Spatio-Temporal Process Monitoring Using Exponentially Weighted Spatial LASSO

Peihua Qiu[1] and Kai Yang[2]

[1]Department of Biostatistics, University of Florida, Gainesville, USA

[2]Division of Biostatistics, Medical College of Wisconsin, Milwaukee, USA

Abstract

Spatio-temporal process monitoring (STPM) has received a considerable attention recently due to its broad applications in environment monitoring, disease surveillance, streaming image processing, and more. Because spatio-temporal data often have complicated structure, including latent spatio-temporal data correlation, complex spatio-temporal mean structure, and nonparametric data distribution, STPM is a challenging research problem. In practice, if a spatio-temporal process has a distributional shift (e.g., mean shift) started at a specific time point, then the spatial locations with the shift are usually clustered in small regions. This kind of spatial feature of the shift has not been considered in the existing STPM literature yet. In this paper, we develop a new STPM method that takes into account the spatial feature of the shift in its construction. The new method combines the ideas of exponentially weighted moving average in the temporal domain for online process monitoring and spatial LASSO in the spatial domain for accommodating the spatial feature of a future shift. It can also accommodate the complicated spatio-temporal data structure well. Both simulation studies and a real-data application show that it can provide a reliable and effective tool for different STPM applications.

Key Words: Data correlation; Exponentially weighted moving average; Nonparametric regression; Spatial LASSO; Spatio-temporal data; Variable selection.

1 Introduction

Spatio-temporal data are common in practice. For instance, disease incidence data of flu, COVID-19 and some other infectious diseases are collected in daily basis at different hospitals and other medical providers in the United States. Daily air quality data of major cities in the United States are also available on the web page of the US Environmental Protection Agency. In these applications, it is important to know whether an underlying spatial process (e.g., the spatial disease incidence)

has a systematic shift in its longitudinal pattern each time after a new batch of spatial data is collected, because such a shift could imply a disease outbreak in the infectious disease example and deterioration of the environment in the air quality example, which are important to detect as early as possible. This is the spatio-temporal process monitoring (STPM) problem that is focused in the current paper. In many STPM applications, such as the ones discussed above, a systematic shift in the spatio-temporal pattern of an underlying spatial process often starts in some small clustered regions (Wang et al. 2018). This kind of spatial feature of the shift should be taken into account in the construction of a control chart for effective STPM. This paper develops such a control chart.

In the literature, there have been some existing methods on STPM. One method is based on the Knox statistic that was first suggested by Knox and Bartlett (1964), and then modified by several authors including Kulldorff and Hjalmars (1999). Its main idea is to detect spatio-temporal clusters by identifying abnormal spatio-temporal interactions from the observed data. Another type of methods are based on the scan statistics (e.g., Kulldorff 1997, Takahashi et al. 2008, Woodall et al. 2008), for detecting spatial or spatio-temporal clusters by using the maximum likelihood ratio test statistics over a collection of the scanning windows. These methods are retrospective, and require various model assumptions on the spatio-temporal data variation and distribution. For instance, seasonal temporal data variation, complex spatial data variation, and spatio-temporal data correlation cannot usually be accommodated by them, and their assumed parametric (e.g., Normal, Poisson, or Negative Binomial) data distributions are rarely valid in practice. In the statistical process control (SPC) literature, there are some recent methods for STPM as well. One such method was suggested by Yan et al. (2018). Its major idea is to decompose the observed spatio-temporal data into three different components, including the functional mean, anomalies, and random errors. The functional mean and anomalies are then represented as linear combinations of certain pre-specified spatio-temporal basis functions, and their estimates are obtained by an optimization algorithm. Selection of the basis functions depends on the expected properties of the functional mean and anomalies, and the random errors are assumed to be independent and identically normally distributed at different locations and times. Another STPM method was suggested by Yang and Qiu (2020), which adopted a cumulative sum (CUSUM) chart that can properly accommodate nonparametric spatio-temporal data variation, spatio-temporal data correlation, and nonparametric data distribution. However, this method treats all spatial locations equally when contructing its charting statistic, and has not taken into account the spatial feature of a future

process shift discussed earlier. Thus, it has room for improvement. To make use of spatial information in the observed data, Wang et al. (2018) suggested a spatial-adaptive sampling procedure that combines the wide search and deep search strategies to speed up detection of clustered shifts. This method is designed for online monitoring of big data streams in cases when only partial spatial information can be observed at each time, and it is assumed that all observations are independent and follow a normal distribution. To get rid of the normality assumption, a generalized version of this method was developed in Xian, Wang and Liu (2018) by using some nonparametric rank-based test statistics. But, this generalized method still assumes the observed data to be independent at different times and locations, and thus cannot accommodate complex correlation structure in a spatio-temporal process.

