Statistics and Its Interface Volume 13 (2020) 251–260

A variable selection approach to multiple change-points detection with ordinal data

Chi Kin Lam, Huaqing Jin, Fei Jiang, and Guosheng Yin*,†

Change-point detection has been studied extensively with continuous data, while much less research has been carried out for categorical data. Focusing on ordinal data, we reframe the change-point detection problem in a Bayesian variable selection context. We propose a latent probit model in conjunction with reversible jump Markov chain Monte Carlo to estimate both the number and locations of change-points with ordinal data. We conduct extensive simulation studies to assess the performance of our method. As an illustration, we apply the new method to detect changes in the ordinal data from the north Atlantic tropical cyclone record, which has an indication of global warming in the past decades.

Keywords and phrases: Latent variable, Multiple change-points, Ordinal data, Probit model, Reversible jump Markov chain Monte Carlo.

1. INTRODUCTION

Change-point detection for binary data was first considered in the seminal work of Hinkley and Hinkley [20], paving the way for categorical data analysis with potential change-points to emerge as an active research area. Cumulative sum-type methods for a sequence of binary responses first appeared in Pettitt [34], and were applied to multinomial responses in Wolfe and Chen [39]. Fu and Curnow [11] proposed the maximum likelihood approach for multiple change-points. As an alternative to frequentist methods, Bayesian approaches have also been developed for change-point problems. Chib [6] reformulated the change-point model by introducing a latent discrete state variable which indicates the regime that an observation is drawn from, which was generalized to multinomial data in Park [32]. A Bayesian hierarchical change-point model was developed by Carlin, Gelfand and Smith [4], where the Gibbs sampler is utilized for obtaining marginal posterior densities. Fearnhead [8] and Fearnhead and Liu [9] described algorithms that can simulate samples directly from the exact posterior distribution of the number and positions of change-points, thus avoiding the potential convergence problem in the commonly used Markov chain Monte Carlo (MCMC) methods. Applications of Bayesian change-point approaches to the study of ecology and literary style analysis can be found in Qian, Pan and King [35] and Girón, Ginebra and Riba [14], respectively.

With the growing interest in high-dimensional variable selection problems, we reformulate the change-point detection in the Bayesian variable selection framework [19]. Bayesian variable selection in probit models often relies upon data augmentation, where the probit regression is transformed into linear regression using latent normal variables [1]. By combining the stochastic search variable selection approach [12] with data augmentation, Lee et al. [28] introduced a method for selecting a suitable subset of regressors in the binary regression. For multinomial data, Sha et al. [38] developed a variable selection approach which utilizes the Metropolis–Hastings algorithm. However, the selection of the tuning parameters in the algorithm is often challenging. Significant progress in this direction was achieved by Holmes and Held [21], in which they presented an automated method without tuning. The data-augmentation based approaches tend to mix poorly due to correlation between the auxiliary variables and the regression coefficients. To circumvent such issues, Lamnisos, Griffin and Steel [27] extended the automatic generic sampler [15] to a binary regression model, in which no auxiliary variable is introduced.

Our research is motivated by analyzing north Atlantic tropical cyclone data between years 1851 and 2015, which may provide scientific evidence for climate changes in the past over 160 years [2, 7, 26, 37]. Global warming causes the sea surface temperature to rise [29], and storms tend to be more intense over warm water while losing their intensities as they veer over cool water or land [31]. Mann et al. [30] discovered that the annual storm counts had reached anomalous levels over the past two decades, and Knutson et al. [24] predicted that tropical cyclones would be stronger with higher wind speeds and heavier rains. Robbins et al. [36] investigated whether there exists any discontinuity (i.e., change-points) in the north Atlantic tropical cyclone record under the change-point detection framework.

We base our analysis of change-points on the data set HURDAT2 that can be directly downloaded from the National Oceanic Atmospheric Administration website. The top panel of Figure 1 shows the sequence of all recorded

*Corresponding author.
†ORCID: 0000-0003-3276-1392.

wind speed scales. With the advancement in measurement techniques [25], as well as the dramatic regime shift in the shelf ecosystems in the north Atlantic [17], it is believed that there exist inconsistencies (i.e., sudden changes) in the data. Robbins et al. [36] developed a test statistic for detecting a single change-point in a categorical data sequence and illustrated that there are two change-points corresponding to years 1930 and 1995.

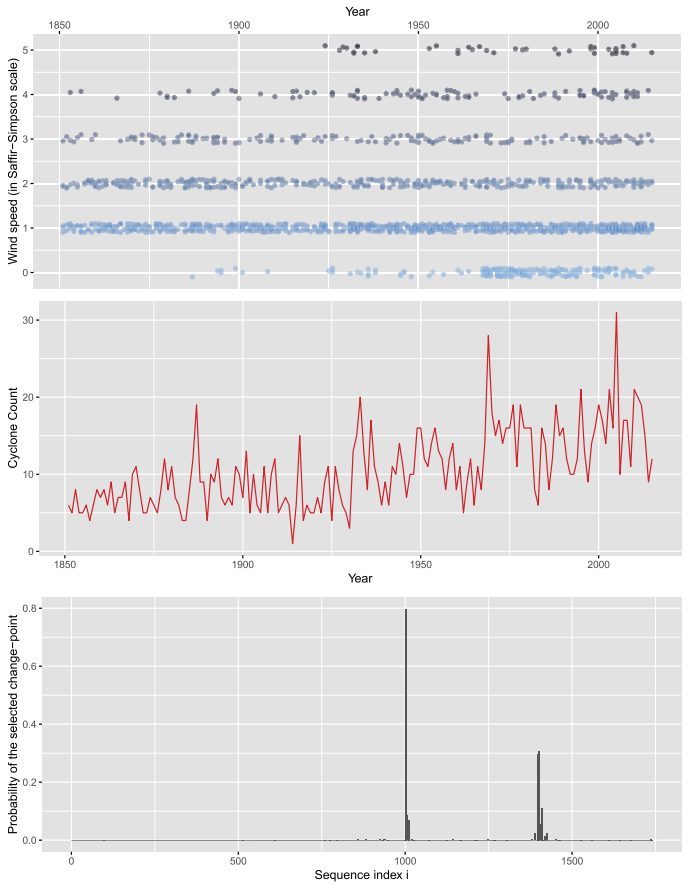

Figure 1. A sequence of the wind speed scales of the 1739 cyclones recorded in years 1851–2015 (top panel), the cyclone count (middle panel) and the posterior probabilities of the locations of the change-points in the HURDAT2 data (bottom panel).

The rest of the article is organized as follows. In Section 2, we describe the probit change-point models for binary and ordinal responses and reparameterize the original models such that the multiple change-point estimation is transformed into a Bayesian variable selection problem. Section 3 provides an introduction of the extended automatic generic sampler, and Section 4 presents simulation studies to illustrate the empirical behavior of the proposed approach. In Section 5, we apply our change-point detection method to the north Atlantic basin cyclone record data, and Section 6 concludes with some remarks.

