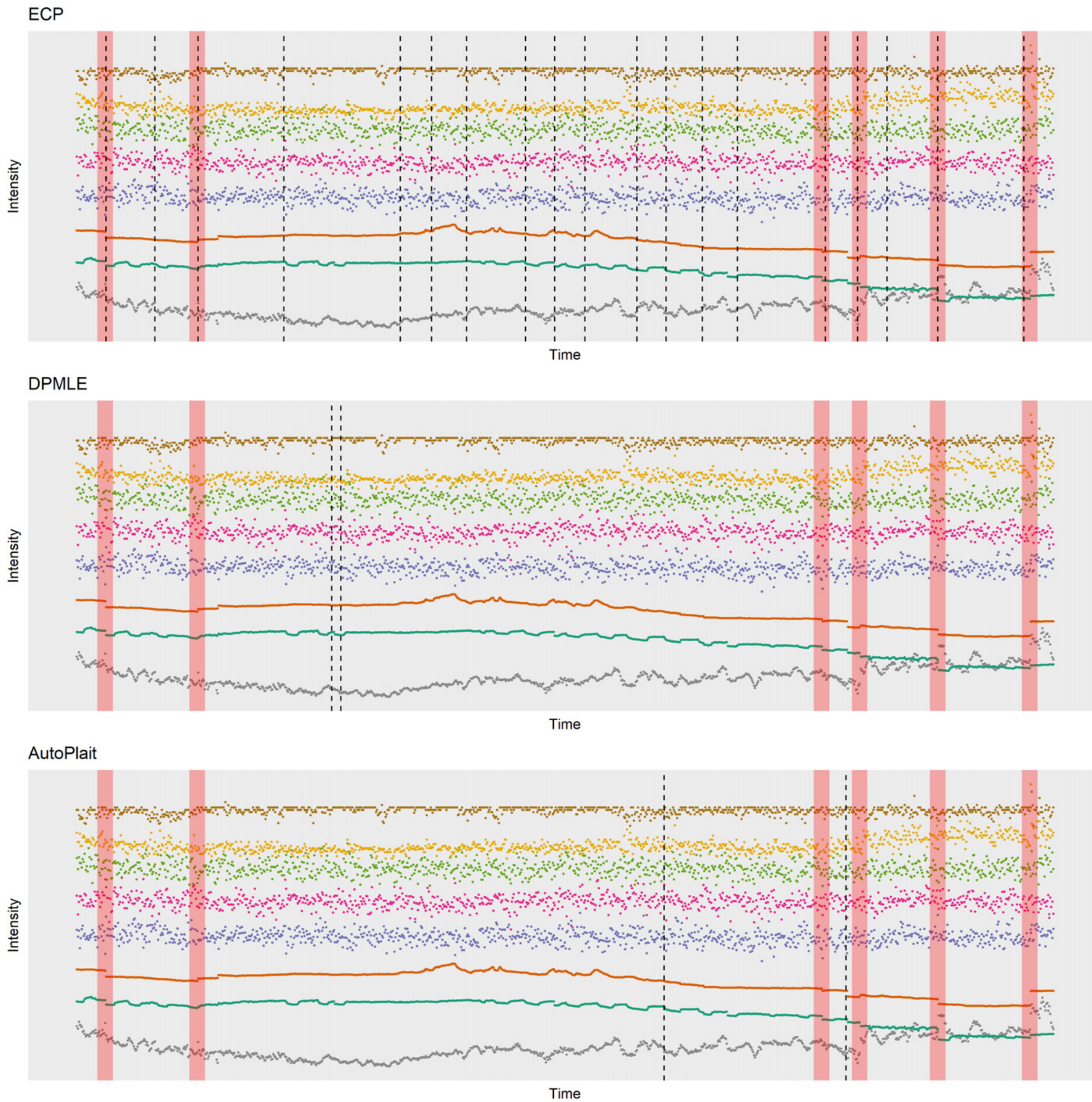

Figure 2. The detection results of three existing methods, namely ECP, DPMLE, AutoPlait, for the wind turbine dataset 1 with $n = 8$ sequences. The black dashed lines are the estimated state shift points and the red shadows are the m_l-neighborhood of the ground truth.

Kao, and Kou 2016), and nonlocal priors, such as the moment prior and inverse moment prior distributions (Johnson and Rossell 2010). The nonlocal priors were first proposed in the Bayesian hypothesis testing framework to improve the speed of the accumulation of the evidence in favor of the true null model (Johnson and Rossell 2010). Jiang, Yin, and Dominici (2018) applied the nonlocal prior to identify the change points in a single sequence of data, which leads to a faster convergence rate compared with the algorithms based on the local priors.

Because the change point detection can be regarded as a special clustering problem, we can use the exchangeable random partition distribution (ERPD) (Pitman 1995) to construct the prior distribution of the segments. The ERPD has been widely used for clustering problems (Lau and Green 2007; Wade et al. 2018), which penalizes the model complexity (Pitman 2002) and automatically selects the number of clusters (McCullagh and Yang 2008). However, ERPD is not directly applicable to the change point detection, because it does not account for the order constraints in the change point problem. Martínez and Mena (2014) proposed to use a modified ERPD, namely the exchangeable random order distribution (EROD), as the prior distribution specifically for the change point detection, which inherits the symmetry and automatic penalization properties of ERPD.

In addition to the mean-change detection, many other applications of change points have been carried out. Matsubara, Sakurai, and Faloutsos (2014) proposed the AutoPlait method to detect the changes in the periodic sequences for identifying the structure changes in the signals. Gharghabi et al. (2017) proposed a fast, low-cost online semantic segmentation for the structure change detection in cyclic data. Bouchard and Badler (2007) introduced a Laban movement-based segmentation method for the motion data. Gong, Medioni, and Zhao (2014) used the kernelized temporal cut method to recognize action changes in the continuous monocular motion sequences.

3. Bayesian Multivariate Change Point Detection

3.1. Probability Model

Suppose there are n sequences of signals measured over time, where correlations exist both between sequences and within sequences. Let Y_{ik} represent the strength of the ith signal at time k, $i = 1, \dots, n$; $k = 1, \dots, T$. Define $\boldsymbol{Y}_k = (Y_{1k}, \dots, Y_{nk})^\top$, and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_T)$ is an $n \times T$ observation matrix and $\boldsymbol{Y}_{(a,b]} = (\boldsymbol{Y}_{a+1}, \dots, \boldsymbol{Y}_b)$ represents the matrix containing the $(a+1)$th to bth columns of $\boldsymbol{Y}$. Let $\mathcal{K} = \{\kappa_0, \dots, \kappa_{p+1}\}$ be a generic notation for a set of p change points, with $\kappa_0 = 0$ and $\kappa_{p+1} = T$, and let $\mathcal{K}_0 = \{\kappa_{00}, \dots, \kappa_{0,p_0+1}\}$ be the true change point set, with $\kappa_{00} = 0$ and $\kappa_{0,p_0+1} = T$. In addition, let $\{N_s, s = 0, \dots, p\} = \{(\kappa_{s+1} - \kappa_s), s = 0, \dots, p\}$ be a collection of numbers of observations between consecutive change points.