In many STPM applications, a process shift often starts in some small clustered regions. For instance, outbreaks of infectious diseases like Ebora and COVID-19 often start in some small communities and then spread to a larger population. For such applications, one intuitive idea is to first find the small regions with possible shifts at each observation time during process monitoring, and then focus on the observed data in these small regions to make a decision about the process status at a given observation time. To this end, variable selection methods like the LASSO and adaptive LASSO procedures (cf., Tibshirani 1996, Zou 2006) should be helpful. In the SPC literature, there have been some discussions on process monitoring based on such variable selection methods. See, for instance, Wang and Jiang (2009), Zou and Qiu (2009), and Zou et al. (2012). However, these process monitoring methods are designed mainly for monitoring high-dimensional data or linear profiles; they are not suitable for monitoring spatio-temporal data because the spatio-temporal data structure is not considered by them. Also, these methods assume that process observations at different time points are independent, which is rarely valid in practice. To overcome these limitations, we develop a new control chart for STPM applications. The new method is based on two major ideas. First, to figure out small regions with possible shifts at each observation time, a spatial LASSO procedure is suggested, which can accommodate spatial data structure during variable selection. Second, to make use of all historical data for effective process monitoring, an exponentially weighted smoothing procedure is suggested in the temporal domain to combine all available spatio-temporal data by the current observation time. These two procedures are then combined seamlessly in the spatio-temporal domain for STPM. To accommodate spatio-temporal data variation and spatio-temporal data correlation, the observed data are first decorrelated and

standardized by a data pre-processing procedure. As far as we know, the proposed new method is the first one that takes into account both spatial features of a process shift and complex spatio-temporal data structures (e.g., spatio-temporal data correlation) in the construction of its control chart. Numerical studies show that it is effective for STPM in various cases considered.

The reminder of the paper is organized as follows. The proposed new method for STPM is described in detail in Section 2. Its numerical performance is evaluated in Section 3 by some simulation studies. The new method is demonstrated in an application to monitor the PM 2.5 concentrations in China in Section 4. Finally, some concluding remarks are given in Section 5.

2 Exponentially Weighted Spatial LASSO for Spatio-Temporal Process Monitoring

This section is organized in three parts. First, an IC model to describe the regular spatio-temporal pattern of the process to monitor is discussed in Subsection 2.1. Then, a data pre-processing procedure for STPM is described in Subsection 2.2. Finally, the proposed new control chart for STPM is described in Subsection 2.3.

2.1 IC model and its estimation

Assume that an IC dataset has been collected before online process monitoring, and $\{z(t_i, \boldsymbol{s}_{ij}), j = 1, \ldots, m_i, i = 1, \ldots, n\}$ are the observations in the IC dataset, where $t_i \in [0, T]$ is the ith observation time, $\boldsymbol{s}_{ij} = (\boldsymbol{s}_{x,ij}, \boldsymbol{s}_{y,ij})^T \in \Omega$ is the jth observation location at time t_i, m_i is the number of spatial locations at t_i, and n is the number of observation times. The IC data are assumed to follow the following nonparametric spatio-temporal model:

$$z(t_i, \boldsymbol{s}_{ij}) = \mu(t_i, \boldsymbol{s}_{ij}) + \varepsilon(t_i, \boldsymbol{s}_{ij}), \quad \text{for } j = 1, \ldots, m_i, i = 1, \ldots, n, \tag{1}$$

where $\mu(t_i, \boldsymbol{s}_{ij})$ is the mean of $z(t_i, \boldsymbol{s}_{ij})$, and $\varepsilon(t_i, \boldsymbol{s}_{ij})$ is a zero-mean random error. The spatio-temporal data correlation is described by the covariance function

$$V(t, t'; \boldsymbol{s}, \boldsymbol{s}') = \text{Cov}\left(z(t, \boldsymbol{s}), z(t', \boldsymbol{s}')\right), \quad \text{for any } t, t' \in [0, T], \boldsymbol{s}, \boldsymbol{s}' \in \Omega.$$

For convenience, $V(t, t; \boldsymbol{s}, \boldsymbol{s})$ is also denoted as $\sigma^2(t, \boldsymbol{s})$, for any $(t, \boldsymbol{s}) \in [0, T] \times \Omega$. In the IC model (1), it is assumed that the mean and covariance functions $\mu(t, \boldsymbol{s})$ and $V(t, t'; \boldsymbol{s}, \boldsymbol{s}')$ are continuous.

Other than that, no specific parametric forms are imposed on them and on the data distribution. Thus, this model is flexible. Estimation of $\mu(t,\boldsymbol{s})$, $\sigma^2(t,\boldsymbol{s})$ and $V(t,t';\boldsymbol{s},\boldsymbol{s}')$ has been discussed in Yang and Qiu (2018, 2019). Their estimates are denoted as $\widehat{\mu}(t,\boldsymbol{s})$, $\widehat{\sigma}^2(t,\boldsymbol{s})$ and $\widehat{V}(t,t';\boldsymbol{s},\boldsymbol{s}')$, respectively, and the definitions of these estimates are also given in Appendix.

2.2 Data pre-processing for spatio-temporal process monitoring

After Model (1) is estimated, the regular (or IC) spatio-temporal pattern of the process under monitoring can be described by the estimated mean, variance and covariance functions $\widehat{\mu}(t,\boldsymbol{s})$, $\widehat{\sigma}^2(t,\boldsymbol{s})$ and $\widehat{V}(t,t';\boldsymbol{s},\boldsymbol{s}')$, for $t,t' \in [0,T]$ and $\boldsymbol{s},\boldsymbol{s}' \in \Omega$. Then, we are ready to monitor the spatio-temporal process online for future process observations. To this end, it is often reasonable in practice to assume that the IC spatio-temporal pattern is periodic in time. For instance, the incidence rate of flu has the seasonality, and it is reasonable to assume that its IC spatio-temporal pattern is yearly periodic. For such applications, the time period $[0,T]$ of the IC data should cover at least one whole period. Without loss of generality, it is assumed here that T is the length of one period, and the IC estimates $\widehat{\mu}(t,\boldsymbol{s})$, $\widehat{\sigma}^2(t,\boldsymbol{s})$ and $\widehat{V}(t,t';\boldsymbol{s},\boldsymbol{s}')$ can be extended in the time domain from $[0,T]$ to $[0,\infty)$ to describe the IC spatio-temporal pattern of the process under monitoring.