2. METHODOLOGY

2.1 Probit model

We consider the problem of detecting multiple change-points in ordinal data using a probit model. Suppose that $\mathbf{y} = (y_1, \ldots, y_n)^\mathsf{T}$ is a data series in which each response y_i depends on a mean parameter β_i. Here, the value of β_i remains constant for $i \in [i_{k-1},\ i_k - 1]$, and only changes at K unknown locations $\{i_1, \ldots, i_K\}$, i.e., $\beta_i = \mu_k$ for $i_{k-1} \le i \le i_k - 1$, $k = 1, \ldots, K+1$, where $i_0 = 1$ and $i_{K+1} = n+1$.

Suppose the response y_i takes a value of 0 or 1. Under the probit model, we assume the probability of $y_i = 1$ to be

$$\Pr(y_i = 1) = 1 - \Pr(y_i = 0) = \Phi(\beta_i), \quad i = 1, \ldots, n, \tag{1}$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. If the responses are ordinal, i.e., if each y_i can take a value from the discrete choices $\{1, \ldots, C\}$, we assume that each y_i is generated from an ordinal probit model,

$$\begin{aligned} \Pr(y_i = c) \quad &= \quad \Phi(\tau_c - \beta_i) - \Phi(\tau_{c-1} - \beta_i), \\ &\qquad i = 1, \ldots, n; \quad c = 1, \ldots, C, \end{aligned}$$

where τ_c is the upper cutoff for $y_i = c$, and $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{C-1} < \tau_C = \infty$. For identification, we fix the first cutpoint $\tau_1 = 0$. Our goal is to detect the locations of the change-points $\{i_1, \ldots, i_K\}$ as well as estimating the true number of change-points K and the μ_k's.

2.2 From change-point detection to variable selection

We first reparameterize our change-point model as follows. Denote $\delta_1 = \beta_1$ and $\delta_i = \beta_i - \beta_{i-1}$ for $i = 2, \ldots, n$, so that

$$\beta_i = \sum_{j=1}^{i} \delta_j, \quad i = 1, \ldots, n.$$

For $k = 0, \ldots, K$, we have

$$\delta_i = \begin{cases} \mu_{k+1} - \mu_k, & \text{if } i = i_k, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mu_0 = 0$, i.e., if i_k is the location of a change-point, then $\delta_{i_k} \neq 0$. In addition, let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^\mathsf{T}$, $\boldsymbol{\delta} = (\delta_1, , \ldots, \delta_n)^\mathsf{T}$, and

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}_{n \times n}, \tag{2}$$

so that $\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\delta}$. With such reparametrization, we can transform the change-point detection problem into a variable selection one by identifying the nonzero regression coefficients

$\delta_i \neq 0$. In the probit regression framework, the matrix $\mathbf{X}$ can be viewed as a design matrix, and the parameter $\boldsymbol{\delta}$ can be interpreted as a vector of regression coefficients with the first element δ_1 representing the intercept. We also introduce a latent indicator vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^\mathsf{T}$ such that $\gamma_i = I(\delta_i \neq 0)$ where $I(\cdot)$ is an indicator function. Equivalently, if there is an $i' > 1$ such that $\gamma_{i'} = 1$, this indicates that i' is one of the change-points. Note that an intercept is included, and thus the value of γ_1 is fixed to be 1. For an indicator vector $\boldsymbol{\gamma}$, we denote the corresponding model as $\mathcal{M}_{\boldsymbol{\gamma}}$.

Under model $\mathcal{M}_{\boldsymbol{\gamma}}$, let $p_{\boldsymbol{\gamma}} = \sum_{i=2}^{n} \gamma_i$, i.e., the number of regressors contained in this model. Let $\{1, h_1, \ldots, h_{p_{\boldsymbol{\gamma}}}\}$ be the subindices corresponding to the nonzero elements in the vector $\boldsymbol{\gamma}$; for example, suppose that $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^\mathsf{T} = (1, 1, 0, 1)^\mathsf{T}$, then $p_{\boldsymbol{\gamma}} = 2$, $h_1 = 2$, and $h_2 = 4$. Define $\boldsymbol{\delta}_{\boldsymbol{\gamma}} = (\delta_1, \delta_{h_1}, \ldots, \delta_{h_{p_{\boldsymbol{\gamma}}}})^\mathsf{T}$, i.e., the vector of regression coefficients for those selected active covariates. In the Bayesian paradigm, we assign a multivariate normal prior distribution to $\boldsymbol{\delta}_{\boldsymbol{\gamma}}$,

$$\boldsymbol{\delta}_{\boldsymbol{\gamma}} \sim \mathrm{N}_{p_{\boldsymbol{\gamma}}+1}(\mathbf{0}, c_0^2 \mathbf{I}_{p_{\boldsymbol{\gamma}}+1}),$$

where c_0^2 is a large number and $\mathbf{I}_j$ is a $j \times j$ identity matrix. If $\mathbf{y}$ is an ordinal sequence, we adopt the transformation $\kappa_c = \log(\tau_c - \tau_{c-1})$, $c = 2, \ldots, C-1$, and denote $\boldsymbol{\kappa} = (\kappa_2, \ldots, \kappa_{C-1})^\mathsf{T}$. We then assign a normal prior distribution $\mathrm{N}(0, 10^2)$ to each κ_c, $c = 2, \ldots, C-1$. Finally, we assign a Bernoulli prior distribution to each element of $\boldsymbol{\gamma}$,

$$\pi(\gamma_i) = \pi_0^{\gamma_i}(1-\pi_0)^{1-\gamma_i}, \quad i = 2, \ldots, n,$$

so that each regressor is included in a model independently with a prespecified common probability π_0.

3. AUTOMATIC GENERIC SAMPLER

3.1 Reversible jump Markov chain Monte Carlo

Our goal is to select a suitable model $\mathcal{M}_{\boldsymbol{\gamma}}$ and estimate the corresponding parameter vector $\boldsymbol{\zeta}_{\boldsymbol{\gamma}}$, where $\boldsymbol{\zeta}_{\boldsymbol{\gamma}} = \boldsymbol{\delta}_{\boldsymbol{\gamma}}$ for binary responses and $\boldsymbol{\zeta}_{\boldsymbol{\gamma}} = (\tau_2, \ldots, \tau_{C-1}, \boldsymbol{\delta}_{\boldsymbol{\gamma}}^\mathsf{T})^\mathsf{T}$ for ordinal responses. We conduct the model searching using the automatic generic (AG) sampler, which can simultaneously explore both the model and parameter spaces.

