



# Asynchronous Changepoint Estimation for Spatially Correlated Functional Time Series

Mengchen WANG<sup>✉</sup>, Trevor HARRIS, and Bo LI

We propose a new solution under the Bayesian framework to simultaneously estimate mean-based asynchronous changepoints in spatially correlated functional time series. Unlike previous methods that assume a shared changepoint at all spatial locations or ignore spatial correlation, our method treats changepoints as a spatial process. This allows our model to respect spatial heterogeneity and exploit spatial correlations to improve estimation. Our method is derived from the ubiquitous cumulative sum (CUSUM) statistic that dominates changepoint detection in functional time series. However, instead of directly searching for the maximum of the CUSUM-based processes, we build spatially correlated two-piece linear models with appropriate variance structure to locate all changepoints at once. The proposed linear model approach increases the robustness of our method to variability in the CUSUM process, which, combined with our spatial correlation model, improves changepoint estimation near the edges. We demonstrate through extensive simulation studies that our method outperforms existing functional changepoint estimators in terms of both estimation accuracy and uncertainty quantification, under either weak or strong spatial correlation, and weak or strong change signals. Finally, we demonstrate our method using a temperature data set and a coronavirus disease 2019 (COVID-19) study.

Supplementary materials accompanying this paper appear online.

**Key Words:** Bayesian hierarchical model; Changepoint; Functional data; Piecewise linear model; Spatial correlation.

## 1. INTRODUCTION

In recent years, there has been a considerable renewed interest in changepoint detection and estimation in many fields, including Climate Science (Lund et al. 2007; Reeves et al. 2007), Finance and Business (Lavielle and Teysiere 2007; Taylor and Letham 2018), and traffic analysis (Kurt et al. 2018). The changepoint problem was first studied by Page (1954) for independently and normally distributed time series. Since then, change-point literature has grown tremendously. Methods for changepoints in time series have been

---

M. Wang (✉) · B. Li, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA (E-mail: [mw33@illinois.edu](mailto:mw33@illinois.edu)).

T. Harris, Department of Statistics, Texas A&M University, College Station TX 77843, USA.

developed for both at most one change and multiple changepoints. Vast methodologies are derived based on the cumulative sum (CUSUM) statistic (e.g., [Wald 1947](#); [Shao and Zhang 2010](#); [Aue and Horváth 2013](#); [Fryzlewicz and Rao 2014](#)) which was first introduced by [Page \(1954\)](#) to detect a shift in the process mean, though other methods have also been proposed (e.g., [Chernoff and Zacks 1964](#); [MacEachern et al. 2007](#); [Sundararajan and Pourahmadi 2018](#)).

With the proliferation of high-frequency data collection and massive data storage in recent years, functional data has become increasingly common, and functional data analysis is an increasingly valuable toolkit. For instance, daily temperature data in a specific year can be considered functional data and analyzed using functional data methods. Consequently, functional time series become prevalent, and they usually contain more information than a single time series. Following the previous example, daily temperature data over, say 50 years, can be treated as a functional time series which is much more informative than an annual average temperature series with 50 observations. As for univariate time series, changepoint detection and estimation for functional time series have received particular interest owing to the rise of high-dimensional time series.

Within the functional data analysis (FDA) literature, changepoint detection has primarily focused on the scenario of at most one change. [Berkes et al. \(2009\)](#) proposed a CUSUM test to detect and estimate changes in the mean of independent functional sequence data. The comprehensive asymptotic properties for their estimation are further studied in [Aue et al. \(2009\)](#). Berkes et al.'s test was then extended to weakly dependent functional data by [Hörmann and Kokoszka \(2010\)](#) and to epidemic changes, for which the observed changes will return to baseline at a later time, by [Aston and Kirch \(2012\)](#). [Zhang et al. \(2011\)](#) introduced a test for changes in the mean of weakly dependent functional data using self-normalization to alleviate the use of asymptotic control. Later, [Sharipov et al. \(2016\)](#) developed a sequential block bootstrap procedure for these methods. Recently, [Aue et al. \(2018\)](#) proposed a fully functional method for finding a change in the mean without losing information due to dimension reduction, thus eliminating restrictions of functional principal component based estimators. Other methods in multiple changepoint detection for functional time series can be seen in [Chiou et al. \(2019\)](#), [Rice and Zhang \(2019\)](#), [Harris et al. \(2020\)](#) and [Li and Ghosal \(2021\)](#).

Environmental data often naturally takes the form of spatially indexed functional data. Again using our temperature data example, if we observe such functional time series at many weather stations in a region, then we have a spatial functional time series. The study for changepoint estimation with spatially indexed functional time series is relatively scant compared to the abundant literature for data not associated with spatial locations. The possible spatial variability and correlation for spatially indexed data present challenges for such data analysis. Spatial data observed in a vast region such as weather data in a state may exhibit heterogeneous characteristics across the spatial domain. Furthermore, it is often not straightforward to model and estimate spatial correlation in statistical analysis. However, spatial correlation, if appropriately taken into account, can effectively improve the statistical inference drawn from the spatial data ([Shand et al. 2018](#)). [Gromenko et al. \(2017\)](#) tackled the changepoint estimation for spatial functional data by assuming a common break time for all functional time series over the spatial domain. They developed a test statistic as a weighted

average of the projected CUSUM with the weights defined as the inverse of the covariance matrix of the spatial data. However, the assumption of a common changepoint over the entire spatial domain can be unrealistic. Other related work on spatial functional data includes a test for the correlation between two different functional data sets observed over the same region (Gromenko et al. 2012), a test for the equality of the mean function in two samples of spatial functional data (Gromenko and Kokoszka 2012), and a nonparametric method to estimate the trend as well as evaluate its significance for spatial functional data (Gromenko and Kokoszka 2013). All those works, similar as Kriging (Matheron 1963), took spatial correlation into account by downweighting the data that are strongly spatially correlated.

To illustrate the limitation of assuming a common changepoint for a large region, we examine the changepoints of the daily minimum temperature in California from 1971 to 2020 obtained from <https://www.ncdc.NOAA.gov/cdo-web/search?datasetid=GHCND>. The data are collected over 207 stations, but only 28 stations have sufficiently complete (<15% missing values) time series for meaningful change point estimation and are presented here. We first use 21 Fourier basis functions to smooth the daily data and then apply the fully functional (FF) method of Aue et al. (2018) to each station. We then test for the existence of changepoints with the FF method and find 16 stations with  $p < 0.1$  after a false discovery rate (FDR) control (Benjamini and Hochberg 1995). The locations of stations and the FF changepoint estimates are shown in Fig. 4a in Sect. 4. The changepoint estimates appear asynchronous, though somewhat spatially clustered. Thus, simply assuming a single common break time would misrepresent the changepoint process and lose information.

We propose an asynchronous changepoint estimation method for simultaneously locating at most one change in each mean function of spatially indexed functional time series. In contrast to estimating a single shared changepoint in Gromenko et al. (2017), our method allows both the break time and the amount of change to vary spatially. In addition, we take spatial correlation into account to strengthen the changepoint estimation and respect the inherent spatial continuity. We derive our method based on the asymptotic properties of the functional CUSUM squared norm process at each location. Rather than directly finding the maximum value of the CUSUM-based processes in previous work, we propose to fit spatially correlated piecewise linear models with two pieces for the CUSUM squared norm process across the spatial domain, and estimate changepoints by where the two pieces meet at each individual location. To our knowledge, this is the first methodological exploration of fitting parametric models for the CUSUM process in order to estimate the spatially correlated changepoints. This new strategy leads to more robust changepoint estimation to noise, and improves the estimation for challenging cases where the changepoints are near the edges. All parameters are jointly specified in a Bayesian hierarchical model, which provides a powerful means for parameter estimation as well as allows us to conveniently quantify the uncertainty of the estimation.

The rest of this paper is organized as follows. In Sect. 2, we first introduce the notations and the properties of the CUSUM squared norm process, then present our proposed method. In Sect. 3, we conduct simulations under different scenarios to evaluate the performance of our proposed model and other competitive methods. Real data analysis on the California minimum temperatures and the COVID-19 dataset is presented in Sect. 4. The paper concludes with a brief discussion in Sect. 5.