Let $\overline{Y}_{i,\kappa_s} = (\kappa_s - \kappa_{s-1})^{-1} \sum_{k=\kappa_{s-1}+1}^{\kappa_s} Y_{ik}$, with $\overline{Y}_{i,\kappa_0} = 0$. We assume

$$\begin{aligned} \frac{Y_{ik} - \overline{Y}_{i,\kappa_s} - \mu_{is}}{\omega_s} \Big| \mathcal{K}, \mu_{is}, \omega_s &\sim \pi_0, \quad \text{for } k \in (\kappa_s, \kappa_{s+1}], \\ \mu_{is} &\sim \pi_\mu(\mu_{is}), \\ \omega_s &\sim \pi_\omega(\omega_s), \\ (N_0, \dots, N_p) &\sim \pi_k(\mathcal{K}), \end{aligned} \tag{1}$$

where π_0 is the likelihood function selected according to the data distribution. It is worth emphasizing that we model the difference $Y_{ik} - \overline{Y}_{i,\kappa_s}$ rather than Y_{ik}, and such a modeling strategy allows us to better control the convergence rates by selecting suitable priors on the mean differences. In model (1), we assume that all the sequences share the same variance for each segment, which can help to reduce the computational burden and numerical error in the optimization procedure. As shown in Table A.1 (supplementary material), using the same variance parameter does not undermine the performance of our algorithm. In practice, when the sequences have very distinct variances, we may standardize the variance for each sequence before the implementation of our detection procedure. Typically, we use a normal likelihood if the data do not contain outliers, and use a t distribution if the data are contaminated with a substantial amount of outliers. In practice, the existence of outliers can be determined by the generalized extreme studentized deviate test (Rosner 1983), as discussed in Section A.3 of the supplementary materials.

3.2. Prior Specifications

Change point detection is closely related to a special clustering algorithm in which each cluster only contains the neighborhood points. This motivates us to consider ERPD (Pitman 1995; Gnedin and Pitman 2006) induced from the Poisson–Dirichlet process as the prior distribution for $(N_0, \dots, N_p)$, which has been widely used in clustering problems (Broderick, Jordan, and Pitman 2013). By choosing $\sigma = 0$ or $\alpha = 0$, the Poisson–Dirichlet process reduces to the Dirichlet or the normalized stable processes, respectively (Martínez and Mena 2014). The ERPD strikes a good balance between the generalization and complexity. As a special case of Gibbs-type priors, it places a tradeoff on the prior distribution between being informative and noninformative about the number of change points (Lijoi, Mena, and Prünster 2007). However, this prior may classify non-neighborhood signals into the same group, and hence the resulting clusters would contradict with the property of the change points. To account for this neighborhood constraint, the probability mass function of ERPD is modified as

$$\pi_k(\mathcal{K}) = \frac{T!}{(p+1)! \prod_{h=0}^{p} N_h!} \cdot \frac{\prod_{s=1}^{p} (\alpha + s\sigma)}{\prod_{j=1}^{T-1} (\alpha + j)} \prod_{h=0}^{p} \prod_{i=1}^{N_h - 1} (i - \sigma), \quad \sigma \in [0, 1),\ \alpha > -\sigma, \tag{2}$$

which corresponds to the probability mass function of EROD (Pitman 2002; Martínez and Mena 2014). The first term in Equation (2) accounts for the neighborhood constraint.

Under the prior distribution in Equation (2), the marginal distribution of the number of change points p can be derived as

$$\Pr(p = l) = \frac{\prod_{i=1}^{l} (\alpha + i\sigma)}{\sigma^{l+1} \prod_{i=1}^{n-1} (\alpha + i)} \frac{1}{(l+1)!} \sum_{j=0}^{l+1} (-1)^j \binom{l+1}{j} \prod_{i=0}^{n-1} (-j\sigma + i).$$

We can select the values of (α, σ) via encoding our prior brief in the number of change points for the data. As discussed in Martínez and Mena (2014), the parameter σ plays a more important role for detecting the change points, and thus it deserves more attention in practice.

The prior for the mean difference, π_μ, is critical for reducing the detection error of the change points. We consider the local prior, nonlocal moment prior and inverse moment prior, respectively defined as follows:

$$\begin{aligned} \pi_{\mu,L}(\mu) &= \mathrm{N}(0, \psi^2), \\ \pi_{\mu,M}(\mu) &= \frac{\mu^{2\nu}}{C_M} \frac{1}{\sqrt{2\pi}} \exp(-\mu^2/2), \\ \pi_{\mu,I}(\mu) &= \frac{r\phi^{q/2}}{\Gamma(q/2r)} |\mu|^{-(q+1)} \exp\{-(\mu^2/\phi)^{-r}\}, \end{aligned}$$

where $\nu \geq 1$, $\psi, r, q, \phi > 0$ and C_M is the normalizing constant. For the nuisance parameters ω_s, we take $\pi_\omega(\omega_s)$ to be a Gamma distribution.

3.3. Posterior Distribution and Change Point Detection

Based on model (1), the marginal likelihood given a change point set $\mathcal{K}$ can be written as

$$\Pr(\boldsymbol{Y}|\mathcal{K}) = \prod_{s=0}^{p}\int\prod_{i=1}^{n}\int\prod_{k\in(\kappa_s,\kappa_{s+1}]}\frac{1}{\omega_s}\pi_0\left(\frac{Y_{ik}-\overline{Y}_{i,\kappa_s}-\mu_{is}}{\omega_s}\right)\pi_\mu(\mu_{is})\mathrm{d}\mu_{is}\pi_\omega(\omega_s)\mathrm{d}\omega_s. \quad (3)$$

Using the prior in Equation (2), the posterior probability of $\mathcal{K}$ is

$$\Pr(\mathcal{K}|\boldsymbol{Y}) \propto \Pr(\boldsymbol{Y}|\mathcal{K})\pi_k(\mathcal{K}). \quad (4)$$

The standard procedure to optimize the posterior is through dynamic programming (Bellman and Roth 1969; Du, Kao, and Kou 2016). However, because the dynamic programming evaluates the signals at every time point, the computation time grows in the order of $O(MT^2)$ where M is the upper bound of the number of change points (Rigaill 2015; Du, Kao, and Kou 2016). When T is large, the computational burden is prohibitive. To alleviate the computational burden, we propose a screening procedure to reduce the search space of the change points.

While we adopt a Bayesian mechanism for change point detection, it is not a fully Bayesian approach. The prior is used to induce penalties on the parameters so as to identify the best set of change points. More specifically, by using a prior on $\mathcal{K}$, the algorithm induces a penalty on the locations and number of change points. Hence, optimizing the posterior distribution of $\mathcal{K}$ automatically provides the best estimates for the locations and number of change points. Furthermore, the prior on the mean difference induces a penalty on μ_{is}, which reduces the false positive rate of detection.