It takes two stages to propose a new model $\mathcal{M}_{\boldsymbol{\gamma}'}$ from the current one $\mathcal{M}_{\boldsymbol{\gamma}}$. At each iteration g, we first draw a sample $m^{(g)}$ from a binomial distribution with the parameters $M-1$ and $\omega_0 \in (0,1)$, i.e., $m^{(g)} \sim \mathrm{Binomial}(M-1, \omega_0)$, where $M \in \mathbb{N}^+$ is a prespecified number indicating the maximum number of variables to be potentially changed from the current model, $\mathbb{N}^+$ refers to the set of all positive integers, and ω_0 is the probability that each variable is selected among the $M-1$ variables. At the second stage, we randomly select one move from the three possible ones: `Add`, `Delete` and `Swap`. If `Add` is selected, then a new model $\mathcal{M}_{\boldsymbol{\gamma}'}$ is constructed by adding $m^{(g)}+1$ new variables to the current model $\mathcal{M}_{\boldsymbol{\gamma}}$; if `Delete` is selected, we remove $m^{(g)}+1$ randomly selected regressors from the current model $\mathcal{M}_{\boldsymbol{\gamma}}$ and construct a new model $\mathcal{M}_{\boldsymbol{\gamma}'}$ with the $p_{\boldsymbol{\gamma}} - m^{(g)} - 1$ remaining regressors in $\mathcal{M}_{\boldsymbol{\gamma}}$; if `Swap` is selected, we randomly select $m^{(g)}+1$ regressors from those in $\mathcal{M}_{\boldsymbol{\gamma}}$ and replace them with another $m^{(g)}+1$ regressors from those excluded from $\mathcal{M}_{\boldsymbol{\gamma}}$. The number of regressors $p_{\boldsymbol{\gamma}'}$ in $\mathcal{M}_{\boldsymbol{\gamma}'}$ using these three moves are $p_{\boldsymbol{\gamma}} + m^{(g)} + 1$, $p_{\boldsymbol{\gamma}} - m^{(g)} - 1$ and $p_{\boldsymbol{\gamma}}$, respectively. According to this scheme, the probability $q(\boldsymbol{\gamma}'|\boldsymbol{\gamma})$ of proposing to move from $\mathcal{M}_{\boldsymbol{\gamma}}$ to $\mathcal{M}_{\boldsymbol{\gamma}'}$ is

$$q(\boldsymbol{\gamma}'|\boldsymbol{\gamma}) = \binom{M-1}{m^{(g)}} \omega_0^{m^{(g)}} (1-\omega_0)^{M-1-m^{(g)}} \times \begin{cases} \dfrac{1}{\binom{p_{\boldsymbol{\gamma}}}{m^{(g)}+1}}, & \text{if } p_{\boldsymbol{\gamma}} > p_{\boldsymbol{\gamma}'} \text{ and } n-1-p_{\boldsymbol{\gamma}} < m^{(g)}+1 \leq p_{\boldsymbol{\gamma}}, \\ \dfrac{1}{\binom{n-1-p_{\boldsymbol{\gamma}}}{m^{(g)}+1}}, & \text{if } p_{\boldsymbol{\gamma}} < p_{\boldsymbol{\gamma}'} \text{ and } p_{\boldsymbol{\gamma}} < m^{(g)}+1 \leq n-1-p_{\boldsymbol{\gamma}}, \\ \dfrac{1}{3\binom{p_{\boldsymbol{\gamma}}}{m^{(g)}+1}}, & \text{if } p_{\boldsymbol{\gamma}} > p_{\boldsymbol{\gamma}'} \text{ and } m^{(g)}+1 \leq \min(p_{\boldsymbol{\gamma}}, n-1-p_{\boldsymbol{\gamma}}), \\ \dfrac{1}{3\binom{n-1-p_{\boldsymbol{\gamma}}}{m^{(g)}+1}}, & \text{if } p_{\boldsymbol{\gamma}} < p_{\boldsymbol{\gamma}'} \text{ and } m^{(g)}+1 \leq \min(p_{\boldsymbol{\gamma}}, n-1-p_{\boldsymbol{\gamma}}), \\ \dfrac{\binom{p_{\boldsymbol{\gamma}}}{m^{(g)}+1}^{-1}}{3\binom{n-1-p_{\boldsymbol{\gamma}}}{m^{(g)}+1}}, & \text{if } p_{\boldsymbol{\gamma}} = p_{\boldsymbol{\gamma}'} \text{ and } m^{(g)}+1 \leq \min(p_{\boldsymbol{\gamma}}, n-1-p_{\boldsymbol{\gamma}}). \end{cases}$$

The probability $q(\boldsymbol{\gamma}|\boldsymbol{\gamma}')$ of proposing to move from $\mathcal{M}_{\boldsymbol{\gamma}'}$ to $\mathcal{M}_{\boldsymbol{\gamma}}$ can be obtained by exchanging $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ accordingly.

Once a new model $\mathcal{M}_{\boldsymbol{\gamma}'}$ is selected, we propose a new parameter $\boldsymbol{\zeta}_{\boldsymbol{\gamma}'}$ based on the current parameter $\boldsymbol{\zeta}_{\boldsymbol{\gamma}}$. When $p_{\boldsymbol{\gamma}'} > p_{\boldsymbol{\gamma}}$, a new random vector $\mathbf{u}_{\boldsymbol{\gamma}}$ of length $p_{\boldsymbol{\gamma}'} - p_{\boldsymbol{\gamma}}$ is generated from a multivariate distribution with probability density function $q_{\boldsymbol{\gamma}}(\mathbf{u}_{\boldsymbol{\gamma}})$. We sample $\mathbf{u}_{\boldsymbol{\gamma}}$ from a multivariate normal distribution $\mathrm{N}_{p_{\boldsymbol{\gamma}'}-p_{\boldsymbol{\gamma}}}(\mathbf{0}, \mathbf{I}_{p_{\boldsymbol{\gamma}'}-p_{\boldsymbol{\gamma}}})$. Let $\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}$ denote the maximum likelihood estimators of the parameter vector $\boldsymbol{\zeta}_{\boldsymbol{\gamma}}$ and the covariance matrix respectively [22, 23]. Let $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{1/2}$ be the lower triangular matrix of the Cholesky decomposition of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}$, so that $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{1/2}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{1/2})^\mathsf{T}$. Denote $\boldsymbol{\nu}_{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{1/2})^{-1}(\boldsymbol{\zeta}_{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}})$. Under the new model $\mathcal{M}_{\boldsymbol{\gamma}'}$, we set the parameter vector $\boldsymbol{\zeta}_{\boldsymbol{\gamma}'}$ to be

$$\boldsymbol{\zeta}_{\boldsymbol{\gamma}'} = \begin{cases} \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}'} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}'}^{1/2}(\boldsymbol{\nu}_{\boldsymbol{\gamma}})_1^{p_{\boldsymbol{\gamma}'}+1}, & \text{if } p_{\boldsymbol{\gamma}'} < p_{\boldsymbol{\gamma}}, \\ \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}'} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}'}^{1/2}\boldsymbol{\nu}_{\boldsymbol{\gamma}}, & \text{if } p_{\boldsymbol{\gamma}'} = p_{\boldsymbol{\gamma}}, \\ \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}'} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}'}^{1/2}(\boldsymbol{\nu}_{\boldsymbol{\gamma}}^\mathsf{T}, \mathbf{u}_{\boldsymbol{\gamma}}^\mathsf{T})^\mathsf{T}, & \text{if } p_{\boldsymbol{\gamma}'} > p_{\boldsymbol{\gamma}}, \end{cases}$$

where $(\cdot)_1^m$ denotes the first m components of a vector. The acceptance probability of moving to model $\boldsymbol{\gamma}'$ is then