## 2. CHANGEPOINT ESTIMATION

### 2.1. NOTATION

Let  $X_{\mathbf{s},t}(u)$  be the functional observation at location  $\mathbf{s} \in \mathcal{D}$ , where  $\mathcal{D}$  is a compact subset in  $\mathbb{R}^d$ , and time  $t \in \mathbb{Z}$ . Each  $X_{\mathbf{s},t}(u) \in L^2([0, 1])$  is a real-valued square integrable function defined without loss of generality on the unit interval  $[0, 1]$ , i.e.,  $u \in [0, 1]$ , and  $\int_0^1 |X_{\mathbf{s},t}^2(u) du| < \infty$ . We assume the functional times series  $X_{\mathbf{s},t}(u)$ ,  $t = 1, \dots, T$ , is generated from the following model:

$$X_{\mathbf{s},t}(u) = \mu_{\mathbf{s}}(u) + \delta_{\mathbf{s}}(u)\mathbb{1}(t > k_{\mathbf{s}}^*) + \varepsilon_{\mathbf{s},t}(u), \quad t \in \mathbb{Z}, \quad (1)$$

where  $\mu_{\mathbf{s}}(u)$  is the baseline mean function that is distorted by the addition of  $\delta_{\mathbf{s}}(u)$  after the break time  $k_{\mathbf{s}}^* \in \{1, \dots, T\}$  at location  $\mathbf{s}$ , and  $\mathbb{1}(A)$  is an indicator function that equals 1 only when event  $A$  is true and zero otherwise. We assume that the functional data at all locations are observed at the same time points. We also assume that the zero-mean error functions  $\varepsilon_{\mathbf{s},t}(u)$  are weakly dependent in time at each location, follow the same distribution at all space and time locations, and are second-order stationary with isotropic correlation in the spatial, temporal, and functional domains. The formal assumption statements and details are deferred to Appendix A. To simplify notation, we suppress  $u$  from the functional random variables such as referring to  $\delta_{\mathbf{s}}(u)$  by  $\delta_{\mathbf{s}}$  when there is no risk of confusion.

Changepoint detection in the mean, at each location  $\mathbf{s}$ , can be formulated into the following hypothesis test:

$$H_0 : \delta_{\mathbf{s}} = 0 \text{ versus } H_A : \delta_{\mathbf{s}} \neq 0, \quad (2)$$

where  $\delta_{\mathbf{s}} = 0$  means  $\delta_{\mathbf{s}}(u) = 0$  for all  $u \in [0, 1]$  and otherwise  $\delta_{\mathbf{s}} \neq 0$ . [Aue et al. \(2018\)](#) proposed a fully functional approach to testing the hypothesis (2) for each location  $\mathbf{s}$  based on the functional CUSUM defined as

$$S_{\mathbf{s},T,k}(u) = \frac{1}{\sqrt{T}} \left\{ \sum_{t=1}^k X_{\mathbf{s},t}(u) - \frac{k}{T} \sum_{t=1}^T X_{\mathbf{s},t}(u) \right\}, \quad k = 0, \dots, T, \quad (3)$$

for which the two empty sums  $S_{\mathbf{s},T,0}(u) = S_{\mathbf{s},T,T}(u) = 0$ . Noting that the  $L^2$  norm of the CUSUM statistic,  $\|S_{\mathbf{s},T,k}\|$ , as a function of  $k$  tends to be large at the true break date motivates a max-type test statistic for detecting a change in the mean function:

$$TS_T(\mathbf{s}) = \max_{1 \leq k \leq T} \|S_{\mathbf{s},T,k}(u)\|^2. \quad (4)$$

If a changepoint is detected, [Aue et al. \(2018\)](#) further provided an estimator for the break time  $k_{\mathbf{s}}^*$ :

$$\hat{k}_{\mathbf{s}}^* = \min \left\{ k : \|S_{\mathbf{s},T,k}(u)\| = \max_{1 \leq k' \leq T} \|S_{\mathbf{s},T,k'}(u)\| \right\}.$$

The CUSUM test based on Eq. (4) allows the functional time series to be  $m$ -dependent and requires notably weaker assumptions than the functional principal component-based methods. The CUSUM statistic is shown to be powerful (Page 1954; MacEachern et al. 2007) in detecting mean shift of univariate time series. For functional time series, the CUSUM is also the basis of many other changepoint detection methods (Berkes et al. 2009; Hörmann and Kokoszka 2010; Aston and Kirch 2012; Sharipov et al. 2016; Gromenko et al. 2017).

**2.2. PROPERTIES OF SPATIAL CUSUM PROCESS**

Most previous methods consider changepoint detection in a single functional time series, and thus may have limited power when directly applied for the spatially indexed functional data that exhibit spatial correlation. While Gromenko et al. (2017) took spatial correlation into account, their assumption of a single shared changepoint can be too restrictive for data observed in a large spatial domain. We aim to develop a flexible and efficient method to estimate spatially varying break time  $k_s^*$  jointly for all locations while taking advantage of spatial correlation in the changepoint estimation. Due to the power of CUSUM statistic in changepoint detection, our method will employ the CUSUM as the building block.

Since our method is derived from the asymptotic properties of CUSUM processes for spatially indexed functional time series, we first study those properties before introducing our model in Sect. 2.3. To simplify notation, let

$$Y_{T,k}(\mathbf{s}) = \|S_{\mathbf{s},T,k}(u)\|^2, \quad k = 0, \dots, T. \tag{5}$$

The notation  $Y_{T,k}(\mathbf{s})$  emphasizes that  $Y$  is a spatially varying random process. By definition,  $Y_{T,k}(\mathbf{s}) = 0$  when  $k = 0$  and  $k = T$ . Since  $Y_{T,k}(\mathbf{s})$  largely preserves the changepoint information (Aue et al. 2018), our method will be based on  $Y_{T,k}(\mathbf{s})$ , which reduces the functional sequence  $X_{\mathbf{s},t}(u)$  at each location into a time series  $Y_{T,k}(\mathbf{s}), k = 0, \dots, T$ . The spatial functional sequence thus reduces into a spatiotemporal random process.

We then study the characteristics of the spatiotemporal process  $Y_{T,k}(\mathbf{s})$ . Let  $\lambda_l$  and  $\psi_l(u)$  be the eigenvalues and eigenfunctions of the error process  $\epsilon_{\mathbf{s},t}(u)$  in Eq. (1). The formal definition is deferred to Section S2 of the Supplement. Let  $q = k/T$  be the scaled time point.

**Lemma 1.** *Under the null hypothesis of no changepoint at location  $\mathbf{s}$ , we have*

$$Y_{T,k}(\mathbf{s}) \xrightarrow{D} \sum_{l=1}^{\infty} \lambda_l B_l^2(q) \text{ as } T \rightarrow \infty,$$

where  $(B_l : l \in \mathbb{N})$  are independent and identically distributed (iid) standard Brownian bridges defined on  $[0, 1]$ ,  $E\{\sum_{l=1}^{\infty} \lambda_l B_l^2(q)\} = q(1 - q) \sum_{l=1}^{\infty} \lambda_l$  and  $\text{var}\{\sum_{l=1}^{\infty} \lambda_l B_l^2(q)\} = 2q^2(1 - q)^2 \sum_{l=1}^{\infty} \lambda_l^2$ .

**Proposition 1.** *Under the alternative hypothesis that there is one changepoint  $k_s^*$  at location  $\mathbf{s}$  and the corresponding change function is  $\delta_s(u)$ , we have*

$$\sqrt{\{Y_{T,k}(\mathbf{s})\}} - \sqrt{\{Z_{T,k}(\mathbf{s})\}} \xrightarrow{\mathcal{P}} 0,$$

for a random process  $Z_{T,k}(\mathbf{s})$  with

$$E\{Z_{T,k}(\mathbf{s})\} = \begin{cases} q(1-q) \sum_{l=1}^{\infty} \lambda_l + Tq^2 \|\delta_s(u)\|^2 \left(1 - \frac{k_s^*}{T}\right)^2, & \text{if } k \leq k_s^*; \\ q(1-q) \sum_{l=1}^{\infty} \lambda_l + T(1-q)^2 \|\delta_s(u)\|^2 \left(\frac{k_s^*}{T}\right)^2, & \text{if } k > k_s^*, \end{cases} \quad (6)$$

and

$$\text{var}\{Z_{T,k}(\mathbf{s})\} = \begin{cases} aq^2(1-q)^2 + b_s Tq^3(1-q) \left(1 - \frac{k_s^*}{T}\right)^2, & \text{if } k \leq k_s^*; \\ aq^2(1-q)^2 + b_s Tq(1-q)^3 \left(\frac{k_s^*}{T}\right)^2, & \text{if } k > k_s^*, \end{cases} \quad (7)$$

where

$$a = 2 \sum_{l=1}^{\infty} \lambda_l^2 \text{ and } b_s = 4 \sum_{l=1}^{\infty} \lambda_l \left\{ \int_0^1 \psi_l(u) \delta_s(u) du \right\}^2.$$

Assumption 2 in Appendix A that assumes identically distributed error functions implies both  $\lambda_l$  and  $\psi_l$  are invariant across  $\mathbf{s}$  and  $t$ , so all locations share the same parameter  $a$  which represents the feature of the long-run variance, whereas  $b_s$  depends on change functions that may vary across different locations. Proofs of Lemma 1 and Proposition 1 are deferred to Section S4 of the Supplement.