3.4. Screening Candidate Points

Through the screening step, we can reduce the search space of the change points to a subset of time points which is guaranteed to cover the true change points asymptotically. This leads to a substantial reduction in the computational time when there are much fewer candidate points than the total measurement times in the data.

Let $\overline{Y}_{ik} = m_I^{-1}\sum_{l=k-m_I+1}^{k} Y_{il}$, where m_I is a prespecified window size and the m_I-neighborhood is defined as the set $\{\boldsymbol{Y}_l : l \in (k-m_I, k+m_I)\}$. We construct a local scan statistic,

$$R_k = \prod_{i=1}^{n}\frac{\int\prod_{l=k+1}^{k+m_I}\exp\{-(Y_{il}-\overline{Y}_{ik}-\mu)^2\}\pi_\mu(\mu)\mathrm{d}\mu}{\prod_{l=k+1}^{k+m_I}\exp\{-(Y_{il}-\overline{Y}_{ik})^2\}}. \quad (5)$$

A large value of R_k favors the kth signal vector to be the only change point in its m_I-neighborhood. Thus, $R_k \to \infty$ if k is a true change point, and $R_k \to 0$ if there is no change point in the m_I-neighborhood of the kth signal. Based on this property, we develop Algorithm 1 for selecting the candidate points.

3.5. Change Point Detection With Candidate Points

Let $\mathcal{H}(m_I) = \{\tau_0, \ldots, \tau_{N+1}\}$ be the set containing the candidate points where $\{\tau_i\}_{i=1}^{N}$ are obtained by Algorithm 1 and $\tau_0 = 0$ and $\tau_{N+1} = T$. Given $\mathcal{K}$, we can define the utility function, $U(\mathcal{K}|\boldsymbol{Y}) = \prod_{s=0}^{p} u(\boldsymbol{Y}_{(\kappa_s,\kappa_{s+1}]}, s)$, where

$$u(\boldsymbol{Y}_{(\kappa_s,\kappa_{s+1}]}, s) = \int\prod_{i=1}^{n}\int\prod_{k\in(\kappa_s,\kappa_{s+1}]}\frac{1}{\omega_s}\pi_0\left(\frac{Y_{ik}-\overline{Y}_{i\kappa_s}-\mu_{is}}{\omega_s}\right)\pi_\mu(\mu_{is})\mathrm{d}\mu_{is}\pi_\omega(\omega_s)\mathrm{d}\omega_s \times\frac{(\alpha+s\sigma)}{(s+1)}\frac{\prod_{i=1}^{N_s-1}(i-\sigma)}{N_s!}, \quad (6)$$

and $u(\boldsymbol{Y}_{(\kappa_s,\kappa_{s+1}]}, s)$ can be regarded as the utility function for segment $\boldsymbol{Y}_{(\kappa_s,\kappa_{s+1}]}$.

The estimator for the set of change points is given by

$$\widehat{\mathcal{K}} = \mathrm{argmax}_{\mathcal{K}\subseteq\mathcal{H}(m_I)} U(\mathcal{K}|\boldsymbol{Y}).$$

To optimize $U(\mathcal{K}|\boldsymbol{Y})$ over $\mathcal{H}(m_I)$ via dynamic programming, we require that $u(\boldsymbol{Y}_{(\kappa_s,\kappa_{s+1}]}, s)$ only depends on $\{\kappa_s, \kappa_{s+1}, s\}$ given $\mathcal{H}(m_I)$, so we replace $\overline{Y}_{i,\kappa_s}$ in Equation (6) by $\widetilde{Y}_{i,\kappa_s} = (\tau_l - \tau_{l-1})^{-1}\sum_{k\in(\tau_{l-1},\tau_l]} Y_{ik}$ with $\tau_l = \kappa_s, \kappa_s \in \mathcal{K}$. Note that $\widetilde{Y}_{i,\kappa_s}$ is the sample mean of the segment $(\tau_{l-1}, \tau_l]$, while $\overline{Y}_{i,\kappa_s}$ is the sample mean of the segment $(\kappa_{s-1}, \kappa_s]$. Both are consistent estimators of the mean of signals in the segment $(\kappa_{s-1}, \kappa_s]$. The dynamic programming procedure is presented as Algorithm A.1 in the Supplementary Materials, which has computational time $O(MN^2)$, in comparison with $O(MT^2)$ of the existing algorithms (Rigaill 2015; Du, Kao, and Kou 2016).

Algorithm 1 Screening Candidate Points

(i) For each $k \in [m_I, T-m_I]$, compute R_k.
(ii) If $R_k = \max\{R_j, j \in (k-m_I, k+m_I)\}$, then k is selected as a candidate point.

4. Theoretical Properties

In this section, we present the theoretical properties of the BHM method. Lemma 1 shows that $\mathcal{H}(m_I)$ covers $\mathcal{K}_0$ with probability one.

Lemma 1. Suppose that regularity conditions (1)–(2) in Section C of the supplementary materials hold. Let η_{is}^2 be the variance of Y_{ik} in the sth segment based on the true change points. Assume $m_I^{1/2}\delta_I/\eta \to \infty$, where $\eta = \max_{\{i=1,\ldots,n;s=0,\ldots,p_0\}}\eta_{is}$, δ_I is defined in condition (2) and $m_I/(\log(T))^{1+\epsilon} \to c > 0$ for $\epsilon > 0$. If $\min_{\{i=0,\ldots,p_0\}}(\kappa_{0,i+1}-\kappa_{0,i}) > m_I$, then for each $\kappa_{0j} \in \mathcal{K}_0$, there is a $\tau \in \mathcal{H}(m_I)$, such that $\Pr\{\kappa_{0j} \in (\tau-m_I, \tau+m_I)\} \to 1$ as $T \to \infty$.

The proof of Lemma 1 is given in Section C of the supplementary materials. The condition $m_I/(\log(T))^{1+\epsilon} \to c > 0$ regulates the selection of m_I, which grows no slower than $\log(T)$. Lemma 1 indicates that $\mathcal{H}(m_I)$ should cover $\mathcal{K}_0$ asymptotically, while the cardinality of $\mathcal{H}(m_I)$, $N+2$, is far less than T. Therefore, when performing the dynamic programming on the smaller set $\mathcal{H}(m_I)$, the algorithm guarantees a positive probability to select the true change points and at the same time improves the computational efficiency.

Furthermore, Theorem 1 shows that the screening step would accelerate the computational speed without sacrificing the statistical consistency. Let $\mathcal{K}_0 = \{\kappa_{01}, \dots, \kappa_{0p_0}\}$ be the true set of change points with $\min_{\{i=0,\dots,p_0\}}(\kappa_{0,i+1} - \kappa_{0,i}) > m_I$, and let $\widehat{\mathcal{K}} = \{\hat{\kappa}_1, \dots, \hat{\kappa}_{\hat{p}}\}$ be the estimated set of change points.