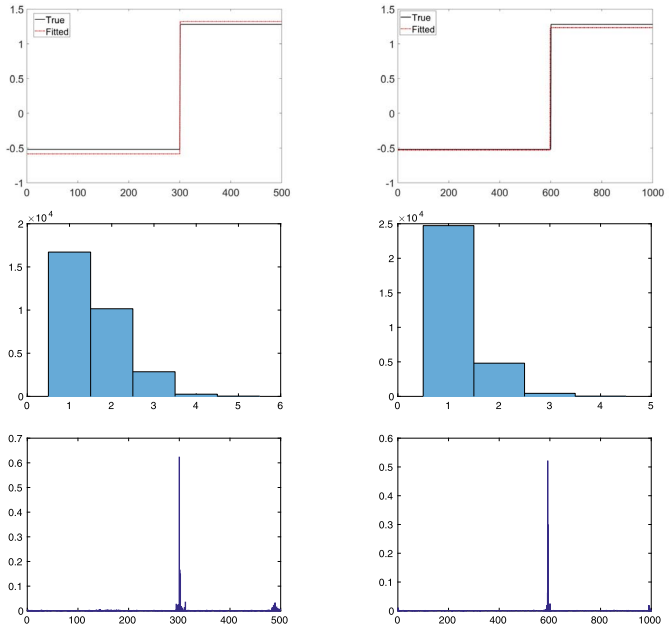

Figure 2. Simulation results under scenario 1 (binary data with one change-point) with sample size $n = 500$ (left) and $n = 1000$ (right). For each sample size, we present the fitted curve in comparison with the true curve (top panel), the posterior histogram of the number of change-points (middle panel) and the posterior probability of the location of the change-point (bottom panel).

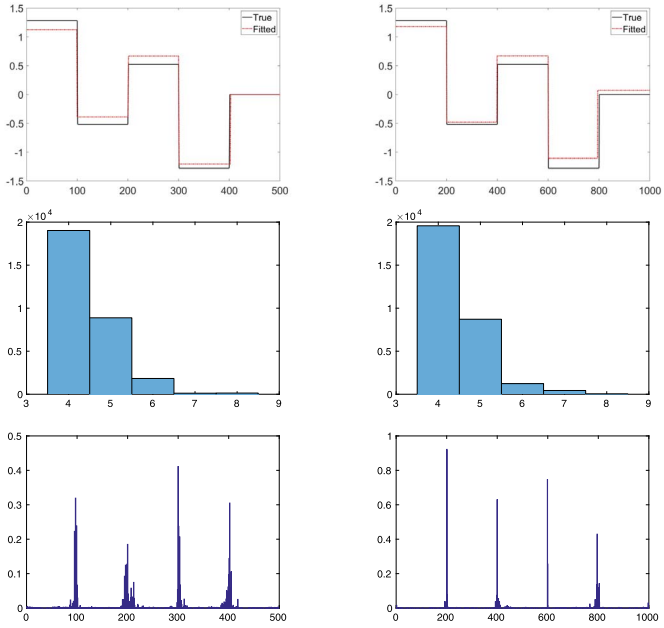

Figure 3. Simulation results under scenario 2 (binary data with four change-points) with sample size $n = 500$ (left) and $n = 1000$ (right). For each sample size, we present the fitted curve in comparison with the true curve (top panel), the posterior histogram of the number of change-points (middle panel) and the posterior probabilities of the locations of four change-points (bottom panel).

$\min(1, A_{\boldsymbol{\gamma}\to\boldsymbol{\gamma}'})$, where

$$A_{\boldsymbol{\gamma}\to\boldsymbol{\gamma}'} = \frac{\pi(\boldsymbol{\zeta}_{\boldsymbol{\gamma}'}, \boldsymbol{\gamma}'|\mathbf{y})q(\boldsymbol{\gamma}|\boldsymbol{\gamma}')|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}'}^{1/2}|}{\pi(\boldsymbol{\zeta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}|\mathbf{y})q(\boldsymbol{\gamma}'|\boldsymbol{\gamma})|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{1/2}|} \times \begin{cases} q_{\boldsymbol{\gamma}'}(\mathbf{u}_{\boldsymbol{\gamma}'}), & \text{if } p_{\boldsymbol{\gamma}'} < p_{\boldsymbol{\gamma}}, \\ 1, & \text{if } p_{\boldsymbol{\gamma}'} = p_{\boldsymbol{\gamma}}, \\ q_{\boldsymbol{\gamma}}(\mathbf{u}_{\boldsymbol{\gamma}})^{-1}, & \text{if } p_{\boldsymbol{\gamma}'} > p_{\boldsymbol{\gamma}}. \end{cases}$$

3.2 Posterior inference

After the iterations are completed using the AG sampler, N posterior samples are collected, which form the basis for inference. Denote the jth posterior indicator vector and posterior sample from the AG sampler for the parameters $\boldsymbol{\beta}$ as $\boldsymbol{\gamma}^{(j)}$ and $\boldsymbol{\beta}^{(j)}$ respectively, with $\boldsymbol{\beta}^{(j)} = (\beta_{(1)}^{(j)}, \ldots, \beta_{(n)}^{(j)})^\top$. We first compute the posterior inclusion probabilities, $(p_1^{\text{post}}, \ldots, p_n^{\text{post}})^\top$, for each point, and then obtain the estimated change-points via the non-maximum suppression (NMS) procedure described as follows.

- Determine a window size h and a threshold probability p.
- For $j = h, h+1, \ldots, n-h$, if $p_j^{\text{post}} = \max\{p_l^{\text{post}}, l \in (j-h, j+h]\} \geq p$, then j is selected as an estimated change-point.

In our simulations and real data application, it is found that our method is not sensitive to the choices of h and p. For simplicity, we choose $h = 30$ and $p = 0.1$ as default. After implementation of the NMS step, an estimated model $\mathcal{M}_{\hat{\boldsymbol{\gamma}}}$ is obtained. Let $\{\hat{i}_0, \hat{i}_1, \ldots, \hat{i}_{\hat{K}}, \hat{i}_{\hat{K}+1}\}$ be the set of estimated change-points, where $\hat{i}_0 = 1$ and $\hat{i}_{\hat{K}+1} = n+1$. We estimate the parameters μ_k for segment k as the posterior median of the corresponding samples, i.e., $\hat{\mu}_k = \text{Median}\left(\{\beta_{(s)}^{(j)}\}_{s=i_{k-1},\ldots,i_k-1}^{j=1,\ldots,N}\right)$, and the 95% highest posterior density (HPD) interval can be obtained accordingly. The cutoff points $\{\tau_2, \ldots, \tau_{C-1}\}$ can be estimated simply based on the posterior samples, as the number of cutoff points remains the same under difference models.