The asymptotics in Proposition 1 indicates that we can use the mean and variance of  $\sqrt{\{Z_{T,k}(\mathbf{s})\}}$  to approximate those of  $\sqrt{\{Y_{T,k}(\mathbf{s})\}}$  at a large  $T$ . However, the calculation of the first two moments for  $\sqrt{\{Z_{T,k}(\mathbf{s})\}}$  is rather involved compared to that for  $Z_{T,k}(\mathbf{s})$  due to the square root operator. To bypass that difficulty, we propose to use the mean and variance of  $Z_{T,k}(\mathbf{s})$  to approximate those of the  $Y_{T,k}(\mathbf{s})$  process. This is not a viable solution in theory; however, the approximations are sufficient to serve our purpose of understanding the mean and variance structure of  $Y_{T,k}(\mathbf{s})$  for deriving our method. To evaluate how well (6) and (7) approximate the mean and variance of the  $Y_{T,k}(\mathbf{s})$ , respectively, we conduct simulations at four different settings composed of two different  $T$ 's and two signal-to-noise ratio (SNR) values that will be introduced in Sect. 3.1. The details of the simulation can be found in Section S5.1 of the Supplement. Figure 1 compares the empirical mean and variance from the simulations with their theoretical approximations. For all scenarios we considered, the approximations seem to match with the empirical result well, especially in the mean function.

The expression in Eq. (6) shows that when  $T$  is large the mean of the  $Y_{T,k}(\mathbf{s})$  sequence attains its peak at the changepoint. This is indeed the basis of the test in Aue et al. (2018). Figure 1 also shows that the  $Y_{T,k}(\mathbf{s})$  sequence starts from exactly zero on both ends and then peaks at the true changepoint  $0.6T$ . In addition, Eq. (6) indicates that the stronger the change signal is, the more pointed the peak tends to be, which is corroborated by the comparison between the means of  $Y_{T,k}(\mathbf{s})$  at two different SNR values. The variance of  $Y_{T,k}(\mathbf{s})$  also

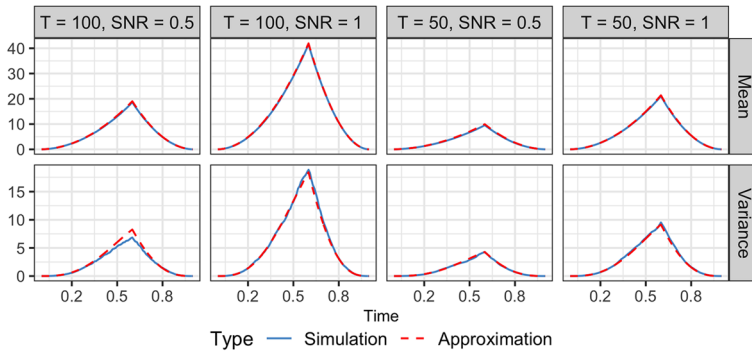


Figure 1. Mean and variance of the  $Y_{T,k}(s)$  process based on simulation results (blue solid) and the proposed theoretical approximation (red dashed).

starts from zero at the two ends and then increases toward the center. However, there is no theoretical evidence that the variance should maximize at the changepoint. Indeed, we find the peak of the variance is not necessarily located at the changepoint, though this particular simulation shows so.

The properties of  $Y_{T,k}(s)$  enlighten us to estimate the break time by fitting a piecewise linear model with two pieces for the  $Y_{T,k}(s)$ ,  $0 \leq k \leq T$  sequence at each location. The two pieces are expected to be joined at the break time. Figure S1 in Section S5.2 of the Supplement illustrates this idea using simulated  $Y_{T,k}(s)$  processes. Due to the constraint of being zeroes on both ends, the two pieces can be modeled by one slope parameter, and a stronger change signal will lead to a steeper slope. Although the mean function in Eq. (6) suggests a piecewise quadratic model, for simplicity and the robustness of linear models, we choose the piecewise linear model which suffices for our purpose of capturing the peak of the  $Y_{T,k}(s)$  process. In order to correctly quantify the uncertainty of the fitted piecewise linear model and thus the uncertainty of the changepoint estimation, it is important to feed the regression model with the appropriate variance structure. We model the variance of the piecewise linear model following Eq. (7).

If the functional data are observed at nearby locations, their break times are expected to be similar due to spatial dependency, so is the amount of change. What these similarities pass to the piecewise linear models is that the locations of the joints and the slopes of the models at two neighboring locations tend to be respectively similar. This suggests us to borrow information from neighbors when estimating the changepoint at one specific location.

Given the above considerations, we propose a Bayesian hierarchical model to jointly estimate spatially varying changepoints together with their uncertainty for all locations that have changepoints. In practice, we can first apply any changepoint detection method at each location and then employ FDR to adjust the p-values to decide which locations show significant evidence of having a changepoint. If the number of spatial locations is large, the mirror procedure developed by Yun et al. (2020) can be an effective alternative to the classic FDR control.

### 2.3. BAYESIAN HIERARCHICAL MODEL

Assume changepoints are detected at locations  $\mathbf{s}_1, \dots, \mathbf{s}_N$ . At each of those locations, we fit a two-piece piecewise linear model with only one slope parameter for  $Y_{T,k}(\mathbf{s})$ ,  $k = 1, \dots, T-1$ , due to the constraint of  $Y_{T,k}(\mathbf{s}) = 0$  for  $k = 0$  and  $k = T$ . We model the slope parameters and the joints of the two pieces as spatially correlated processes to account for the spatial correlation in the break time and change amount of the changepoints. Let  $c(\mathbf{s}) = k_s^*/T \in (0, 1)$  be the scaled location specific changepoint. We propose the following model:

**Stage I** Likelihood of the  $Y_{T,k}(\mathbf{s})$  process:

$$Y_{T,k}(\mathbf{s}) = \beta(\mathbf{s})[\{c(\mathbf{s}) - 1\}q + \{q - c(\mathbf{s})\}\mathbb{1}\{q \geq c(\mathbf{s})\}] + e_k(\mathbf{s}), \quad k = 1, \dots, T-1,$$

where  $\beta(\mathbf{s}) < 0$  is the spatially varying piecewise linear model coefficient, and the error process  $e_k(\mathbf{s})$  is assumed to be a zero-mean spatially correlated Gaussian process. We further assume a space-time separable covariance structure for errors for simplicity, as is widely used in spatiotemporal modeling (Haas 1995; Hoff 2011). We denote the entire error process as

$$\mathbf{e} = (e_1(\mathbf{s}_1), \dots, e_1(\mathbf{s}_N), e_2(\mathbf{s}_1), \dots, e_2(\mathbf{s}_N), \dots, e_{T-1}(\mathbf{s}_1), \dots, e_{T-1}(\mathbf{s}_N))^T,$$

and assume

$$\mathbf{e} \sim N(\mathbf{0}_{N(T-1)}, \mathbf{\Omega}^{1/2} \mathbf{\Gamma}_t \otimes \mathbf{\Gamma}_s \mathbf{\Omega}^{1/2}),$$

where  $\mathbf{0}_{N(T-1)}$  is a vector of  $N(T-1)$  zeros and

$$\mathbf{\Omega} = \text{diag} \left( \omega_1^2(\mathbf{s}_1), \dots, \omega_1^2(\mathbf{s}_N), \dots, \omega_{T-1}^2(\mathbf{s}_1), \dots, \omega_{T-1}^2(\mathbf{s}_N) \right),$$

with

$$\omega_k^2(\mathbf{s}) = \begin{cases} aq^2(1-q)^2 + b_s\{1-c(\mathbf{s})\}^2Tq^3(1-q), & \text{if } q \leq c(\mathbf{s}); \\ aq^2(1-q)^2 + b_sc(\mathbf{s})^2Tq(1-q)^3, & \text{if } q > c(\mathbf{s}). \end{cases}$$