Let $\{z(t_i^*,\boldsymbol{s}_{ij}^*), j=1,\ldots,m_i^*, i=1,2,\ldots\}$ be the spatio-temporal process observations for online monitoring, where $\{t_i^* \in (T,\infty), i=1,2,\ldots\}$ are observation times and $\{\boldsymbol{s}_{ij}^* \in \Omega, j=1,\ldots,m_i^*, i=1,2,\ldots\}$ are observation locations. These observations could be spatio-temporally correlated and their IC means could be spatio-temporally varied. Thus, they cannot be monitored properly by the conventional SPC charts because the conventional charts are designed for cases when the IC process means are a constant and the IC process observations at different times and/or locations are independent. It has been well demonstrated in the literature that the conventional charts are unreliable for monitoring serially correlated data (cf., Apley and Tsung 2002, Qiu et al. 2020). To overcome this difficulty, we propose a data pre-processing procedure for data decorrelation and standardization, described in the following pseudo algorithm.

- When $i=1$, let $\mathbf{Z}(t_1^*) = (z(t_1^*,\boldsymbol{s}_{11}^*),\ldots,z(t_1^*,\boldsymbol{s}_{1m_1^*}^*))^T$, $\widehat{\varepsilon}(t_1^*,\boldsymbol{s}_{1j}^*) = z(t_1^*,\boldsymbol{s}_{1j}^*) - \widehat{\mu}(t_1^*,\boldsymbol{s}_{1j}^*)$, for $1 \le j \le m_1^*$, and $\widehat{\boldsymbol{\varepsilon}}(t_1^*) = (\widehat{\varepsilon}(t_1^*,\boldsymbol{s}_{11}^*),\ldots,\widehat{\varepsilon}(t_1^*,\boldsymbol{s}_{1m_1^*}^*))^T$. Then, the decorrelated and standardized data at time t_1^* is defined to be $\widehat{\boldsymbol{e}}(t_1^*) = (\widehat{e}(t_1^*,\boldsymbol{s}_{11}^*),\ldots,\widehat{e}(t_1^*,\boldsymbol{s}_{1m_1^*}^*))^T = \widehat{\boldsymbol{\Sigma}}_1^{-1/2}\widehat{\boldsymbol{\varepsilon}}(t_1^*)$, where $\widehat{\boldsymbol{\Sigma}}_1$ is an estimate of $\boldsymbol{\Sigma}_1 = \mathrm{Cov}\,(\mathbf{Z}(t_1^*),\mathbf{Z}(t_1^*))$ obtained from $\widehat{V}(t,t';\boldsymbol{s},\boldsymbol{s}')$.

- For $i > 1$, let $\mathbf{Z}(t_i^*) = (z(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, z(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T$, $\mathbf{Z}_{i-1} = \left(\mathbf{Z}^T(t_1^*), \ldots, \mathbf{Z}^T(t_{i-1}^*)\right)^T$, $\widehat{\varepsilon}(t_i^*, \boldsymbol{s}_{ij}^*) = z(t_i^*, \boldsymbol{s}_{ij}^*) - \widehat{\mu}(t_i^*, \boldsymbol{s}_{ij}^*)$, for $1 \le j \le m_i^*$, $\widehat{\boldsymbol{\varepsilon}}(t_i^*) = (\widehat{\varepsilon}(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \widehat{\varepsilon}(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T$, and $\widehat{\boldsymbol{\varepsilon}}_{i-1} = (\widehat{\boldsymbol{\varepsilon}}^T(t_1^*), \ldots, \widehat{\boldsymbol{\varepsilon}}^T(t_{i-1}^*))^T$. Then, the decorrelated and standardized data at time t_i^* is defined to be

$$\widehat{\boldsymbol{e}}(t_i^*) = \left(\widehat{e}(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \widehat{e}(t_i^*, \boldsymbol{s}_{im_i^*}^*)\right)^T = \widehat{\boldsymbol{\Sigma}}_{ii\cdot i-1}^{-1/2} \left(\widehat{\boldsymbol{\varepsilon}}(t_i^*) - \widehat{\boldsymbol{\Sigma}}_{i-1,i}^T \widehat{\boldsymbol{\Sigma}}_{i-1,i-1}^{-1} \widehat{\boldsymbol{\varepsilon}}_{i-1}\right),$$

where $\widehat{\boldsymbol{\Sigma}}_{ii\cdot i-1} = \widehat{\boldsymbol{\Sigma}}_i - \widehat{\boldsymbol{\Sigma}}_{i-1,i}^T \widehat{\boldsymbol{\Sigma}}_{i-1,i-1}^{-1} \widehat{\boldsymbol{\Sigma}}_{i-1,i}$, and $\widehat{\boldsymbol{\Sigma}}_i$, $\widehat{\boldsymbol{\Sigma}}_{i-1,i}$ and $\widehat{\boldsymbol{\Sigma}}_{i-1,i-1}$ are the estimates of $\boldsymbol{\Sigma}_i = \mathrm{Cov}\left(\mathbf{Z}(t_i^*), \mathbf{Z}(t_i^*)\right)$, $\boldsymbol{\Sigma}_{i-1,i} = \mathrm{Cov}\left(\mathbf{Z}_{i-1}, \mathbf{Z}(t_i^*)\right)$ and $\boldsymbol{\Sigma}_{i-1,i-1} = \mathrm{Cov}\left(\mathbf{Z}_{i-1}, \mathbf{Z}_{i-1}\right)$, respectively, all of which are obtained from $\widehat{V}(t, t'; \boldsymbol{s}, \boldsymbol{s}')$.