4. SIMULATION STUDY

To investigate the performance of the proposed approach, we conduct simulation studies for both binary and ordinal data. The parameters used in Sections 4 and 5 are $c_0 = 5$, $M = 3$, $\omega_0 = 0.25$, and π_0 is chosen such that the mean model size is 5, which are the default parameter values of the MATLAB codes in the supplemental materials of Lamnisos, Griffin and Steel [27]. More detailed discussion on π_0 and c_0 is given in Section 5. For each scenario, we run the AG sampler for 43000 iterations with the first 13000 draws discarded.

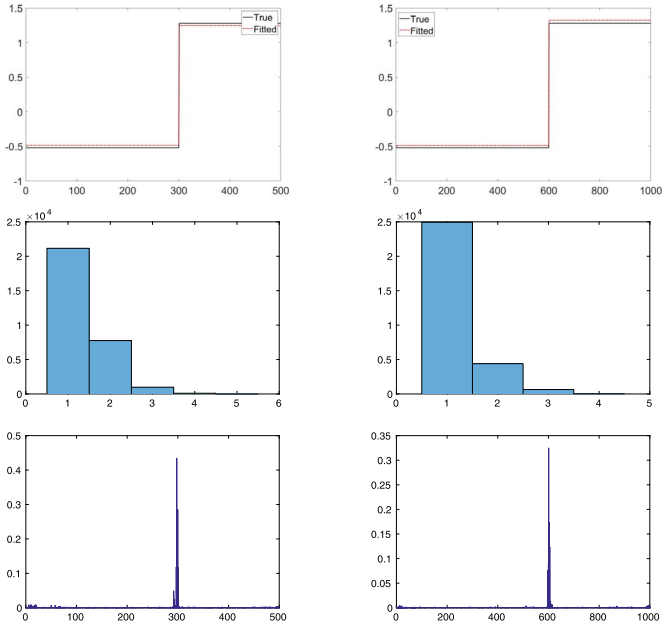

Figure 4. Simulation results under scenario 3 (ordinal data with four categories and one change-point) with sample size $n = 500$ (left) and $n = 1000$ (right). For each sample size, we present the fitted curve in comparison with the true curve (top panel), the posterior histogram of the number of change-points (middle panel) and the posterior probability of the location of the change-point (bottom panel).

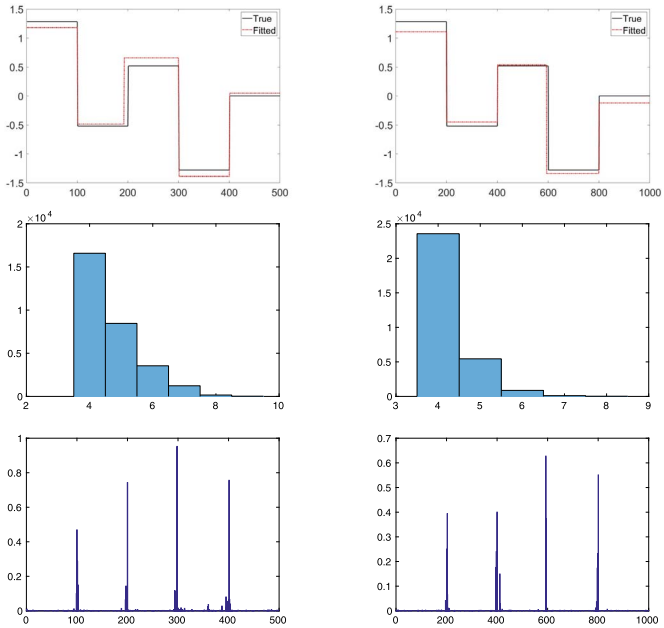

Figure 5. Simulation results under scenario 4 (ordinal data with four categories and four change-points) with sample size $n = 500$ (left) and $n = 1000$ (right). For each sample size, we present the fitted curve in comparison with the true curve (top panel), the posterior histogram of the number of change-points (middle panel) and the posterior probabilities of the locations of four change-points (bottom panel).

We first consider a sequence of binary observations $\{y_i\}$ where $y_i \sim \text{Bernoulli}\{\Phi(\beta_i)\}$ as in (1), and

$$\beta_i = \begin{cases} \mu_1, & 1 \le i \le 0.2n, \\ \mu_2, & 0.2n < i \le 0.4n, \\ \mu_3, & 0.4n < i \le 0.6n, \\ \mu_4, & 0.6n < i \le 0.8n, \\ \mu_5, & 0.8n < i \le n. \end{cases}$$

In scenario 1, the parameter vector is $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (-0.52, -0.52, -0.52, 1.28, 1.28)$, which contains a single change-point at $i = 0.6n + 1$. In scenario 2, the parameter vector is $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (1.28, -0.52, 0.52, -1.28, 0.00)$, which contains four change-points at $i = 0.2n+1, 0.4n+1, 0.6n+1, 0.8n+1$ respectively. Under each scenario, sample sizes $n = 500$ and $n = 1000$ are considered. Figures 2 and 3 present the posterior median estimates of β_i, $i = 1, \ldots, n$, which match well with their true values. The histograms show that the distributions of the number of change-points are centered around the true values 1 and 4, respectively. The posterior inclusion probabilities of each regressor $\mathbf{x}_i$, where $\mathbf{x}_i$ is the ith column of $\mathbf{X}$ in (2), are also presented in Figures 2 and 3. Under these two scenarios, we can see that the regressors at the locations of change-points have significantly higher inclusion probabilities than the others. With the increase of sample size from $n = 500$ to $n = 1000$, the posterior inclusion probabilities tend to be more concentrated on these regressors.

In scenarios 3 and 4, we consider that each y_i is an ordinal response. We simulate y_i with probabilities

$$\begin{aligned} \Pr(y_i = c) &= \Phi(\tau_c - \beta_i) - \Phi(\tau_{c-1} - \beta_i), \\ & \quad i = 1, \ldots, n; \ c = 1, \ldots, C, \end{aligned}$$

where $C = 4$, $\tau_2 = 0.25$ and $\tau_3 = 0.75$. The values of $\boldsymbol{\beta}$ and the sample sizes n in scenarios 3 and 4 are the same as those in scenarios 1 and 2. Compared with scenarios 1 and 2, these two scenarios are more challenging due to the increase of response categories. Figures 4 and 5 show the estimation results by the proposed method, which illustrate that even though two additional cutpoint parameters are involved in the iteration procedure, it can still provide reliable inference on the true values of $\boldsymbol{\beta}$ as well as the number and locations of change-points.