The variance term  $\mathbf{\Omega}$  follows the theoretical approximation in Eq. (7) to represent the uncertainty of  $Y_{T,k}(\mathbf{s})$ . Parameters  $a$  and  $b_s$  are complex functions of unknown eigenvalues, eigenfunctions and change functions. We will directly treat them as unknown nuisance parameters in our model. This also gives us the leverage of being less dependent on the exact form of the approximation but rather following its basic structure. The pure temporal correlation matrix  $\mathbf{\Gamma}_t$  and pure spatial correlation matrix  $\mathbf{\Gamma}_s$  can be governed by any valid correlation function such as exponential or Matérn function (Stein 2012). For simplicity, we assume an exponential covariance function for both matrices:

$$\mathbf{\Gamma}_t(k,k') = \exp\left(-\frac{|k-k'|}{T\phi_t}\right), \quad \mathbf{\Gamma}_s(j,j') = \exp\left(-\frac{\|\mathbf{s}_j - \mathbf{s}_{j'}\|}{\phi_s}\right),$$



where  $\phi_t$  and  $\phi_s$  are range parameters for temporal and spatial correlation, respectively.

As shown earlier by the asymptotic and numerical results, the shape of the piecewise linear model is influenced by the change function and changepoint. To respect the fact that the nearby locations tend to have similar changepoints and change functions, we regulate  $\boldsymbol{\beta} = (\beta(\mathbf{s}_1), \dots, \beta(\mathbf{s}_N))^T$  and  $\mathbf{c} = (c(\mathbf{s}_1), \dots, c(\mathbf{s}_N))^T$  by a correlated process. Since  $\mathbf{b} = (b_{s_1}, \dots, b_{s_N})^T$  also depends on the change function, it is governed by a correlated process as well. Because the dependency in  $\boldsymbol{\beta}$ ,  $\mathbf{c}$  and  $\mathbf{b}$  all arise from the spatial dependency in the data, it is not unreasonable to assume these parameters share one correlation matrix  $\boldsymbol{\Sigma}(\phi)$  to retain parsimony of the model. Considering the constraints that the slope  $\beta(\mathbf{s})$  is negative, changepoint  $c(\mathbf{s})$  is between 0 and 1, and the parameters  $a$  and  $b_s$  in the variance part are positive, we construct the following priors:

**Stage II Priors:**

$$\begin{aligned} \log(-\boldsymbol{\beta}) &\sim N(\boldsymbol{\mu}_\beta, \sigma_\beta^2 \boldsymbol{\Sigma}(\phi)), \\ \Phi^{-1}(\mathbf{c}) &\sim N(\boldsymbol{\mu}_c, \sigma_c^2 \boldsymbol{\Sigma}(\phi)), \\ \log(a) &\sim N(\mu_a, \sigma_a^2), \\ \log(\mathbf{b}) &\sim N(\boldsymbol{\mu}_b, \sigma_b^2 \boldsymbol{\Sigma}(\phi)), \end{aligned}$$

where  $\boldsymbol{\Sigma}(\phi)_{nn'} = \exp(-\|\mathbf{s}_n - \mathbf{s}_{n'}\|/\phi)$ .

All parameters  $\boldsymbol{\mu}_\beta$ ,  $\boldsymbol{\mu}_c$ ,  $\mu_a$  and  $\boldsymbol{\mu}_b$  take values in  $\mathbb{R}$ , so we choose a normal distribution with large variance as their weak hyperpriors. The variance parameters  $\sigma_i^2, i = \beta, c, a, b$ , are all given a conjugate inverse gamma hyperprior. We choose  $\text{IG}(0.1, 0.1)$  because it provides relatively reasonable range for the variances of  $\boldsymbol{\beta}$ ,  $\mathbf{c}$ ,  $a$ , and  $\mathbf{b}$ , and we find the hyperprior has little influence on all those parameters, in particular, the  $\mathbf{c}$  parameter of our primary interest.

The range parameters  $\phi$ ,  $\phi_s$  and  $\phi_t$  are positive, so we choose an exponential hyperprior for them but set a different hyperparameter for  $\phi_t$ , given that the spatial and temporal domains have different characteristics.

The hyperpriors for unknown parameters are given in Section S6 of the Supplement. We use the Markov chain Monte Carlo (MCMC) algorithm to obtain posterior samples from the model. Gibbs sampling is utilized to sample the posteriors for  $\sigma_\beta^2, \sigma_c^2, \sigma_a^2$  and  $\sigma_b^2$ , while the Metropolis-Hasting-within-Gibbs algorithm is implemented for the remaining parameters. For parameters  $\boldsymbol{\beta}$ ,  $\mathbf{c}$ ,  $a$  and  $\mathbf{b}$  that have constraints on their range, we first transform them to the real line before sampling. The derivation of the posterior full conditional distributions can be found in Section S6 of the Supplement.

### 3. SIMULATION STUDY

We conduct simulations to evaluate the accuracy of our changepoint estimation, as well as the coverage and the length of the credible interval. We also explore how the strength of spatial correlation and change signal influence performance. To further study the properties of our method, we compare it with other competitive methods.

### 3.1. DATA GENERATION

We randomly select  $N = 50$  locations in a  $10 \times 10$  spatial domain as the rejection region  $\mathcal{D}_R$  resulting from a changepoint detection algorithm adjusted by the FDR control. Due to the joint estimation for all locations of our method, the false discoveries, i.e., the null locations falsely classified as alternatives, may undermine the estimation. To mimic false discoveries at a typical rate of 0.1, we randomly select a cluster of  $N_0 = 5$  locations among the  $N$  locations to be the falsely classified null locations. At each location, we consider  $T = 50$  time points and generate  $T$  functional data,  $X_{s,t}(u) : u \in [0, 1]$  for  $t = 1, \dots, T$ , as defined in Eq. (1). Without loss of generality, we assume the mean curves,  $\mu_{s_1}, \dots, \mu_{s_N}$ , to be zero functions. Thus, the data generation mainly involves simulating error functions, break time and change functions, except that at those  $N_0$  locations, the change function  $\delta_s$  is set to be zero.

**Error functions:** Although we allow the error functions to be weakly dependent, using temporally independent error functions in simulation studies is very common (Horváth et al. 2013; Aue et al. 2018). In particular, Aue et al. (2018) repeated their simulation with the first-order functional autoregressive errors, and found the results generally remain the same as those from the independent errors. This is because the power of  $Y_{T,k}(\mathbf{s})$  process for changepoint detection is insensitive to the error correlation structure. The  $Y_{T,k}(\mathbf{s})$  process is temporally correlated even if  $X_{s,t}(u)$  is temporally independent. We therefore adopt temporally independent error functions in our simulation. For each location, we generate  $T$  error functions  $\varepsilon_{s,t}$  as follows:

$$\varepsilon_{s,t}(u) = \sum_{l=1}^L \xi_{s,t}^l v_l(u), \quad t = 1, \dots, T, \quad \mathbf{s} \in \mathcal{D}_R,$$

where  $L = 21$  is the number of Fourier basis functions,  $v_l(u)$  is the  $l$ th Fourier basis function, and  $\xi_{s,t}^l$  is the coefficient for  $v_l(u)$  at location  $\mathbf{s}$  and time point  $t$ .

Define  $\boldsymbol{\xi}_t^l = (\xi_{s_1,t}^l, \dots, \xi_{s_N,t}^l)^T$  for any  $l$  between 1 and  $L$ , and assume  $\boldsymbol{\xi}_t^l \sim N(\mathbf{0}_N, 1/(2m^3)\boldsymbol{\Sigma})$ . To ensure curve smoothness, we set  $m = 1$  if  $l = 1$ ,  $m = l/2$  if  $l$  is even, and  $m = (l - 1)/2$  if  $l$  is odd and  $l \geq 3$ . The derivation of  $m$  and details of basis functions are deferred to Section S7 of the Supplement. To ensure the error functions be spatially correlated, we assume that the  $N \times N$  matrix  $\boldsymbol{\Sigma}$  is governed by  $\boldsymbol{\Sigma}(\phi)_{ij} = \exp(-\|\mathbf{s}_i - \mathbf{s}_j\|/\phi)$  for a range parameter  $\phi$ .