By using the above data pre-processing procedure, the original process observations $\{z(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*, i = 1, 2, \ldots\}$ are transformed to the decorrelated and standardized observations $\{\widehat{e}(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*, i = 1, 2, \ldots\}$. It can be checked that the decorrelated and standardized observations would be asymptotically uncorrelated with each other, and each has the asymptotic mean 0 and the asymptotic variance 1 when the process under monitoring is IC.

2.3 New spatio-temporal process monitoring chart

After data pre-processing, we are ready to monitor the decorrelated and standardized observations $\{\widehat{e}(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*\}$, for $i \ge 1$. To this end, by the general principles of sequential process monitoring (cf., Qiu 2014, Chapters 4 and 5), we should use as much available information as possible about the process under monitoring, and we should give less weights to process observations collected farther away from the current observation time. To accomplish these goals, the following exponentially weighted kernel smoothing (EWKS) procedure can be considered:

$$\arg\min_{a \in \mathbb{R}} \sum_{k=1}^{i} \sum_{j=1}^{m_k^*} \left[\widehat{e}(t_k^*, \boldsymbol{s}_{kj}^*) - a\right]^2 K_s\left(d_E(\boldsymbol{s}_{kj}^*, \boldsymbol{s})/h\right) (1-\lambda)^{(t_i^* - t_k^*)}, \tag{2}$$

where t_i^* is the current observation time, $\boldsymbol{s} \in \Omega$ is a given spatial location, $K_s(\cdot)$ is a kernel function, $h > 0$ is a bandwidth, $d_E(\cdot, \cdot)$ denotes the Euclidean distance, and $\lambda \in (0, 1]$ is a weighting parameter. The solution to a of the minimization procedure (2), denoted as $\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s})$, is the kernel smoothing estimate of the mean of $\widehat{e}(t_i^*, \boldsymbol{s})$, denoted as $\mu_{\widehat{e}}(t_i^*, \boldsymbol{s})$. In (2), the kernel function $K_s(\cdot)$ is chosen to be a decreasing function in $[0, \infty)$ so that process observations collected at places farther away from the given location $\boldsymbol{s}$ would receive less weights in the minimization problem. Similarly,

the exponential weights $(1-\lambda)^{(t_i^*-t_k^*)}$ are used in (2) so that process observations collected at times farther away from the current time t_i^* would receive less weights. Also, it can be checked that $\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s})$ is a linear combination of all current and previous process observations, with the weights exponentially decaying for older observations.

From the above intuitive explanation, the EWKS procedure (2) can be regarded as a combination of the exponentially weighted moving average (EWMA) procedure commonly used in the SPC literature and the local kernel smoothing procedure in nonparametric regression (cf., Qiu 2005). The bandwidth h used in (2) can be chosen by minimizing the following leave-one-out cross-validation (CV) criterion:

$$CV(h) = \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} \left[\widehat{e}(t_i^*, \boldsymbol{s}_{ij}^*) - \widetilde{\mu}_{\widehat{e},-(ij)}(t_i^*, \boldsymbol{s}_{ij}^*) \right]^2, \tag{3}$$

where $\widetilde{\mu}_{\widehat{e},-(ij)}(t_i^*, \boldsymbol{s}_{ij}^*)$ is the estimate of $\mu_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)$ obtained by the EWKS procedure (2) after the data point $\widehat{e}(t_i^*, \boldsymbol{s}_{ij}^*)$ is deleted from the estimation.

Intuitively, when the process under monitoring is IC at t_i^*, the values of $\{|\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)|, j = 1, \ldots, m_i^*\}$ should be all small. Thus, a natural charting statistic for process monitoring is $\widetilde{T}_i = \widetilde{\boldsymbol{\mu}}_{\widehat{e},i}^T \widetilde{\Sigma}_{\widetilde{\boldsymbol{\mu}}_{\widehat{e},i}}^{-1} \widetilde{\boldsymbol{\mu}}_{\widehat{e},i}$, where $\widetilde{\boldsymbol{\mu}}_{\widehat{e},i} = (\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T$ and $\widetilde{\Sigma}_{\widetilde{\boldsymbol{\mu}}_{\widehat{e},i}}$ is the estimated covariance matrix of $\widetilde{\boldsymbol{\mu}}_{\widehat{e},i}$. But, this statistic treats all observation locations $\{\boldsymbol{s}_{ij}^*\}$ equally, and it would not be effective for many STPM applications in which a process shift starts in small clustered regions in the design space Ω. To overcome this limitation, one idea is to first detect the small regions with possible process shifts at t_i^*, and then construct the charting statistic based on the detected small regions. To this end, the LASSO variable selection procedure (Tibshirani 1996) can be considered. However, considering the fact that the spatial locations with possible process shifts are usually spatially clustered small regions, instead of a set of isolated spatial locations scattered in Ω, we suggest using a spatial LASSO procedure (e.g., Samarov et al. 2015) that takes into account the spatial data structure when detecting small clustered regions with possible process shifts. By combining the spatial LASSO idea with the EWKS procedure (2), we suggest the following penalized EWKS procedure for estimating $\{\mu_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)\}$:

$$\begin{aligned} &\underset{a_1,\ldots,a_{m_i^*}}{\arg\min} \sum_{j=1}^{m_i^*} \sum_{k=1}^{i} \sum_{l=1}^{m_k^*} \left[\widehat{e}(t_k^*, \boldsymbol{s}_{kl}^*) - a_j \right]^2 K_s\left(d_E(\boldsymbol{s}_{kl}^*, \boldsymbol{s}_{ij}^*)/h \right) (1-\lambda)^{(t_i^*-t_k^*)} \\ &+ \gamma_1 \sum_{j=1}^{m_i^*} \varpi_{1j} |a_j| + \gamma_2 \sum_{j=1}^{m_i^*} \varpi_{2j} \left| a_j - \frac{\sum_{1\le l\le m_i^*} K_s(d_E(\boldsymbol{s}_{ij}^*, \boldsymbol{s}_{il}^*)/h) a_l}{\sum_{1\le l\le m_i^*} K_s(d_E(\boldsymbol{s}_{ij}^*, \boldsymbol{s}_{il}^*)/h)} \right|, \end{aligned} \tag{4}$$

where $\varpi_{1j} = 1/|\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)|$ and $\varpi_{2j} = 1/\left|\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*) - \frac{\sum_{1\le l\le m_i^*} K_s(d_E(\boldsymbol{s}_{ij}^*, \boldsymbol{s}_{il}^*)/h)\widetilde{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{il}^*)}{\sum_{1\le l\le m_i^*} K_s(d_E(\boldsymbol{s}_{ij}^*, \boldsymbol{s}_{il}^*)/h)}\right|$ are the adaptive LASSO weights (cf., Zou 2006), and $\gamma_1, \gamma_2 > 0$ are two tuning parameters. In (4), the last term is the adaptive spatial LASSO penalty. In this term, $\left|a_j - \frac{\sum_{1\le l\le m_i^*} K_s(d_E(\boldsymbol{s}_{ij}^*, \boldsymbol{s}_{il}^*)/h)a_l}{\sum_{1\le l\le m_i^*} K_s(d_E(\boldsymbol{s}_{ij}^*, \boldsymbol{s}_{il}^*)/h)}\right|$ measures the difference between the estimated process mean at the jth location with a weighted average of the estimated process means at neighboring locations at the current time t_i^*. Thus, this term penalizes the difference among estimated means at neighboring locations, which is desirable for applications in which process shifts are spatially clustered. By a reparameterization, it can be shown that the minimization problem (4) is a generalized LASSO procedure, and it can be solved by using the path algorithm suggested by Tibshirani and Taylor (2011). The estimates from (4) are denoted as $\{\widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*\}$.

It should be pointed out that our proposed penalized EWKS procedure (4) is quite different from the existing spatial LASSO procedures in the literature (e.g., Huang et al. 2010, Samarov et al. 2015). First, most existing spatial LASSO procedures are designed for spatial model estimation, while the procedure (4) is design for sequential process monitoring. So, the exponential weights $(1-\lambda)^{(t_i^* - t_k^*)}$ used in the first term of (4) to combine the collected data at different time points is completely new. Second, to capture the spatial structure, most existing spatial LASSO procedures either impose a spatial prior distribution on all coefficients $\{a_j, j = 1, 2, \ldots, m_i^*\}$ used in (4) when defining their objective functions, or first compute the L_1 difference between a_j and its neighboring coefficients in a pre-specified neighborhood for each j and then use their average as a penalty term. In (4), this has been simplified by using the kernel function $K_s(\cdot)$. Third, the kernel function $K_s(\cdot)$ is also used in the first term of (4) to facilitate data smoothing, which is not considered in most existing spatial LASSO procedures.

In the penalized EWKS procedure (4), the bandwidth h can be chosen by the CV procedure (3), and the two tuning parameters γ_1 and γ_2 can be chosen by the following BIC procedure (Schwarz 1978):

$$BIC(\gamma_1, \gamma_2) = \log\left(\sum_{j=1}^{m_i^*}\left[\widehat{e}(t_i^*, \boldsymbol{s}_{ij}^*) - \widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)\right]^2\right) + \log(m_i^*)\sum_{j=1}^{m_i^*} I(\widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*) \neq 0). \tag{5}$$

Then, γ_1 and γ_2 are chosen by minimizing $BIC(\gamma_1, \gamma_2)$. The BIC procedure (5) is used here because of its consistency property in selecting the true model (cf., Yang 2005).

From the construction of the penalized EWKS procedure (4), it can be seen that the estimates

$\{\widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*\}$ have taken into account the previous process observations through its EWKS component (i.e., the first term in (4)). The procedure (4) is an adaptive LASSO procedure and its shrinkage property guarantees that many elements in $\{\widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*\}$ would be zero and only those at locations with process shifts would be non-zero. So, to detect process shifts, it is natural to consider

$$T_i = \widehat{\boldsymbol{\mu}}_{\widehat{e},i}^T \widehat{\Sigma}_{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}}^{-1} \widehat{\boldsymbol{\mu}}_{\widehat{e},i},$$

where $\widehat{\boldsymbol{\mu}}_{\widehat{e},i} = (\widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T$, and $\widehat{\Sigma}_{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}}$ is an estimate of the covariance matrix of $\widehat{\boldsymbol{\mu}}_{\widehat{e},i}$. Then, the charting statistic of our proposed new method for STPM is the following standardized version of T_i:

$$ST_i = \frac{T_i - \widehat{\mathrm{E}}(T_i)}{\sqrt{\widehat{\mathrm{Var}}(T_i)}}, \quad \text{for } i \geq 1, \tag{6}$$

where $\widehat{\mathrm{E}}(T_i)$ and $\widehat{\mathrm{Var}}(T_i)$ are the estimates of the mean and variance of T_i, respectively. The chart gives a signal of process shift at time t_i^* if

$$ST_i > L, \tag{7}$$

where $L > 0$ is a control limit. This chart is called the exponentially weighted spatial LASSO (EWSL) chart hereafter.