For each scenario, we adopt the AG sampler to obtain N posterior samples, and then carry out the posterior inference according to the methods described in Section 3.2. We calculate the bias and the posterior standard deviation as well as the 95% HPD interval. Table 1 presents the numerical results under the four scenarios with sample sizes

Table 1. The biases, standard deviations (SD), 95% highest posterior density (HPD) intervals of the model parameters and the deviations of the estimated locations of the change-points from the truth under each scenario with sample size $n = 500$ and $n = 1000$

n	Scenario	Parameter	Bias	SD	95% HPD	Deviation
500	1	μ_1	−0.07	0.20	(−0.77, −0.42)	–
		μ_2	−0.04	0.48	(1.00, 1.60)	0
	2	μ_1	−0.16	0.33	(0.77, 1.51)	–
		μ_2	0.13	0.23	(−0.73, −0.08)	−1
		μ_3	0.15	0.29	(0.22, 1.21)	0
		μ_4	0.08	0.46	(−1.74, −0.74)	0
		μ_5	0.00	0.25	(−0.28, 0.28)	2
	3	μ_1	0.03	0.17	(−0.66, −0.31)	–
		μ_2	−0.04	0.13	(1.04, 1.46)	0
		τ_2	−0.01	0.04	(0.16, 0.32)	–
		τ_3	−0.04	0.07	(0.59, 0.85)	–
	4	μ_1	−0.09	0.26	(0.85, 1.53)	–
		μ_2	0.03	0.22	(−0.83, −0.16)	0
		μ_3	0.13	0.37	(0.34, 1.11)	−8
		μ_4	−0.11	0.63	(−1.79, −0.87)	0
		μ_5	0.05	0.23	(−0.20, 0.30)	0
		τ_2	0.04	0.04	(0.21, 0.38)	–
		τ_3	−0.02	0.06	(0.62, 0.86)	–
1000	1	μ_1	−0.01	0.10	(−0.67, −0.39)	–
		μ_2	−0.04	0.17	(0.99, 1.47)	−3
	2	μ_1	−0.10	0.29	(0.69, 1.67)	–
		μ_2	0.04	0.23	(−0.68, −0.28)	0
		μ_3	0.15	0.24	(0.46, 0.87)	−2
		μ_4	0.17	0.18	(−1.36, −0.85)	0
		μ_5	0.07	0.16	(−0.13, 0.31)	−6
	3	μ_1	0.03	0.08	(−0.59, −0.38)	–
		μ_2	0.04	0.15	(1.17, 1.47)	0
		τ_2	0.01	0.03	(0.21, 0.32)	–
		τ_3	0.00	0.05	(0.66, 0.84)	–
	4	μ_1	−0.17	0.13	(0.91, 1.30)	–
		μ_2	0.07	0.15	(−0.65, −0.27)	1
		μ_3	0.02	0.12	(0.36, 0.72)	0
		μ_4	−0.06	0.23	(−1.62, −1.08)	−7
		μ_5	−0.12	0.11	(−0.29, 0.06)	−1
		τ_2	−0.01	0.03	(0.19, 0.30)	–
		τ_3	−0.01	0.05	(0.66, 0.83)	–

$n = 500$ and $n = 1000$. We can conclude that the estimation of the parameters and locations of the change-points is satisfactory. In addition, both the standard deviation and the length of the 95% HPD interval of each parameter decrease with the increase of sample size n. The convergence of the Markov chain appears to be satisfactory based on Geweke's statistics [13].

5. ATLANTIC BASIN CYCLONE DATA

We apply the proposed change-point detection method to the HURDAT2 data set, which is available from the National Oceanic Atmospheric Administration's website. The data set is composed of information on 1739 Atlantic basin cyclones between years 1851 and 2015. We classify each storm according to the Saffir–Simpson scale, which is a common practice for meteorologists. The storms with wind speeds of 40–73 mph, 74–95 mph, 96–110 mph, 111–130 mph, 131–156 mph, and above 156 mph are categorized as the Saffir–Simpson categories 0, 1, 2, 3, 4, and 5, respectively. A plot of the categorical scales of the storms is provided in the top panel of Figure 1. We also show a time series plot of cyclone count per year from 1851 to 2015 in the middle panel of Figure 1, which indicates an obvious increasing trend of the cyclone count per year. However, as we focus on the categories of the cyclones, we do not take the number of cyclones of each year into consideration.

In all the experiments, we draw 43000 posterior samples using the AG sampler with the first 13000 iterations dis-

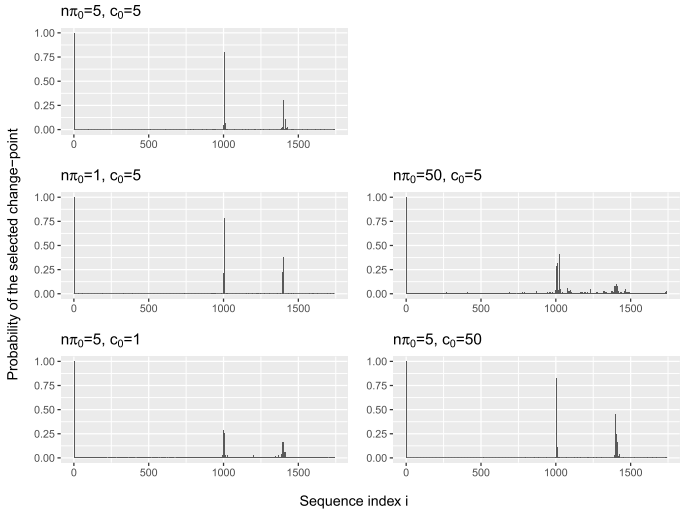

Figure 6. Inclusion probability of each point under different values of π_0 and c_0.

carded as burn-ins, which takes about two and half hours using our Matlab code on a computer with Intel Core i7-4770 CPU and 8GB RAM. By default, we choose $c_0 = 5$ and $n\pi_0 = 5$, while four other configurations are also explored: small or large model sizes with $n\pi_0 = 1$ or 50, and informative or non-informative priors with $c_0 = 1$ or 50. Figure 6 shows that under all the five hyper-parameters settings the inclusion probabilities achieve peaks at around the 1000th and the 1400th observations. However, if the prior expected model size deviates far from the truth (e.g., under $n\pi_0 = 50$ and $c_0 = 5$), or the prior of $\boldsymbol{\delta}_{\boldsymbol{\gamma}}$ is informative (e.g., under $n\pi_0 = 5$ and $c_0 = 1$), the peaks are more dispersive and relatively lower, and the 95% HPD intervals for parameters become wider compared with the other three settings. Under a suitable expected model size and a non-informative prior for $\boldsymbol{\delta}_{\boldsymbol{\gamma}}$, the results are very similar. In practice, we may choose a large c_0 to induce a non-informative prior, and $n\pi_0$ in the same scale of the true model size which can be experimented through preliminary data analysis.

To evaluate the convergence of the posterior samples, we use the method of Geweke [13] which is based on a test for equality of the means of the first and last portions of a Markov chain. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic asymptotically follows the standard normal distribution. Since the number of parameters μ_i's changes under different models, in our diagnostic test we mainly focus on the number of change-points K and the parameters of cutoff points $\{\tau_c\}_{c=2}^{C-1}$. We compute Geweke's statistics of $\{\tau_c\}_{c=2}^{C-1}$ and K based on 30000 posterior samples and the resultant p-values are all larger than 0.1, demonstrating that the Markov chain achieves stationarity as expected.