**Break time:** For the region  $\mathcal{D}_a$  of the  $N_a = 45$  true alternative locations,  $\mathbf{s}_1, \dots, \mathbf{s}_{N_a}$ , we first generate the scaled break times  $\tilde{k}_s^*$ ,  $\mathbf{s} \in \mathcal{D}_a$ , from a truncated multivariate normal distribution such that  $0.15 \leq \tilde{k}_s^* \leq 0.85$  for any  $\mathbf{s} \in \mathcal{D}_a$ :

$$\tilde{\mathbf{k}}^* = (\tilde{k}_{s_1}^*, \dots, \tilde{k}_{s_{N_a}}^*)^T \sim TN(\mathbf{0.5}_{N_a}, \boldsymbol{\Sigma}_a, \mathbf{0.15}_{N_a}, \mathbf{0.85}_{N_a}),$$

where  $\mathbf{d}_n$  is an  $n$ -length vector of all  $d$  values and  $\mathbf{x} \sim TN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_a, \mathbf{b}_l, \mathbf{b}_u)$  means

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}_a, \mathbf{b}_l, \mathbf{b}_u) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_a^{-1}(\mathbf{x} - \boldsymbol{\mu})\}}{\int_{\mathbf{b}_l}^{\mathbf{b}_u} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_a^{-1}(\mathbf{x} - \boldsymbol{\mu})\} d\mathbf{x}}.$$

Again,  $\Sigma_a(\phi)_{ij} = \exp(-\|s_i - s_j\|/\phi)$  for  $s_i, s_j \in \mathcal{D}_a$ . Then, the real break time  $k_s^* = [\tilde{k}_s^* T]$ ,  $s \in \mathcal{D}_a$ , where  $[n]$  denotes rounding  $n$  to its nearest integer. We truncate the scaled break time to ensure there are a reasonable amount of data both before and after the changepoint.

**Change functions:** We generate change functions  $\delta_s, s \in \mathcal{D}_a$ , as follows:

$$\delta_s = \sum_{l=1}^L \eta_s^l v_l, \quad s \in \mathcal{D}_a,$$

where  $v_l$  is the  $l$ th Fourier basis function and  $\eta_s^l$  is the coefficient for  $v_l$  at location  $s$ . Define  $\eta_l = (\eta_{s_1}^l, \dots, \eta_{s_{N_a}}^l)^T$  for any  $l$  between 1 and  $L$ , and assume  $\eta_l \sim N(\rho/m^2 \mathbf{1}_{N_a}, 1/(10m^3) \Sigma_a)$ , where  $m$  and  $\Sigma_a$  follow the definition in the error function and break time, respectively. The parameter  $\rho$  measures the magnitude of the change signal.

To investigate how our model performs under different spatial correlation strengths, we consider both  $\phi = 2$  and 5 which corresponds to relatively weaker and stronger spatial correlation. The correlation between functional data at two locations decays faster under  $\phi = 2$  than that with  $\phi = 5$ . It is also interesting to study the influence of change signal strength on our model performance. We adopt the signal-to-noise ratio (SNR) used in [Aue et al. \(2018\)](#) to measure the strength of the change signal. SNR, the ratio of the magnitude of change function to that of error functions, is defined as

$$\text{SNR} = \frac{\theta(1 - \theta) \|\delta\|^2}{\text{tr}(\mathbf{C}_\epsilon)}, \tag{8}$$

where  $\theta$  is the scaled date of the changepoint, i.e.,  $k_s^*/T$  in our context,  $\delta$  is the change function,  $\mathbf{C}_\epsilon$  is the long-run covariance matrix of the error functions as defined in Equation (S2.1) of the Supplement, and  $\text{tr}(\cdot)$  is the trace function. The estimation procedure for SNR at a single location is detailed in [Aue et al. \(2018\)](#). By setting  $\rho = 1$  and 1.5, we obtain simulated data with mean SNR over all locations in  $\mathcal{D}_a$  being around 0.5 and 1, which corresponds to weaker and stronger signal, respectively.

### 3.2. RESULTS

To evaluate the performance of the proposed method, we examine the root-mean-squared error (RMSE) of the changepoint estimate, the empirical coverage of the credible interval (CI), and the length of CI. For each setting of spatial correlation and SNR, we run 100 simulations. Different locations, changepoints and functional data are generated independently in each simulation. To ensure MCMC chain convergence, we try different initial values of all parameters and run MCMC until the parallel chains converge based on the Gelman–Rubin diagnostic ([Gelman and Rubin 1992](#)). We also apply Geweke’s diagnostic ([Geweke 1991](#)) to determine the burn-in period. Through experimentation, we picked a conservatively large iteration number of 20,000 and a larger than necessary burn-in size of 15,000 for all simulation runs to ensure convergence. We retain every 10th iteration and discard the remainder to

reduce the positive correlation in the original sample. We compute 95% credible intervals as the interval between the 2.5 and 97.5 percentiles of the posteriors for each parameter.

We compare our method to the recent fully functional (FF) method in [Aue et al. \(2018\)](#) and the method particularly designed for spatial functional data in [Gromenko et al. \(2017\)](#) (hereinafter GKR). The GKR method mainly focuses on changepoint detection and does not provide confidence intervals. Thus, our comparison with [Gromenko et al. \(2017\)](#) is only limited to comparing the accuracy of the estimation. The changepoint confidence interval based on the FF method is computed using the R package `fChange`. Although all methods are applied to the functional data at 50 locations, the evaluation metrics are calculated only at the 45 true alternative locations.

The RMSE of the changepoint estimates from all three methods is reported in Section S9 of the Supplement. Unsurprisingly, GKR has significantly higher error rates than the other two methods since it assumes a simultaneous changepoint, whereas the data are generated with spatially varying changepoints. We instead focus on FF and our method in [Fig. 2](#) since they have comparable error rates. Across all four scenarios representing both the weaker and stronger spatial correlation and SNR, our proposed method outperforms FF by reducing the RMSE of the changepoint estimation. When the signal of change is stronger ( $\rho = 1.5$ ), both FF and our method show smaller and more stable RMSE, as expected. When spatial correlation is higher ( $\phi = 5$ ), our method achieves far less estimation error, especially in the challenging situation with a weaker change signal ( $\rho = 1$ ). This implies that our method can use spatial correlation to improve the changepoint estimation. Curiously, the FF method experiences a slight RMSE reduction in the high correlation regime, which turned out to be an artifact of the data generation randomness. Details are reported in Section S9 of the Supplement.

We further report the empirical coverage probability of our 95% credible intervals against the 95% confidence intervals of the FF method, and present the interval lengths of both methods in [Fig. 2](#). Narrow credible or confidence intervals, with empirical coverage close to the nominal level, indicate precise uncertainty quantification. Our credible intervals, based on weakly informative priors, have empirical coverage closer to the nominal level and are narrower than the corresponding FF confidence intervals. We observe that when the change signal is stronger, both FF and our method improve the uncertainty quantification compared to the lower change signal scenarios. Again, our method is apparently able to take advantage of the spatial correlation in changepoint estimation, reflected by shorter credible interval length while better coverage when the spatial correlation becomes stronger. This ability is particularly important when the change signal is weak, because in such cases, methods like FF that do not take spatial correlation into account may face challenges.