In the SPC literature, the performance of an SPC chart is usually evaluated by the IC average run length (ARL), denoted as ARL_0, which is defined to be the average number of time points from the beginning of process monitoring to the signal time when the process is IC, and the out-of-control (OC) ARL, denoted as ARL_1, which is the average number of time points from the occurrence of a process shift to the signal time. When the process is IC, a control chart with a larger ARL_0 value would have a smaller false alarm rate. When the process becomes OC, a control chart with a smaller ARL_1 value would have a better OC performance since it can detect a shift more quickly. Usually, if the ARL_0 value is set to be larger when designing a control chart, then the ARL_1 value would be larger as well for detecting a given shift. So, in practice, the ARL_0 value is often pre-specified when designing a control chart. Then, a chart with a smaller ARL_1 value when detecting a given shift is considered better. It should be pointed out that ARL_0 and ARL_1 values are appropriate to use only in cases with equally spaced observation times. In cases when the observation times are unequally spaced, the IC average time to signals, denoted as ATS_0, and the OC average time to signals, denoted as ATS_1, should be considered (cf., Reynolds et al. 1990).

To use the control chart (6)-(7), the two parameters λ and L need to be specified properly beforehand. Regarding λ, it is similar to the regular weighting parameter used in an EWMA chart. In the SPC literature, it has been well demonstrated that a large value of the weighting parameter is good for detecting relatively large shifts, and a small value is good for detecting relatively small shifts (cf., Qiu 2014, Chapter 5). Based on our extensive numerical studies, this is also valid here if λ is chosen within the range $(0, 0.1]$. It should be pointed out that, in the current penalized EWKS procedure (4), if λ is chosen too large, then the estimate $\widehat{\mu}_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)$ would have a relatively large variability in estimating $\mu_{\widehat{e}}(t_i^*, \boldsymbol{s}_{ij}^*)$, for all i and j, because only a few previous process observations will be used in function estimation in such cases. See Qiu et al. (2010) for a related discussion on this issue. Based on these considerations and our numerical experience, we suggest choosing $\lambda \in [0.02, 0.1]$. Once λ is chosen, the control limit L can be selected based on an IC dataset such that a pre-specified ARL_0 value is reached, which can be accomplished by a block bootstrap procedure described below. From this block bootstrap procedure, the estimates $\widehat{\Sigma}_{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}}$, $\widehat{\mathrm{E}}(T_i)$ and $\widehat{\mathrm{Var}}(T_i)$ used in (6) can also be computed.

i) Decorrelate and standardize the IC data by the data pre-processing procedure discussed in Subsection 2.2, and obtain the decorrelated and standardized observations $\{\widehat{\boldsymbol{e}}(t_i), i = 1, \ldots, n\}$. Then, these decorrelated and standardized observations are used to form $n - \tau + 1$ overlapping blocks $\{B_k, k = 1, \ldots, n - \tau + 1\}$, where τ is a block size and the kth block is defined to be $B_k = \{\widehat{\boldsymbol{e}}(t_i), i = k, k + 1, \ldots, k + \tau - 1\}$.

ii) Sample with replacement from $\{B_k, k = 1, \ldots, n - \tau + 1\}$, the selected blocks are placed one after another, and they form the bth bootstrap sample of process observations for STPM, denoted as $\{\widehat{\boldsymbol{e}}_i^{(b)}, i \geq 1\}$. Repeat this block bootstrap sampling procedure for B times and obtain B bootstrap samples.

iii) Monitor the bth bootstrap sample and compute the values of $\{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}, i \geq 1\}$, denoted as $\{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}^{(b)}, i \geq 1\}$, for $b = 1, \ldots, B$. For each i, use the sample covariance of $\{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}^{(b)}, 1 \leq b \leq B\}$ to estimate $\Sigma_{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}}$, and the resulting estimate is denoted as $\widehat{\Sigma}_{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}}$.

iv) For each b and each i, compute $T_i^{(b)} = (\widehat{\boldsymbol{\mu}}_{\widehat{e},i}^{(b)})^T \widehat{\Sigma}_{\widehat{\boldsymbol{\mu}}_{\widehat{e},i}}^{-1} \widehat{\boldsymbol{\mu}}_{\widehat{e},i}^{(b)}$. Then, $\mathrm{E}(T_i)$ and $\mathrm{Var}(T_i)$ can be estimated by the sample mean and sample variance of $\{T_i^{(b)}, 1 \leq b \leq B\}$, for each i. The estimates are denoted as $\widehat{\mathrm{E}}(T_i)$ and $\widehat{\mathrm{Var}}(T_i)$.

v) For each b, the charting statistic values $\{ST_i, i \geq 1\}$ can be computed by (10) and are denoted

as $\{ST_i^{(b)}, i \geq 1\}$. For a given control limit L, record the IC run length to be $\text{RL}_0^{(b)}(L) = \min\{i, ST_i^{(b)} > L\}$. Then, the ARL_0 value is computed to be $\text{ARL}_0(L) = \frac{1}{B}\sum_{b=1}^{B} \text{RL}_0^{(b)}(L)$.

vi) Use the bisection method to search for L such that the computed value of $\text{ARL}_0(L)$ equals the pre-specified value for ARL_0.