Theorem 1. Suppose that regularity conditions (1)–(4) in Section C of the supplementary materials hold. Assume $m_I^{1/2}\delta_I/\eta \to \infty$ and $m_I/(\log(T))^{1+\epsilon} \to c > 0$ for $\epsilon > 0$, and $\mathcal{K}_0$ is a subset of $\mathcal{H}(m_I)$. Then, as $T \to \infty$,

$$\hat{p} \xrightarrow{\mathcal{P}} p_0 \quad \text{and} \quad \sup_{b\in\mathcal{K}_0} \inf_{a\in\widehat{\mathcal{K}}} |a-b| = O_p(1).$$

The proof of Theorem 1 is delineated in Section C of the supplementary materials. Theorem 1 establishes the consistency of the estimated change points when $\mathcal{H}(m_I)$ covers $\mathcal{K}_0$. Combined with the result in Lemma 1 that each true change point falls in the m_I-neighborhood of at least one candidate point, we have

$$\hat{p} \xrightarrow{\mathcal{P}} p_0 \quad \text{and} \quad \sup_{b\in\mathcal{K}_0} \inf_{a\in\widehat{\mathcal{K}}} |a-b| = O_p(m_I).$$

Hence, Lemma 1 and Theorem 1 imply that as $T \to \infty$, the estimated change points are guaranteed to fall in the m_I-neighborhood of the true change points.

5. Simulation Studies

5.1. Simulation Settings

We conduct simulation experiments to evaluate the properties of the BHM method in comparison with two existing methods. First, we introduce the two main assessment metrics: the over-segmentation error,

$$d(\widehat{\mathcal{K}}|\mathcal{K}_0) = \sup_{b\in\mathcal{K}_0} \inf_{a\in\widehat{\mathcal{K}}} |a-b|,$$

and the under-segmentation error,

$$d(\mathcal{K}_0|\widehat{\mathcal{K}}) = \sup_{b\in\widehat{\mathcal{K}}} \inf_{a\in\mathcal{K}_0} |a-b|.$$

The over- or under-segmentation errors would be larger if we select fewer or more change points than the truth, respectively. The maximum segmentation error is $\max\{d(\widehat{\mathcal{K}}|\mathcal{K}_0), d(\mathcal{K}_0|\widehat{\mathcal{K}})\}$, and the estimation error for p_0 is $|\hat{p} - p_0|$.

The locations of change points are $\mathcal{K}_0 = \{\lfloor T \times r_i \rfloor\}_{i=1}^{p_0}$, with $p_0 = 10$ and

$$\{r_i\}_{i=1}^{10} = \{0.025, 0.155, 0.220, 0.365, 0.395, 0.495, 0.630, 0.725, 0.865, 0.975\}.$$

We adopt $n = 2$ for parameter tuning and $n = 8$ for comparing the BHM method with other methods. We define

$$\mathbf{g}(i,k) = \left[\frac{\{1+\operatorname{sgn}(k-\kappa_{0j})\}(1-\mathbb{I}_{\{j\in A_i\}})}{2}, j = 1,\dots,p_0\right]^\top, \quad i = 1,\dots,n; k = 1,\dots,T,$$

where $A_i = \{(3l + i) \mod p_0,\ l = 0, 1, 2\}$, $\operatorname{sgn}(\cdot)$ is the sign function and $\mathbb{I}_{\{\cdot\}}$ is the indicator function.

The data are generated with different dimensions from the base model,

$$Y_{ik} = 2 + \boldsymbol{d}^\top \mathbf{g}(i,k) + \xi_{ik} \prod_{j=1}^{\mathbf{1}_{p_0}{}^\top \mathbf{g}(i,k)} v_j, \quad i = 1,\dots,n; k = 1,\dots,T, \tag{7}$$

where $\boldsymbol{d} = (2.5, -2.8, 2.4, 2.6, -3, -2.9, 3.1, -2.5, -2.7, 2.6)^\top$ are the mean differences between consecutive segments, $\mathbf{1}_{p_0}$ is a p_0-dimensional vector of 1's, ξ_{ik}'s represent the errors and $[v_j]_{j=1}^{p_0}$ controls the homogeneity of the variances for different segments. If $[v_j]_{j=1}^{p_0} = \mathbf{1}_{p_0}$, the model yields a homogeneous variance across the segments.

By selecting distinct $[v_j]_{j=1}^{p_0}$ and different error distributions, we can obtain different data-generating models. The models used in the simulation studies are summarized as follows:

I. $[v_j]_{j=1}^{p_0} = \mathbf{1}_{p_0}$ and independent standard normal errors.
II. $[v_j]_{j=1}^{p_0} = (0.6, 2, 2/3, 0.6, 2, 2/3, 0.6, 2, 2/3, 0.6)$ and independent standard normal errors.
III. $[v_j]_{j=1}^{p_0} = \mathbf{1}_{p_0}$ and independent $t(5)$ errors with unit variance.
IV. $[v_j]_{j=1}^{p_0} = \mathbf{1}_{p_0}$ and independent skewed normal errors with slant parameter 4 and variance 1 (O'Hagan and Leonard 1976).
V. $[v_j]_{j=1}^{p_0} = \mathbf{1}_{p_0}$ and autocorrelated standard normal errors under a moving average (MA) model, $\xi_{ik} = a_k - 1.5a_{k-1}$ with $a_k \sim \mathrm{N}(0, 4/13)$.
VI. $[v_j]_{j=1}^{p_0} = \mathbf{1}_{p_0}$ and standard normal errors with correlated sequences of correlation coefficient 0.3.

The top panel of Figure 3 presents the mean of the simulated data under the base model (7) with $T = 400$ and $n = 8$. At some change points, the mean shifts only happen in certain sequences while the rest keep unchanged. For illustration, we also display the simulated sequences under model I in the bottom panel of Figure 3.

We choose $\alpha = 1, \sigma = 0.3$ in the prior π_k, as they yield the smallest maximal segmentation errors under models I and II as shown in Figure A.1 of the supplementary materials. We choose π_ω to be Gamma(1, 1).

5.2. Tuning Parameters

We use the simulation to find a suitable window size m_I for our method. Lemma 1 indicates that m_I should grow no slower than $\log(T)$. Hence, we select $m_I = \{\log(T)\}^{1.5} h$, where h is a constant to be determined numerically. Figure 4 exhibits the relationship between h and $|\hat{p} - p_0|$, and that between h and the maximum segmentation error under models I and II with $T = 400$ and $n = 2$, respectively. Clearly, $h = 0.55$ leads to the overall smallest errors for both models. In general, BHM works well for $h \in [0.5, 0.6]$.