[Figure 3](#) shows the 95% credible and confidence intervals from randomly chosen simulation runs in two different settings. [Figure 3a](#) is associated with the stronger spatial correlation and the stronger change signal when both our and the FF method have the best performance among all the settings. In this scenario, the confidence interval from the FF method performs slightly worse than the credible interval of our method and the RMSE from FF is competitive. Nevertheless, it is still seen that when the true changepoints are close to the edges, the FF method tends to miss true values and results in longer credible intervals, while our method consistently captures all changepoints well regardless of their positions. Besides,

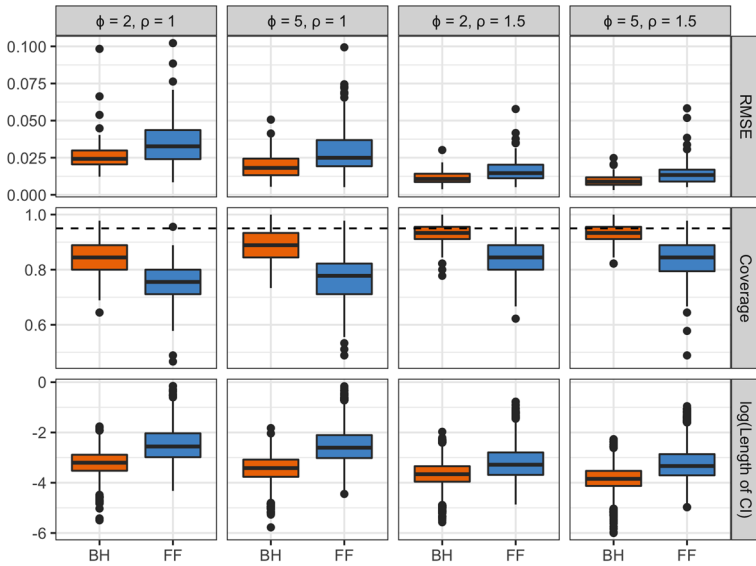


Figure 2. Boxplots of RMSE, the empirical coverage probability of 95% credible and confidence intervals and the logarithms of interval length under four settings. The range parameter  $\phi = 2$  and  $\phi = 5$  represent weaker and stronger spatial correlation, and  $\rho = 1$  and  $\rho = 1.5$  represent weaker and stronger change signals, respectively. “BH” is our proposed Bayesian hierarchical model and “FF” refers to the fully functional method in [Aue et al. \(2018\)](#).

for many locations, even though the estimate from the FF method is close to the true value, their confidence interval often appears too long to be informative. Figure 3b corresponds to the case with the stronger spatial correlation and weaker signal. Both methods perform satisfactorily when the true changepoint is near  $0.5T$ . However, the FF method in this scenario struggles to capture the changepoint as well as quantify the uncertainty when the real changepoint is slightly extreme toward both ends. In contrast, our method still retains its power in those situations by providing accurate estimates and informative credible intervals.

Under the null hypothesis of no changepoint, the variability of  $Y_{T,k}(s)$  process is large in the middle and reaches its peak at  $k = 0.5T$ . Even if the changepoint exists and perturbs the variation, the variance in the middle still tends to be higher due to the intrinsic properties of  $Y_{T,k}(s)$ , though the peak may not occur at the center. Since the FF method only searches for the maximum value of the  $Y_{T,k}(s)$  process, it could be vulnerable to the large variance often dwelling around the center of the duration. When the real changepoint is off-center and the signal is weak, high variance near the center can lead to spurious maxima in the  $Y_{T,k}(s)$  process. In contrast, our method attempts to identify the changepoint with a piecewise linear model, which is more robust to variance. Furthermore, our method allows us to borrow the neighborhood information to estimate the changepoint, which is particularly helpful for challenging situations such as change signal being weak or changepoints close to the edges.

It is worth noting that we use 21 basis functions, the same number as to smooth the daily minimum temperature in California, to smooth the simulated data. We also evaluated the sensitivity of our method to the presmoothing procedure by smoothing the simulated data with  $L = 5, 7, 11, 21, 31$  Fourier basis functions, respectively. We found that the

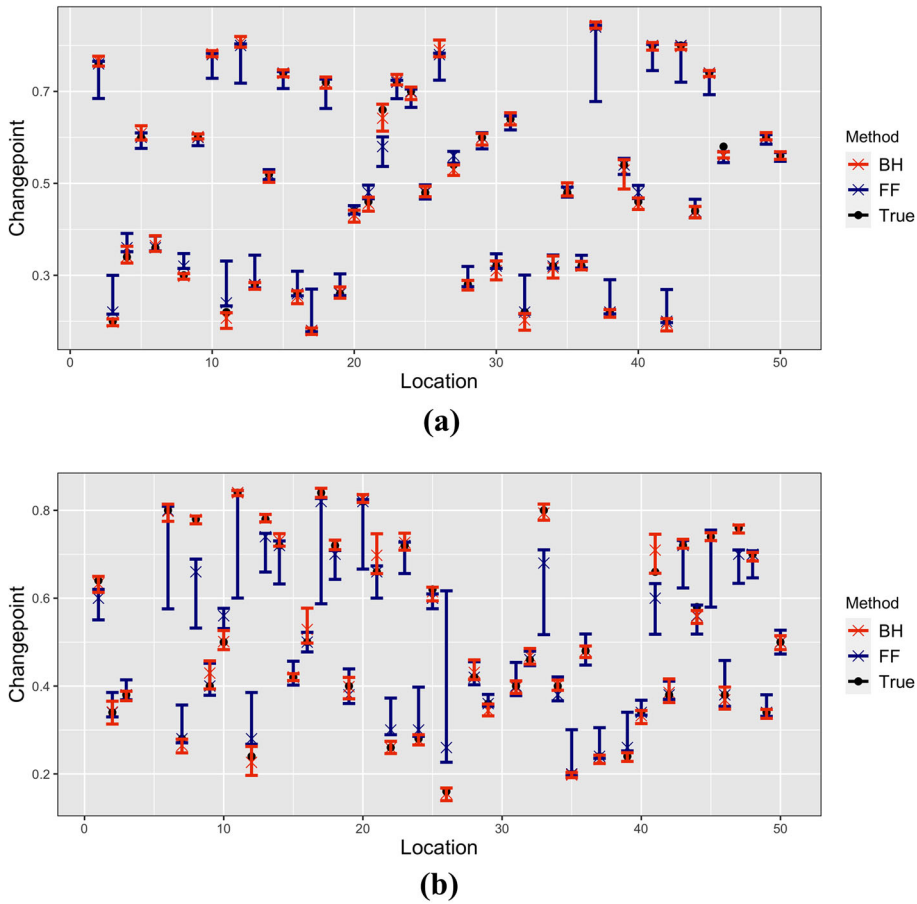


Figure 3. 95% credible and confidence intervals (vertical “I”), changepoint estimates (cross), and true changepoints (black dot) at alternative locations. Labeling of the procedures is the same as that in Fig. 2. **a** A simulation from the setting  $\phi = 5$  and  $\rho = 1.5$ . Our model has coverage 93.3% and RMSE 0.0068, while the FF method has coverage 80% and RMSE 0.0146. **b** A simulation from the setting  $\phi = 5$  and  $\rho = 1$ . Our model has coverage 93.3% and RMSE 0.0130, while the FF method has coverage 66.7% and RMSE 0.0363.

performance of our method is insensitive to a particular choice of the number of basis functions. More details can be seen in Section S9 of the Supplement.

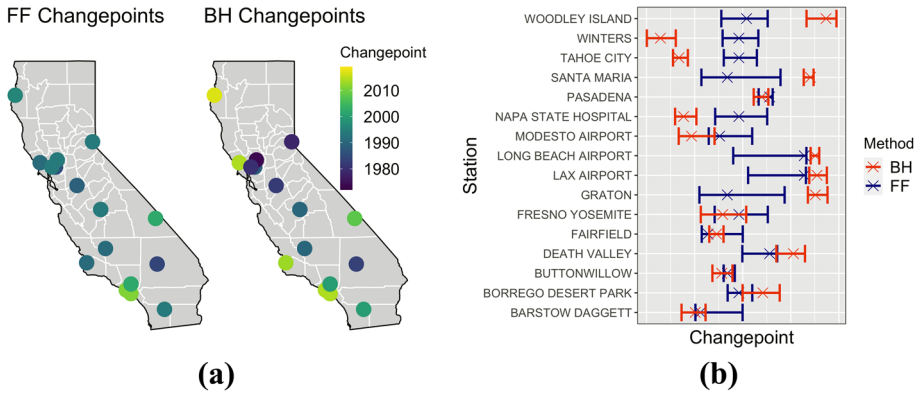


Figure 4. **a** Changepoint estimates from the FF method (left) and our proposed BH method (right). **b** 95% credible intervals from our method and 95% confidence intervals from the FF method.

### 4. REAL DATA EXAMPLES

We demonstrate our method on two datasets and, again, compare our results with the FF detector of Aue et al. (2018). The first dataset is the temperature profiles introduced in Sect. 1, and the second dataset records COVID-19 positive cases by age in Illinois during the spring of 2021.

#### 4.1. CALIFORNIA MINIMUM TEMPERATURE

As described in Sect. 1, we have daily minimum temperature profiles at 207 locations in California from 1971 to 2020. Due to the high degree of missingness in many sites, we only retain 28 stations that have at least 85% complete profiles each year. Each profile is then smoothed with 21 Fourier basis functions. We apply the FF method to further subset the number of stations down to 16, each with a p-value below the 0.1 cutoff after FDR correction.