It should be pointed out that the IC dataset for estimating the IC model (1) could be different from the one for determining the control limit L. Based on extensive numerical studies, we find that the proposed method could have a more reliable IC performance, in the sense that its calculated ARL_0 value would be closer to the nominal ARL_0 level, if two different IC datasets are used for these two purposes. In addition, we would like to mention that the estimated covariance matrix $\widehat{\Sigma}_{\widehat{\boldsymbol{\mu}}_{\widehat{\varepsilon},i}}$ obtained from the above block bootstrap procedure may not be positive definite. To modify it to be positive definite, we suggest using the matrix modification procedure discussed in Yang and Qiu (2019) that is based on matrix projection to the space of all positive definite matrices, which can be implemented by using the command **nearPD()** in the R package *Matrix*.

3 Simulation Studies

In this section, we evaluate the numerical performance of the proposed chart EWSL, using Monte Carlo simulations. In the simulation examples, let us assume that $[0, T] = [0, 1]$, the IC observation times are $\{t_i = i/n, i = 1, \ldots, n\}$, and the observation locations at each observation time are $\{\boldsymbol{s}_j, j = 1, \ldots, m\}$, which are unchanged over time and equally spaced in $\Omega = [0, 1] \times [0, 1]$. In the IC model (1), it is assumed that $(n, m) = (300, 64)$ unless stated otherwise and

$$\mu(t, \boldsymbol{s}) = \cos(2\pi t) + \exp[-\{(s_x - 0.5)^2 + (s_y - 0.5)^2\}/2] + 1.$$

The IC mean function $\mu(t, \boldsymbol{s})$ is presented in the 1st column of Figure 1 when $t = 0.3$ (1st row) and 0.6 (2nd row). The random errors are generated in one of the following six ways with different spatio-temporal data correlation and distribution.

Case I: The random errors $\{\varepsilon(t_i, \boldsymbol{s}_j)\}$ are independent and identically distributed (i.i.d.) with the common distribution to be standard normal.

Cases II and III: Let $\boldsymbol{\varepsilon}(t_i) = (\varepsilon(t_i, \boldsymbol{s}_1), \ldots, \varepsilon(t_i, \boldsymbol{s}_m))^T$. Then, $\boldsymbol{\varepsilon}(t_i)$ is generated from the vector AR(1) model $\boldsymbol{\varepsilon}(t_i) = \rho_t \boldsymbol{\varepsilon}(t_{i-1}) + (1 - \rho_t^2)^{1/2} \boldsymbol{\eta}(t_i)$, where $\boldsymbol{\eta}(t_i) = (\eta(t_i, \boldsymbol{s}_1), \ldots, \eta(t_i, \boldsymbol{s}_m))^T$ is

a realization of a temporally independent Gaussian spatial process whose spatial correlation is described by the covariance function $\mathrm{Cov}\,(\eta(t_i, \boldsymbol{s}_j), \eta(t_i, \boldsymbol{s}_l)) = \exp\{-d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/\rho_s\}$, and $\rho_t, \rho_s > 0$ are constants. Then, it can be checked that, for any $1 \le i, k \le n$ and $1 \le j, l \le m$, the covariance between $z(t_i, \boldsymbol{s}_j)$ and $z(t_k, \boldsymbol{s}_l)$ is

$$V(t_i, t_k; \boldsymbol{s}_j, \boldsymbol{s}_l) = \rho_t^{|k-i|} \exp\{-d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/\rho_s\}.$$

It can be seen that the parameters (ρ_t, ρ_s) control the temporal and spatial data correlation; the larger their values, the stronger the data correlation. To consider cases with different levels of spatio-temporal data correlation, (ρ_t, ρ_s) are chosen to be (0.25, 0.1) in Case II and (0.5, 0.2) in Case III.

Case IV: The random errors are generated from the vector AR(2) model $\boldsymbol{\varepsilon}(t_i) = 0.5\boldsymbol{\varepsilon}(t_{i-1}) + 0.25\boldsymbol{\varepsilon}(t_{i-2}) + \boldsymbol{\eta}(t_i)$, where $\boldsymbol{\eta}(t_i)$ is the the same as that in Case II, except that its spatial covariance function is $\rho(\boldsymbol{s}, \boldsymbol{s}') = \exp\left[-10[(\boldsymbol{s} - \boldsymbol{s}')^T \mathbf{M} (\boldsymbol{s} - \boldsymbol{s}')]^{1/2}\right]$, for any $\boldsymbol{s}, \boldsymbol{s}' \in \Omega$, where $\mathbf{M} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. The spatial data correlation described by this covariance function is nonstationary.

Case V: The random errors are generated from the model $\boldsymbol{\varepsilon}(t_i) = 0.25\boldsymbol{\varepsilon}(t_{i-1}) + (1 - 0.25^2)^{1/2}\boldsymbol{\eta}(t_i)$, where $\boldsymbol{\eta}(t_i) = \boldsymbol{\eta}_1(t_i) + \boldsymbol{\eta}_2(t_i)$, $\{\boldsymbol{\eta}_1(t_i)\}$ are temporally independent Gaussian spatial processes whose spatial correlation is described by $\mathrm{Cov}(\eta_1(t_i, \boldsymbol{s}_j), \eta_1(t_i, \boldsymbol{s}_l)) = \exp\{-10 d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)\}$, and $\{\boldsymbol{\eta}_2(t_i, \boldsymbol{s}_j)\}$ are i.i.d. with the standardized $t(3)$ distribution.