Table 2. The posterior medians, standard errors (SE) and 95% highest posterior density (HPD) intervals of the model parameters in the analysis of the HURDAT2 data set

Parameter	Median	SE	95% HPD
μ_1	1.44	0.05	(1.32, 1.53)
μ_2	0.65	0.08	(0.52, 0.77)
μ_3	1.23	0.08	(1.09, 1.36)
τ_2	1.51	0.05	(1.42, 1.57)
τ_3	2.19	0.06	(2.08, 2.27)
τ_4	2.64	0.06	(2.54, 2.75)
τ_5	3.28	0.08	(3.11, 3.44)

Among a total of 30000 posterior samples, 96.8% of them indicate existence of two change-points, 3.0% indicate three change-points, and the remaining indicate four change-points. The marginal inclusion probabilities of the regressors in the bottom panel of Figure 1 demonstrate that one change-point is at the 1006th observation (year 1967) and the other is at the 1404th observation (year 1995). The estimates of the parameters are presented in Table 2.

As shown in Figure 7, we further calculate the probability of a storm being classified into each category in the three detected segments. Figure 8 exhibits the time series plots of posterior medians of $\Pr(y_i = c)$ for each class, for which the computation is based on a full posterior manner that accounts for the uncertainty in the total number of change-points. From these figures, we can see that the estimated probability of a storm to belong to category 0 in 1967–1995 is substantially larger than those in the other two segments. This is due to an unphysical increase of the record of these storms, which may be the consequence of the progressive technology and analysis protocols in the meteorological satellite era beginning in the 1960s [25]. Prior to the 1960s, a tropical cyclone could only be detected when it was physically encountered by humans either at sea or on land.

We proceed to compare the estimation results in the 1967–1995 and 1995–2015 segments. We can deduce from Figures 7 and 8 that the estimated probabilities of the storms with categories higher than 1 in the recent two decades have increased significantly. The top panel and the middle panel of Figure 1 demonstrate that the number of the storms increased between 1995 and 2015, while the number of category 0 storms has declined significantly compared with the counterpart in the 1967–1995 segment. A possible explanation for this phenomenon is the dramatic regime shift in the shelf ecosystems of the northwest Atlantic, perhaps due to natural environmental variability and human activities, e.g., rampant overfishing in this region [10, 16, 17].

6. CONCLUDING REMARKS

To detect potential changes in the Atlantic basin cyclone record, we have developed a Bayesian approach for the analysis of probit models with multiple change-points. By in-

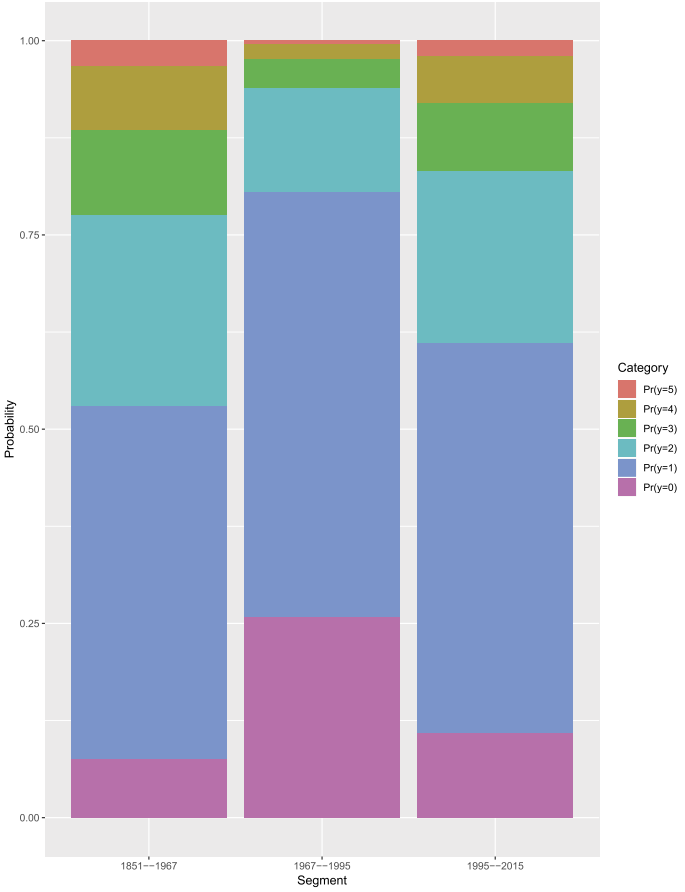

Figure 7. Estimated probabilities of the six wind speed categories in the three detected segments.

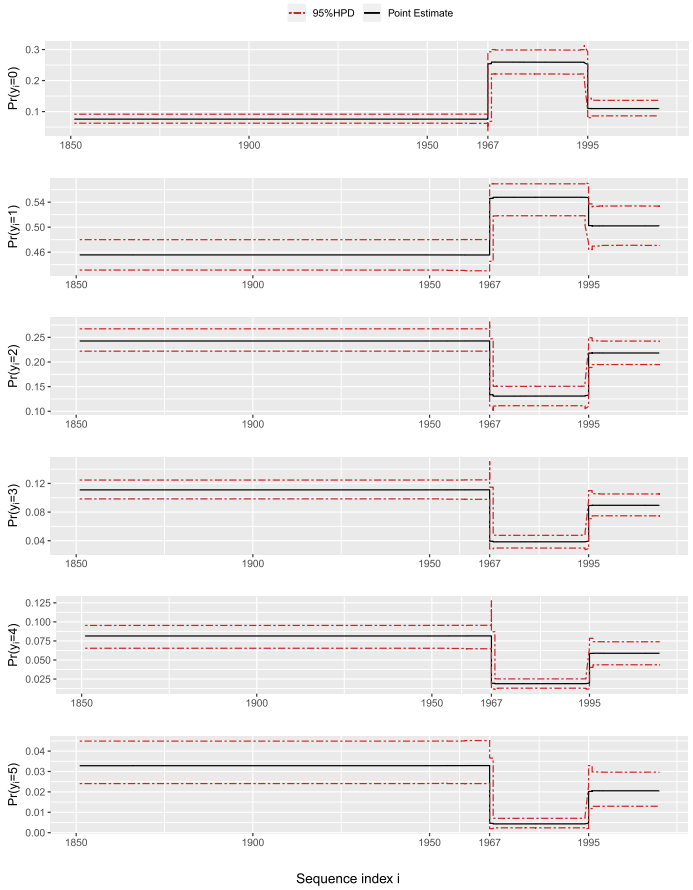

Figure 8. Time series plots of the posterior probability $\Pr(y_i = c)$ *for each class.*

troducing a virtual regressor on each observation, we reformulate the original change-point detection problem into a Bayesian variable selection one and tackle it by using a reversible-jump Markov chain Monte Carlo algorithm. The objective is to simultaneously obtain a sparse vector with the nonzero components matching the locations of the change-points and make inference in each of the detected segments. With the algorithm developed in this paper, we identify two change-points in Atlantic basin hurricane record circa 1967 and 1995, which is consistent with the findings in Robbins et al. [36] based on the frequentist cumulative sum-type method. We thus offer evidence in support of the recent increase in the frequency of intense storms from a Bayesian perspective.