As an example, the daily minimum temperature profile at Los Angeles International Airport in 1980 together with the smoothed curve using the 21 Fourier basis functions are shown in Figure S6 of the Supplement. When applying our method to this data, we determine the burn-in size using Geweke’s diagnostic and check the MCMC convergence with Gelman–Rubin diagnostic, which guides us to run 30,000 iterations with a burn-in size of 20,000. Again we use a thinning interval of 10. The changepoint posterior estimates are shown in Fig. 4a, and the credible intervals for each station are shown in Fig. 4b, together with the FF estimates and their confidence intervals for a comparison.

Both Figs. 4a and b indicate the FF break date estimates concentrate near the middle of the interval, while our method freely finds changepoints all along the interval. The same phenomenon was observed in the simulation studies. Our estimates also preserve the spatial continuity of the naturally dependent temperature process, as evidenced by the changepoint estimates in Fig. 4a Stations close in space tend to have changepoints close in time. Accu-

rate changepoint estimates and informative credible intervals can help us more profoundly understand the climate dynamics and the threat of tipping points in the climate system.

## 4.2. COVID-19 DATA IN ILLINOIS

As we all know, the coronavirus emerged as mainly attacking the older adults, but then it is observed that the age distribution of COVID-19 cases moved toward younger ones. One interesting question in studying how COVID-19 cases evolve is to identify when the age distribution changes. To investigate this question in our state, we obtain the daily COVID-19 cases for all counties in Illinois between 01/01/2021 and 04/05/2021, 95 days in total, from the Illinois Department of Public Health (<https://www.dph.illinois.gov/covid19/data-portal>). The data reports the number of cumulative confirmed and probable positive cases in 9 age groups (< 20, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80+, Unknown). After exploratory data analysis, we eliminate the age group “Unknown” because this category only contains very few cases and the numbers are often incoherent. Although eight data points are considered low for representing a function, they contain all the information that the public health department concerns.

For each county, we first calculate the daily new cases for each age group and then scale them by the total number of daily new cases to approximate the age distribution. We consider the daily age distribution over time as a functional time series, and our goal is to detect and locate any changepoints. We smooth the data using seven Fourier basis functions, and an example of smoothed data is shown in Section S10 of the Supplement. Again, we first use the FF test and FDR control to identify the counties that show evidence of change; 28 such counties are identified. Champaign County has the adjusted  $p$ -value 0.102, only barely above the threshold 0.1. Since Champaign County is the 10th largest among the 102 counties in Illinois in terms of population, and it has a large young age group due to a major public university being in this county, we also include Champaign for changepoint estimation.

We apply both FF and our method to the data over the counties that are expected to have changepoints. For simplicity, we use an exponential covariance function to model the dependence between county level parameters, and use the county geographical center to calculate distance, though conditional or simultaneous autoregressive models are usually more typical for areal aggregated data. We do not expect the results will be sensitive to the choice of the covariance model due to the scatter of the 29 counties. Using the same convergence diagnostic as for the previous temperature dataset, we run MCMC for 50,000 iterations and take the first 40,000 as the burn-in, then we thin the rest using the stepsize 10 to obtain the posterior samples.

The changepoint estimates from both methods are shown in Fig. 5a, and the 95% credible intervals and confidence intervals are shown in Fig. 5b. Again, our method is able to estimate changepoints close to the boundaries, and our credible intervals are shorter than the FF confidence intervals.

To illustrate how the age distribution changes, we further plot the functional time series and their mean functions colored in two groups, whether before or after the changepoint, using Champaign and Peoria as two examples. In Fig. 6, we can see in both counties, the ratio of younger people getting the coronavirus increases and that of the elder drops.



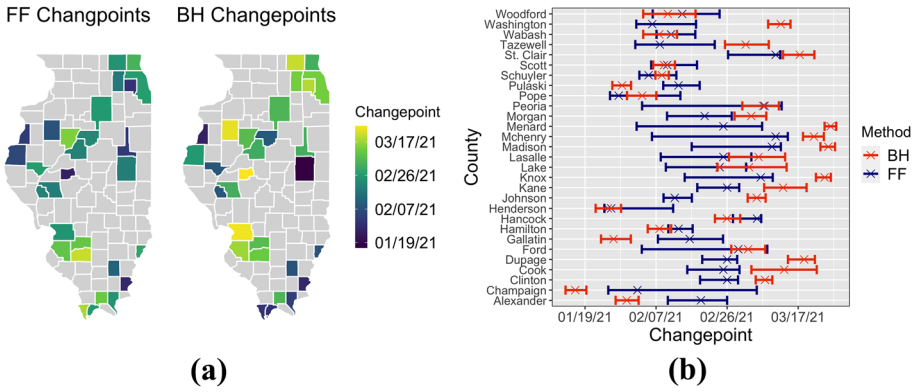


Figure 5. **a** Changepoint estimates from the FF method (left) and our proposed BH method (right) over the 29 counties. **b** Credible intervals from our method and confidence intervals from the FF method.

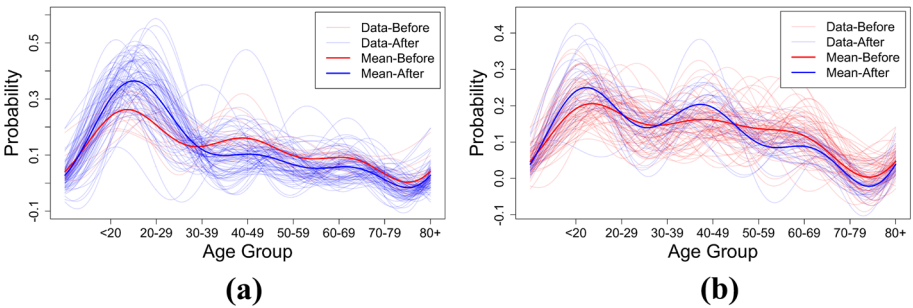


Figure 6. Functional time series of COVID-19 age distribution in **a** Champaign County and **b** Peoria County. The light red and blue curves represent the functional data before and after the changepoint, respectively. The solid red and blue curves are the respective mean of the light red and blue curves.

For Champaign County, where the University of Illinois Urbana-Champaign is located, a changepoint is detected on January 16, 2021, by our method. According to the school calendar, University residence halls were open for the spring semester on January 17. So it was approximately the time when students in < 20 and 20–29 age groups began to gather at the university. This could be one factor for cases shifting to the younger-age groups for this county.

### 5. DISCUSSION

We proposed an effective approach under the Bayesian framework to simultaneously estimate heterogeneous changepoints for spatially indexed functional time series. Our method differs significantly from the existing methods in two aspects. First, we utilize spatial correlation to synthesize information over the whole spatial domain instead of focusing on a single location, e.g., [Aue et al. \(2018\)](#). Second, we allow spatially varying changepoints for different locations, instead of assuming a single shared changepoint across all locations ([Gromenko et al. 2017](#)). In addition, formulating changepoint estimation into fitting spatially correlated piecewise linear models makes our method more robust to noise and more

powerful for changepoints close to edges. We also show that our method produces precise, informative, and intuitive credible intervals of the changepoint.

Our method currently only focuses on the changepoint estimation, after the rejection region of changepoint detection has been identified. In future work, we would like to develop a more compact approach by incorporating detection and estimation in a single model to remove dependence on auxiliary methods for detection.

### ACKNOWLEDGEMENTS

Bo Li is partially supported by NSF-DMS-1830312, NSF-DGE-1922758, and NSF-DMS-2124576 from the National Science Foundation. We would like to thank the editor and anonymous reviewer for their valuable comments that greatly help improve the manuscript.

*[Received March 2022. Revised September 2022. Accepted October 2022. Published Online October 2022.]*

### APPENDIX A: ASSUMPTIONS

Following Assumption 1 in Aue et al. (2018), we allow the error functions  $\varepsilon_{s,t}(u) \in L^2([0, 1])$  to be weakly dependent in time by assuming they are  $L^p - m$ -approximable for some  $p > 2$ . Assumption 1 essentially means that for any location  $\mathbf{s}$  the error functions  $\varepsilon_{s,t}(u)$  are weakly dependent.