We also compare the performances of using different priors for the mean difference under model I, including the normal prior ($\pi_{\mu,L}$) with $\psi = 2$, moment prior ($\pi_{\mu,M}$) with $\nu = 1$ and inverse moment prior ($\pi_{\mu,I}$) with $q = \phi = 2$ and $r = 0.6$.

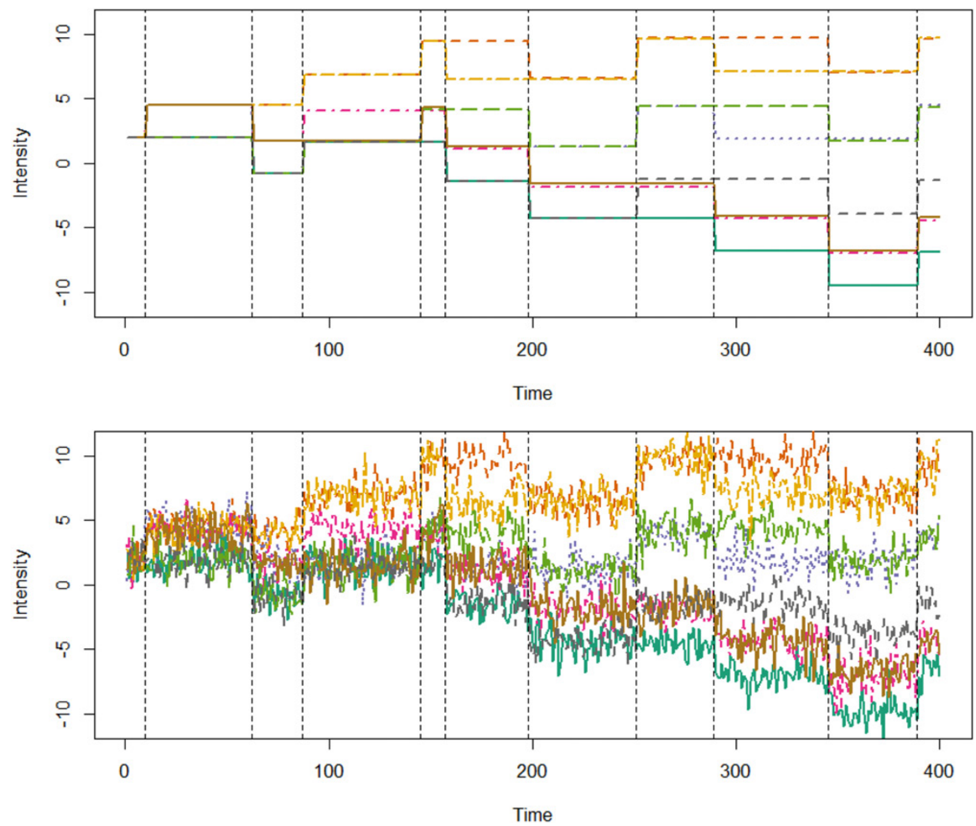

Figure 3. The mean of the simulated data (top) and randomly generated data under model I (bottom) with sample size $T = 400$ and $n = 8$. The vertical dashed lines indicate the true change points.

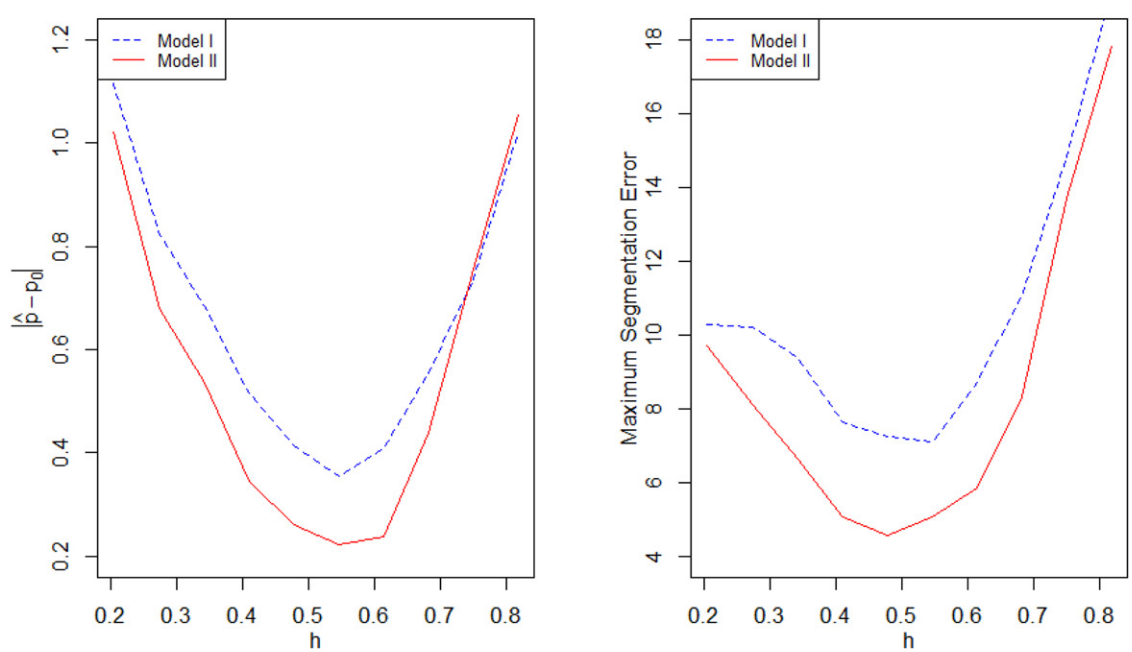

Figure 4. The absolute difference $|\hat{p} - p_0|$ (left) and the maximum segmentation error (right) versus h over 500 simulations with sample size $T = 400$, $n = 2$ under models I and II, respectively.

For fair comparisons, we select the best tuning parameters for each prior, so as to achieve the lowest segmentation error under model I, as shown in Figure A.2 (supplementary material).

Figure A.3 (supplementary material) shows the simulation results under model I with $n = 2$, where both $|\hat{p} - p_0|$ and the maximum segmentation error decrease as the sample size increases. Further, the two nonlocal priors have similar performances and both outperform the local prior by yielding smaller errors, especially when the sample size is larger than 350.

Table 1. Comparison of BHM-FIX, BHM-MPP, ECP and DPMLE when $n = 8$ under the six data-generating models over 500 simulations.