There are several possible future directions for this work. Along with the probit model, the approach proposed in this paper can be easily applied to the change-point analysis in other types of generalized linear models *mutatis mutandis*, for which the logit model for multinomial data and the Poisson log-linear model for count data are natural candidates. With the approach proposed by Bartolucci, Scaccia and Mira [3], we may compare alternative change-point models by calculating the corresponding Bayes factors with the output of the reversible-jump algorithm. Instead of the Gaussian priors, we may also employ Laplace priors for the regression coefficients in our proposed approach, leading to a fully Bayesian Lasso framework [5, 18, 33].

ACKNOWLEDGEMENT

We thank the associate editor, the two referees, and the Editor-in-Chief Professor Yuedong Wang for their many constructive and insightful comments that have led to significant improvements in the article. The research was supported in part by a grant (grant number 17307218) from the Research Grants Council of Hong Kong.

Received 14 May 2019

REFERENCES

[1] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679. MR1224394

[2] Allison, I., Bindoff, N. L., Bindschadler, R. A., Cox, P. M., de Noblet, N., England, M. H., Francis, J. E., Gruber, N.,

Haywood, A. M., Karoly, D. J., and Kaser, G. (2011). *The Copenhagen Diagnosis: Updating the World on the Latest Climate Science*. Elsevier.

[3] Bartolucci, F., Scaccia, L., and Mira, A. (2006). Efficient Bayes factor estimation from the reversible jump output. *Biometrika* **93**, 41–52. MR2277738

[4] Carlin, B. P., Gelfand, A. E., and Smith, A. F. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* **41**, 389–405.

[5] Chen, X., Wang, Z. J., and McKeown, M. J. (2011). A Bayesian Lasso via reversible-jump MCMC. *Signal Processing* **91**, 1920–1932. MR2858445

[6] Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics* **86**, 221–241. MR1649222

[7] Derksen, C. and Brown, R. (2012). Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. *Geophysical Research Letters* **39**, L19504.

[8] Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing* **16**, 203–213. MR2227396

[9] Fearnhead, P., and Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 589–605. MR2370070

[10] Frank, K. T., Petrie, B., Choi, J. S., and Leggett, W. C. (2005). Trophic cascades in a formerly cod-dominated ecosystem. *Science* **308**, 1621–1623.

[11] Fu, Y. X. and Curnow, R. N. (1990). Maximum likelihood estimation of multiple change-points. *Biometrika* **77**, 563–573. MR1087847

[12] George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

[13] Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Vol. 196). Minneapolis, MN: Federal Reserve Bank of Minneapolis, Research Department. MR1380276

[14] Girón, J., Ginebra, J., and Riba, A. (2005). Bayesian analysis of a multinomial sequence and homogeneity of literary style. *The American Statistician* **59**, 19–30. MR2109428

[15] Green, P. J. (2003). Trans-dimensional Markov chain Monte Carlo. *Oxford Statistical Science Series*, 179–198. MR2082410

[16] Greene, C. H. and Pershing, A. J. (2007). Climate drives sea change. *Science* **315**, 1084–1085.

[17] Greene, C. H., Pershing, A. J., Cronin, T. M., and Ceci, N. (2008). Arctic climate change and its impacts on the ecology of the North Atlantic. *Ecology* **89**, S24–S38.

[18] Hans, C. (2009). Bayesian Lasso regression. *Biometrika* **96**, 835–845. MR2564494

[19] Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* **105**, 1480–1493. MR2796565

[20] Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change-point in a sequence of binomial variables. *Biometrika* **57**, 477–488. MR0275556

[21] Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* **1**, 145–168. MR2227368

[22] Jansen, J. (1991). Fitting regression models to ordinal data. *Biometrical Journal* **33**, 807–815.

[23] Johnson, V. E. and Albert, J. H. (2006). *Ordinal Data Modeling*. Springer Science and Business Media. MR1683018

[24] Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J. P., Srivastava, A. K., and Sugi, M. (2010). Tropical cyclones and climate change. *Nature Geoscience* **3**, 157–163.

[25] Kossin, J. P., Olander, T. L., and Knapp, K. R. (2013). Trend analysis with a new global record of tropical cyclone intensity. *Journal of Climate* **26**, 9960–9976.

[26] Kwok, R. and Rothrock, D. A. (2009). Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008. *Geophysical Research Letters* **36**, L15501.

[27] Lamnisos, D., Griffin, J. E., and Steel, M. F. (2009). Transdimensional sampling algorithms for Bayesian variable selection in classification problems with many more variables than observations. *Journal of Computational and Graphical Statistics* **18**, 592–612. MR2751643

[28] Lee, K. E., Sha, N., Dougherty, E. R., Vannucci, M. and Mallick, B. K. (2003). Gene selection: A Bayesian variable selection approach. *Bioinformatics* **19**, 90–97.

[29] Levitus, S., Antonov, J. I., Boyer, T. P., Locarnini, R. A., Garcia, H. E., and Mishonov, A. V. (2009). Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophysical Research Letters* **36**, L07608.

[30] Mann, M. E., Woodruff, J. D., Donnelly, J. P., and Zhang, Z. (2009). Atlantic hurricanes and climate over the past 1,500 years. *Nature* **460**, 880–883.

[31] Mendelsohn, R., Emanuel, K., Chonabayashi, S., and Bakkensen, L. (2012). The impact of climate change on global tropical cyclone damage. *Nature Climate Change* **2**, 205–209.

[32] Park, J. H. (2011). Changepoint analysis of binary and ordinal probit models: An application to bank rate policy under the interwar gold standard. *Political Analysis* **19**, 188–204.

[33] Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103**, 681–686. MR2524001

[34] Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika* **67**, 79–84. MR0570508

[35] Qian, S. S., Pan, Y., and King, R. S. (2004). Soil total phosphorus threshold in the Everglades: A Bayesian changepoint analysis for multinomial response data. *Ecological Indicators* **4**, 29–37.

[36] Robbins, M. W., Lund, R. B., Gallagher, C. M., and Lu, Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *Journal of the American Statistical Association* **106**, 89–99. MR2816704

[37] Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C., Wallace, D. W., Tilbrook, B., and Millero, F. J. (2004). The oceanic sink for anthropogenic CO_2. *Science* **305**, 367–371.

[38] Sha, N., Vannucci, M., Tadesse, M. G., Brown, P. J., Dragoni, I., Davies, N., Roberts, T. C., Contestabile, A., Salmon, M., Buckley, C., and Falciani, F. (2004). Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. *Biometrics* **60**, 812–819. MR2089459

[39] Wolfe, D. A. and Chen, Y. S. (1990). The changepoint problem in a multinomial sequence. *Communications in Statistics-Simulation and Computation* **19**, 603–618. MR1075465

Chi Kin Lam
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong
China
E-mail address: cklamsta@connect.hku.hk

Huaqing Jin
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong
China
E-mail address: jinhuaqing@hku.hk

Fei Jiang
Department of Epidemiology and Biostatistics
University of California
San Francisco, CA
USA
E-mail address: Fei.Jiang@ucsf.edu

Guosheng Yin
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong
China
E-mail address: gyin@hku.hk