**Assumption 1.** For all spatial locations  $\mathbf{s} \in \mathcal{D}$ , the error functions  $(\varepsilon_{s,t} : t \in \mathbb{Z})$  satisfy

(a) there is a measurable space  $S$  and a measurable function  $g : S^\infty \rightarrow L^2([0, 1])$ , where  $S^\infty$  is the space of infinite sequences  $(\zeta_{s,t}, \zeta_{s,t-1}, \dots)$  with  $(\zeta_{s,t} : t \in \mathbb{Z})$  taking values in  $S$ , such that  $\varepsilon_{s,t} = g(\zeta_{s,t}, \zeta_{s,t-1}, \dots)$  for  $t \in \mathbb{Z}$ , given a sequence of iid random variables  $(\zeta_{s,t} : t \in \mathbb{Z})$ ;

(b) there are  $m$ -dependent sequences  $(\varepsilon_{s,t,m} : t \in \mathbb{Z})$  such that for some  $p > 2$ ,

$$\sum_{m=0}^{\infty} \{E(\|\varepsilon_{s,t} - \varepsilon_{s,t,m}\|^p)\}^{1/p} < \infty,$$

where  $\varepsilon_{s,t,m} = g(\zeta_{s,t}, \dots, \zeta_{s,t-m+1}, \zeta_{s,t,m,t-m}^*, \zeta_{s,t,m,t-m-1}^*, \dots)$  with  $\zeta_{s,t,m,j}^*$  being independent copies of  $\zeta_{s,0}$  independent of  $(\zeta_{s,t} : t \in \mathbb{Z})$ .

This assumption covers most commonly used stationary functional time series models, such as functional auto-regressive and auto-regressive moving average processes. We additionally assume that all error functions are generated from the same distribution as in Assumption 2.

**Assumption 2.** The errors  $(\varepsilon_{s,t} : \mathbf{s} \in \mathcal{D}, t \in \mathbb{Z})$  are identically distributed random fields on  $[0, 1]$ .

Assumption 2 indicates that the error functions at all time points and all locations follow the same distribution. Under Model (1), the only changes observed in a functional time series are due to  $\delta_s(u)$ , i.e., changes in the mean of the functional sequence. Therefore,

all other aspects of the distribution, such as the variance, are required to remain the same. While seemingly restrictive, requiring the moments to not change simultaneously is common in functional time series (Gromenko et al. 2017) and required for identifiability even in univariate change point estimation (Horváth 1993). Practically, Assumption 2 also allows to share variance parameters across spatial locations when estimating the properties of error functions. Finally, we assume that the error process is stationary and isotropic.

**Assumption 3.** The errors  $(\varepsilon_{\mathbf{s},t} : \mathbf{s} \in \mathcal{D}, t \in \mathbb{Z})$  form a mean zero, second-order stationary and isotropic random field. Formally,

$$E\{\varepsilon_{\mathbf{s},t}(u)\} = 0, \quad \text{cov}\{\varepsilon_{\mathbf{s},t}(u), \varepsilon_{\mathbf{s}',t'}(u')\} = C(\|\mathbf{s} - \mathbf{s}'\|, t - t', u - u'),$$

where  $\|\mathbf{s} - \mathbf{s}'\|$  is the Euclidean distance between spatial locations  $\mathbf{s}$  and  $\mathbf{s}'$ .

Assumption 3 essentially means the covariance between any two observations only depends on their distance in each dimension, regardless of their locations and relative orientation.

## REFERENCES

- Aston JA, Kirch C (2012) Detecting and estimating changes in dependent functional data. *J Multivar Anal* 109:204–220
- Aue A, Gabrys R, Horváth L, Kokoszka P (2009) Estimation of a change-point in the mean function of functional data. *J Multivar Anal* 100(10):2254–2269
- Aue A, Horváth L (2013) Structural breaks in time series. *J Time Ser Anal* 34(1):1–16
- Aue A, Rice G, Sönmez O (2018) Detecting and dating structural breaks in functional data without dimension reduction. *J R Stat Soc Ser B Stat Methodol* 80(3):509–529
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodological)* 57(1):289–300
- Berkes I, Gabrys R, Horváth L, Kokoszka P (2009) Detecting changes in the mean of functional observations. *J R Stat Soc Ser B Stat Methodol* 71(5):927–946
- Chernoff H, Zacks S (1964) Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann Math Stat* 35(3):999–1018
- Chiou J-M, Chen Y-T, Hsing T (2019) Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation. *Ann Appl Stat* 13(3):1430–1463
- Fryzlewicz P, Rao SS (2014) Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *J R Stat Soc Ser B Stat Methodol*, pp 903–924
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7(4):457–472
- Geweke JF, et al (1991) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis
- Gromenko O, Kokoszka P (2012) Testing the equality of mean functions of ionospheric critical frequency curves. *J R Stat Soc Ser C Appl Stat* 61(5):715–731
- Gromenko O, Kokoszka P (2013) Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination. *Comput Stat Data Anal* 59:82–94
- Gromenko O, Kokoszka P, Reimherr M (2017) Detection of change in the spatiotemporal mean function. *J R Stat Soc Ser B Stat Methodol* 79(1):29–50
- Gromenko O, Kokoszka P, Zhu L, Sojka J (2012) Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Ann Appl Stat*, pp 669–696

- Haas TC (1995) Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J Am Stat Assoc* 90(432):1189–1199
- Harris T, Li B, Tucker JD (2020). Scalable multiple changepoint detection for functional data sequences. arXiv preprint [arXiv:2008.01889](https://arxiv.org/abs/2008.01889)
- Hoff PD (2011) Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Anal* 6(2):179–196
- Hörmann S, Kokoszka P (2010) Weakly dependent functional data. *Ann Stat* 38(3):1845–1884
- Horváth L (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann Stat*, pp 671–680
- Horváth L, Kokoszka P, Reeder R (2013) Estimation of the mean of functional time series and a two-sample problem. *J R Stat Soc Ser B Stat Methodol* 75(1):103–122
- Kurt B, Yıldız Ç, Ceritli TY, Sankur B, Cemgil AT (2018) A bayesian change point model for detecting sip-based ddos attacks. *Digit Signal Process* 77:48–62
- Lavielle M, Teysiere G (2007) Adaptive detection of multiple change-points in asset price volatility. Long memory in economics. Springer, pp 129–156
- Li X, Ghosal S (2021) Bayesian change point detection for functional data. *J Stat Plan Inference* 213:193–205
- Lund R, Wang XL, Lu QQ, Reeves J, Gallagher C, Feng Y (2007) Changepoint detection in periodic and autocorrelated time series. *J Clim* 20(20):5178–5190
- MacEachern SN, Rao Y, Wu C (2007) A robust-likelihood cumulative sum chart. *J Am Stat Assoc* 102(480):1440–1447
- Matheron G (1963) Principles of geostatistics. *Econ Geol* 58(8):1246–1266
- Page ES (1954) Continuous inspection schemes. *Biometrika* 41(1/2):100–115
- Reeves J, Chen J, Wang XL, Lund R, Lu QQ (2007) A review and comparison of changepoint detection techniques for climate data. *J Appl Meteorol Climatol* 46(6):900–915
- Rice G, Zhang C (2019) Consistency of binary segmentation for multiple change-points estimation with functional data. arXiv preprint [arXiv:2001.00093](https://arxiv.org/abs/2001.00093)
- Shand L, Li B, Park T, Albarracín D (2018) Spatially varying auto-regressive models for prediction of new human immunodeficiency virus diagnoses. *J R Stat Soc Ser C Appl Stat* 67(4):1003–1022
- Shao X, Zhang X (2010) Testing for change points in time series. *J Am Stat Assoc* 105(491):1228–1240
- Sharipov O, Tewes J, Wendler M (2016) Sequential block bootstrap in a hilbert space with application to change point analysis. *Can J Stat* 44(3):300–322
- Stein ML (2012) Interpolation of spatial data: some theory for kriging. Springer
- Sundararajan RR, Pourahmadi M (2018) Nonparametric change point detection in multivariate piecewise stationary time series. *J Nonparametric Stat* 30(4):926–956
- Taylor SJ, Letham B (2018) Forecasting at scale. *Am Stat* 72(1):37–45
- Wald A (1947) *Sequential Analysis*. Wiley, New York
- Yun S, Zhang X, Li B (2020) Detection of local differences in spatial characteristics between two spatiotemporal random fields. *J Am Stat Assoc*, pp 1–16
- Zhang X, Shao X, Hayhoe K, Wuebbles DJ et al (2011) Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron J Stat* 5:1765–1796

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.