Data-generating		$\hat{p} - p_0$							Segmentation error	
Model	Method	≤ -3	-2	-1	0	1	2	≥ 3	$d(\widehat{\mathcal{K}}\|\mathcal{K}_0)$	$d(\mathcal{K}_0\|\widehat{\mathcal{K}})$
I	BHM-FIX	0	0	0	500	0	0	0	0.14 (0.38)	0.14 (0.38)
	BHM-MPP	0	0	0	500	0	0	0	0.08 (0.29)	0.08 (0.29)
	ECP	0	0	0	471	26	3	0	0.18 (0.43)	1.10 (4.11)
	DPMLE	500	0	0	0	0	0	0	78.32 (20.14)	0.08 (0.32)
II	BHM-FIX	0	0	0	500	0	0	0	0.05 (0.23)	0.05 (0.23)
	BHM-MPP	0	0	0	500	0	0	0	0.02 (0.15)	0.02 (0.13)
	ECP	0	0	0	477	21	2	0	0.08 (0.29)	0.12 (0.35)
	DPMLE	499	1	0	0	0	0	0	52.14 (1.97)	3.85 (0.36)
III	BHM-FIX	0	0	0	500	0	0	0	0.17 (0.40)	0.17 (0.40)
	BHM-MPP	0	0	0	500	0	0	0	0.11 (0.32)	0.11 (0.32)
	ECP	0	0	0	469	28	3	0	0.20 (0.45)	1.20 (4.41)
	DPMLE	500	0	0	0	0	0	0	72.02 (21.76)	0.12 (0.37)
IV	BHM-FIX	0	0	0	500	0	0	0	0.10 (0.31)	0.11 (0.34)
	BHM-MPP	0	0	0	500	0	0	0	0.10 (0.31)	0.09 (0.30)
	ECP	0	0	0	474	22	4	0	0.15 (0.40)	0.20 (0.45)
	DPMLE	500	0	0	0	0	0	0	75.60 (21.87)	5.80 (0.51)
V	BHM-FIX	0	0	0	500	0	0	0	0.10 (0.32)	0.11 (0.32)
	BHM-MPP	0	0	0	500	0	0	0	0.10 (0.27)	0.08 (0.27)
	ECP	0	0	0	500	0	0	0	0.15 (0.38)	0.17 (0.38)
	DPMLE	500	0	0	0	0	0	0	73.56 (21.39)	0.07 (0.26)
VI	BHM-FIX	0	0	0	486	14	0	0	1.02 (1.65)	1.60 (3.86)
	BHM-MPP	0	0	0	488	12	0	0	0.67 (0.83)	1.04 (2.57)
	ECP	0	0	0	476	20	4	0	0.45 (0.64)	1.25 (3.86)
	DPMLE	500	0	0	0	0	0	0	77.70 (22.67)	0.10 (0.31)

Note: Standard deviations are given in parentheses.

5.3. *Comparison With Other Methods*

We compare BHM with the ECP and DPMLE mean-change detection methods under all the six data-generating models with sample size $T = 400$ and $n = 8$. For the ECP and DPMLE methods, we adopt the default parameters from the original articles, respectively. The BHM method uses a normal likelihood for all models except for model III where a t-distribution likelihood is used. The moment prior $\pi_{\mu,M}$ is chosen for mean differences according to the results in Figure A.3. We adopt two methods to select parameters (α, σ, ν, h), as they are essential for BHM. (i) We set $(\alpha, \sigma, \nu, h) = (1, 0.3, 1, 0.55)$ as they deliver satisfactory performances in the simulation studies (denoted as BHM-FIX). (ii) We tune parameters (α, σ, ν, h) to maximize the posterior probability $\Pr(\mathcal{K}|\boldsymbol{Y})$ (denoted as BHM-MPP). Table 1 shows the results for $n = 8$ under the six models. The BHM-MPP consistently outperforms BHM-FIX under all the six models, indicating that maximization of the posterior probability is a more effective method to select hyper-parameters for the BHM method in practice. Nevertheless, the performance of BHM-FIX is only slightly inferior to that of BHM-MPP under all six settings, suggesting that the selected parameters in Section 5.2 achieve high estimation accuracy with small $|\widehat{p} - p_0|$ and segmentation errors. In practice, $(\nu, h) = (1, 0.55)$ can be set as the default parameters while the values of (α, σ) should be selected via the prior belief on the number of change points or the MPP method, as they rely on the number and locations of change points.

Overall, the proposed BHM-FIX and BHM-MPP are superior to other methods by yielding the smallest $|\widehat{p} - p_0|$ and segmentation errors under all the six models. Although incorrect models are adopted under models IV, V, and VI, the BHM-FIX and BHM-MPP still perform better than the competitive methods due to the robustness property.

6. Wind Turbine Data

For illustration, we apply our BHM method to detect the changes in the wind turbine data available from a wind turbine anomaly detection contest. For the details of the contest, refer to the link *http://www.caict.ac.cn/kxyj/qwfb/bps/201804/t20180426_158519.htm*). It includes a total of seven datasets which are manually labeled for the wind turbine failure times serving as the ground truth. Each dataset contains eight sequences ($n = 8$), corresponding to wind speed, cabin temperature, environment temperature, accelerated speed along horizontal and vertical directions and the ng5 temperatures from three pitches. In each dataset, there are 4–8 change points, that is, $p_0 \in [4, 8]$ in the sequences. The lengths of the sequences are from 400 to 1000, that is, $T \in [400, 1000]$. Under the moment prior, we choose (α, σ, ν, h) to yield the largest posterior probability $\Pr(\mathcal{K}|\boldsymbol{Y})$ under the BHM-MPP. We first use an outlier detection method introduced in Section A.3 of the supplementary materials to evaluate the data distribution. As there are significant amount of outliers in the data and thus the normal likelihood may not fit the data well, we adopt the t distribution to construct the likelihood in the BHM method. For comparison, we also implement other mean-change point detection methods, including ECP and DPMLE, and the popular pattern detection method called AutoPlait (Matsubara, Sakurai, and Faloutsos 2014).

We use three metrics to compare different methods, while considering an estimator within (or outside) the m_I-neighbourhood of a true change point as a true positive (or false positive) detection.

1. Precision (P): proportion of the estimated change points that are true change points. If the method yields no estimated change point, the precision is defined as 0.

Table 2. The running times based on the Intel core i7-7700K CPU and detection results of BHM-MPP, ECP, DPMLE and AutoPlait on seven wind turbine datasets, and the overall results are the average of those for the seven datasets.

Metrics	BHM-MPP	ECP	DPMLE	AutoPlait	BHM-MPP	ECP	DPMLE	AutoPlait
		Dataset 1				Dataset 2		
Time (sec)	161.1	21.9	435.2	2.8	32.9	6.3	58.6	1.2
Precision	0.625	0.316	0	0	0.500	0.300	0	1.000
Recall	0.833	1.000	0	0	0.750	0.750	0	0.250
F1 score	0.715	0.480	0	0	0.600	0.429	0	0.400
		Dataset 3				Dataset 4		
Time (sec)	31.2	5.2	58.6	2.2	28.2	7.2	58.7	1.8
Precision	0.429	0.300	0.444	1.000	0.333	0.214	0	0
Recall	0.750	0.750	1.000	0.250	0.750	0.750	0	0
F1 score	0.545	0.429	0.615	0.400	0.462	0.333	0	0
		Dataset 5				Dataset 6		
Time (sec)	70.3	13.3	156.2	2.8	30.3	5.8	58.8	2.6
Precision	0.375	0.176	0	0	0.875	0.667	0.727	0
Recall	0.750	0.750	0	0	0.875	1.000	1.000	0
F1 score	0.500	0.286	0	0	0.875	0.800	0.842	0
		Dataset 7				Overall		
Time (sec)	15.8	2.0	13.0	1.4	52.8	8.8	119.9	2.1
Precision	0.429	0.375	0.429	0	0.509	0.322	0.441	0.500
Recall	0.750	0.750	0.750	0	0.794	0.853	0.441	0.059
F1 score	0.545	0.500	0.545	0	0.620	0.468	0.441	0.105

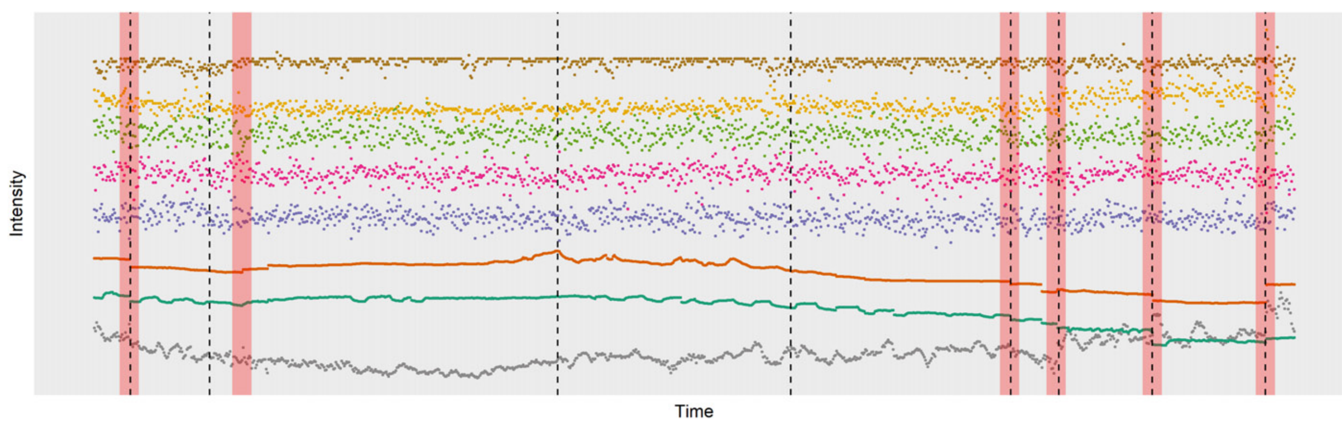

Figure 5. Detection results of the BHM-MPP method for the wind turbine dataset 1 with $n = 8$ sequences. The black dashed lines are the estimated state shift points and the red shadows are the m_l-neighborhood of the ground truth.

2. Recall (R): proportion of true change points detected by an algorithm.
3. F1 score: $\mathrm{F1} = 2\mathrm{RP}/(\mathrm{R} + \mathrm{P})$. When $\mathrm{R} = \mathrm{P} = 0$, the F1 score is defined as 0.

Table 2 shows the results from the BHM-MPP, ECP, DPMLE, AutoPlait methods on the seven datasets. Among the four methods, BHM-MPP yields the best performance in five out of seven datasets (i.e., datasets 1, 2, 4, 5, and 6) in terms of the F1 score. ECP yields a competitive recall but a much lower precision; DPMLE does not provide satisfactory precision or recall, especially for datasets 1, 2, 4, and 5, where the algorithm misses all the change points; AutoPlait barely captures any true change points in all the datasets. Figure 5 shows the eight data sequences of the wind turbine dataset 1 and the detection results using the BHM-MPP method. Compared with the results of other methods in Figure 2, the BHM-MPP method clearly yields the best performance.

We also report the running time of the four methods for each dataset based on the Intel core i7-7700k CPU in Table 2. While the AutoPlait leads to unsatisfactory performance, it consumes the shortest running time. Among the three mean-change detection methods, the nonparametric ECP is fastest. Due to the dynamic programming procedure, the computational burden of the BHM and DMPLE is relatively heavier, requiring longer computational time compared to the other two methods. However, it is worth noting that all AutoPlait, ECP, and DPMLE are implemented with C/C++ languages while our BHM method is implemented with the R language.

7. Conclusion

Motivated by the wind turbine data, we propose a BHM-based algorithm to detect mean changes for multivariate data sequences. Our method borrows the information across different sequences using the exchangeable random order prior. Furthermore, BHM reduces the detection errors by applying the nonlocal priors to the mean difference. It also eases the computational burden by employing an initial screening stage for selecting the candidate points. We show the asymptotic consistency of the proposed method from both theoretical and numerical perspectives. As an illustration, we apply the BHM method to detect the anomalies in the wind turbine data where the BHM method shows robust outcomes and yields the best performance in most of the datasets in terms of the F1 score compared with other competitive methods considered in the article.

Supplementary Materials

PDF file containing additional simulation results, the dynamic programming algorithm as well as the proofs of Lemma 1 and Theorem 1. *RCode_and_data*: Zip file containing the R code for implementing the BHM method as well as the real data used in the article in RData form.

References

Barry, D., and Hartigan, J. A. (1993), "A Bayesian Analysis for Change Point Problems," *Journal of the American Statistical Association*, 88, 309–319. [178]

Bellman, R., and Roth, R. (1969), "Curve Fitting by Segmented Straight Lines," *Journal of the American Statistical Association*, 64, 1079–1084. [177,181]

Bertolino, F., Racugno, W., and Moreno, E. (2000), "Bayesian Model Selection Approach to Analysis of Variance Under Heteroscedasticity," *Journal of the Royal Statistical Society*, Series D, 49, 495–502. [178]

Bouchard, D. and Badler, N. (2007), "Semantic Segmentation of Motion Capture Using Laban Movement Analysis," in *International Workshop on Intelligent Virtual Agents*, New York: Springer. [180]

Broderick, T., Jordan, M. I., and Pitman, J. (2013), "Cluster and Feature Modeling from Combinatorial Stochastic Processes," *Statistical Science*, 28, 289–312. [180]

Du, C., Kao, C. L. M., and Kou, S. C. (2016), "Stepwise Signal Extraction via Marginal Likelihood," *Journal of the American Statistical Association*, 111, 314–330. [177,178,179,181]

Fuentes-García, R., Mena, R. H., and Walker, S. G. (2019), "Modal posterior clustering motivated by Hopfield's network," *Computational Statistics & Data Analysis*, 137, 92–100. [178]

Gharghabi, S., Ding, Y., Yeh, C. C. M., Kamgar, K., Ulanova, L., and Keogh, E. (2017), "Matrix Profile VIII: Domain Agnostic Online Semantic Segmentation at Superhuman Performance Levels," *Proceedings - IEEE International Conference on Data Mining (ICDM)*, pp. 117–126. [180]

Gnedin, A., and Pitman, J. (2006), "Exchangeable Gibbs Partitions and Stirling Triangles," *Journal of Mathematical Sciences*, 138, 5674–5685. [180]

Gong, D., Medioni, G., and Zhao, X. (2014), "Structured Time Series Analysis for Human Action Segmentation and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 1414–1427. [180]

Gray, C. S. and Watson, S. J. (2010), "Physics of Failure Approach to Wind Turbine Condition Based Maintenance," *Wind Energy*, 13, 395–405. [178]

Hawkins, D. M. (2001), "Fitting Multiple Change-Point Models to Data," *Computational Statistics & Data Analysis*, 37, 323–341. [178]

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003), "The Changepoint Model for Statistical Process Control," *Journal of Quality Technology*, 35, 355–366. [178]

Hinkley, D. V. (1971), "Inference About the Change-Point From Cumulative Sum Tests," *Biometrika*, 58, 509–523. [178]

Jiang, F., Yin, G., and Dominici, F. (2018), "Bayesian Model Selection Approach to Boundary Detection With Non-Local Priors," in *Advances in Neural Information Processing Systems*, pp. 1978–1987. [179]

Johnson, V. E. and Rossell, D. (2010), "On the Use of Non-Local Prior Densities in Bayesian Hypothesis Tests," *Journal of the Royal Statistical Society*, Series B, 72, 143–170. [179]

Kim, K., Parthasarathy, G., Uluyol, O., Foslien, W., Sheng, S., and Fleming, P. (2011), "Use of SCADA Data for Failure Detection in Wind Turbines," in Energy Sustainability Conference and Fuel Cell Conference, Vol. 54686, Washington, D.C., August 7–10, 2011. [178]

Lau, J. W. and Green, P. J. (2007), "Bayesian Model-Based Clustering Procedures," *Journal of Computational and Graphical Statistics*, 16, 526–558. [179]

Lavielle, M. (2005), "Using Penalized Contrasts for the Change-Point Problem," *Signal Processing*, 85, 1501–1510. [178]

Lijoi, A., Mena, R. H., and Prünster, I. (2007), "Controlling the Reinforcement in Bayesian Non-Parametric Mixture Models," *Journal of the Royal Statistical Society)*, Series B, 69, 715–740. [180]

Maboudou-Tchao, E. M., and Hawkins, D. M. (2013), "Detection of Multiple Change-Points in Multivariate Data," *Journal of Applied Statistics*, 40, 1979–1995. [177]

Manogaran, G., and Lopez, D. (2018), "Spatial Cumulative Sum Algorithm With Big Data Analytics for Climate Change Detection," *Computers & Electrical Engineering*, 65, 207–221. [178]

Martínez, A. F., and Mena, R. H. (2014), "On a Nonparametric Change Point Detection Model in Markovian Regimes," *Bayesian Analysis*, 9, 823–858. [178,179,180]

Matsubara, Y., Sakurai, Y., and Faloutsos, C. (2014), "Autoplait: Automatic Mining of Co-Evolving Time Sequences," in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, New York: ACM. [177,180,184]

Matteson, D. S., and James, N. A. (2014), "A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data," *Journal of the American Statistical Association*, 109, 334–345. [177]

McCullagh, P., and Yang, J. (2008), "How Many Clusters?" *Bayesian Analysis*, 3, 1–19. [179]

Muggeo, V. M., and Adelfio, G. (2010), "Efficient Change Point Detection for Genomic Sequences of Continuous Measurements," *Bioinformatics*, 27, 161–166. [178]

Muñoz, C. Q. G., Jiménez, A. A., and Márquez, F. P. G. (2018), "Wavelet Transforms and Pattern Recognition on Ultrasonic Guides Waves for Frozen Surface State Diagnosis," *Renewable Energy*, 116, 42–54. [178]

O'Hagan, A., and Leonard, T. (1976), "Bayes Estimation Subject to Uncertainty About Parameter Constraints," *Biometrika*, 63, 201–203. [182]

Pitman, J. (1995), "Exchangeable and Partially Exchangeable Random Partitions," *Probability Theory and Related Fields*, 102, 145–158. [179,180]

——— (2002), "Combinatorial Stochastic Processes," Technical report, Dept. Statistics, UC Berkeley. [179,180]

Qiu, P. (2013), *Introduction to Statistical Process Control*, CRC Press. [178]

Rigaill, G. (2015), "A Pruned Dynamic Programming Algorithm to Recover the Best Segmentations With 1 to $K_{\max}$ Change-Points." *Journal de la Société Francaise de Statistique*, 156, 180–205. [178,181]

Rosner, B. (1983), "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, 25, 165–172. [180]

Tautz-Weinert, J. and Watson, S. J. (2016), "Using SCADA Data for Wind Turbine Condition Monitoring–A Review," *IET Renewable Power Generation*, 11, 382–394. [178]

Truong, C., Oudre, L., and Vayatis, N. (2018), "A Review of Change Point Detection Methods," arXiv:1801.00718. [178]

Tsiamyrtzis, P. and Hawkins, D. M. (2005), "A Bayesian Scheme to Detect Changes in the Mean of a Short-Run Process," *Technometrics*, 47, 446–456. [178]

——— (2008), "A Bayesian EWMA Method to Detect Jumps at the Start-up Phase of a Process," *Quality and Reliability Engineering International*, 24, 721–735. [178]

Wade, S., Ghahramani, Z. (2018), "Bayesian Cluster Analysis: Point Estimation and Credible Balls (With Discussion)," *Bayesian Analysis*, 13, 559–626. [179]

Wilkinson, M., Darnell, B., Van Delft, T., and Harman, K. (2014), "Comparison of Methods for Wind Turbine Condition Monitoring With SCADA Data," *IET Renewable Power Generation*, 8, 390–397. [178]

Yang, W., Court, R., and Jiang, J. (2013), "Wind Turbine Condition Monitoring by the Approach of SCADA Data Analysis," *Renewable Energy*, 53, 365–376. [178]

Yao, Y.-C. (1988), "Estimating the Number of Change-Points via Schwarz'Criterion," *Statistics & Probability Letters*, 6, 181–189. [178]

Yao, Y.-C., and Au, S.-T. (1989), "Least-Squares Estimation of a Step Function," *Sankhyā: The Indian Journal of Statistics, Series A*, 51, 370–381. [178]

Zamba, K., and Hawkins, D. M. (2006), "A Multivariate Change-Point Model for Statistical Process Control," *Technometrics*, 48, 539–549. [178]

Zhang, N. R., and Siegmund, D. O. (2007), "A Modified Bayes Information Criterion With Applications to the Analysis of Comparative Genomic Hybridization Data," *Biometrics*, 63, 22–32. [178]

Zou, C., Yin, G., Feng, L., and Wang, Z. (2014), "Nonparametric Maximum Likelihood Approach to Multiple Change-Point Problems," *The Annals of Statistics*, 42, 970–1002. [178]