Case VI: The random errors are generated from the model $\boldsymbol{\varepsilon}(t_i) = 0.25\boldsymbol{\varepsilon}(t_{i-1}) + (1 - 0.25^2)^{1/2}\boldsymbol{\eta}(t_i)$, where $\boldsymbol{\eta}(t_i) = \boldsymbol{\eta}_1(t_i) + \boldsymbol{\eta}_2(t_i)$, $\{\boldsymbol{\eta}_1(t_i)\}$ are generated in the same way as that in Case V, and $\{\boldsymbol{\eta}_2(t_i, \boldsymbol{s}_j)\}$ are i.i.d. with the standardized $\chi^2(3)$ distribution.

In each of the above six cases, the random errors $\{\varepsilon(t_i, \boldsymbol{s}_j)\}$ are all standardized to have mean 0 and variance 1. From the above description, it can be seen that Case I is the conventional case without any spatio-temporal data correlation that is considered by many existing methods. Cases II and III have the AR(1) temporal data correlation and a stationary spatial data correlation, with different levels of spatio-temporal data correlation. Case IV has the AR(2) temporal data correlation and a nonstationary spatial data correlation. Case V is the same as Case II, except that its AR(1) model

has an extra error term with the standardized $t(3)$ distribution, which is heavy-tailed. Case VI is the same as Case V, except that the extra error term has the standardized $\chi^2(3)$ distribution which is skewed.

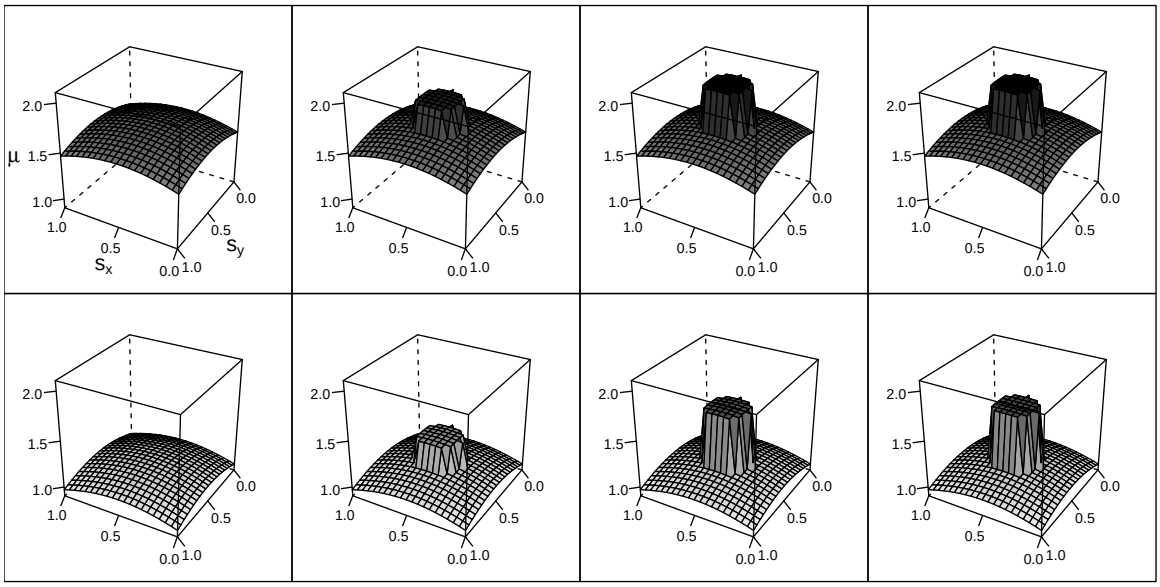

Figure 1: IC mean function (1st column) and three types of OC mean functions (2nd, 3rd and 4th columns) when $t = 0.3$ (1st row) and 0.6 (2nd row).

After the process under monitoring becomes OC, the process observations are assumed to follow the OC model

$$z(t, \boldsymbol{s}) = \mu(t, \boldsymbol{s}) + \delta(t, \boldsymbol{s}) + \varepsilon(t, \boldsymbol{s}), \quad \text{for } (t, \boldsymbol{s}) \in [0, 1] \times \Omega,$$

where $\delta(t, \boldsymbol{s})$ is the shift function. In this section, the following three types of shift functions are considered, and each type has four different shift magnitudes:

Type (i): $\delta(t, \boldsymbol{s}) = 0.25\nu \mathrm{I}(\boldsymbol{s} \in \Delta)$, for $\nu = 1, 2, 3, 4$;

Type (ii): $\delta(t, \boldsymbol{s}) = (0.25 + 0.5\nu t)\mathrm{I}(\boldsymbol{s} \in \Delta)$, for $\nu = 1, 2, 3, 4$;

Type (iii): $\delta(t, \boldsymbol{s}) = \{0.25 + 0.5t + 0.1\nu d_E(\boldsymbol{s}, \boldsymbol{s}_0)\}\, \mathrm{I}(\boldsymbol{s} \in \Delta)$, for $\nu = 1, 2, 3, 4$,

where $\boldsymbol{s}_0 = (0.5, 0.5)^T$ and $\Delta = \{\boldsymbol{s}, d_E(\boldsymbol{s}, \boldsymbol{s}_0) \leq d\} \subset \Omega$ denotes the spatial region where shifts occur. Without further specification, d is chosen to be 0.2. After the occurrence of a shift with size $\delta(t, \boldsymbol{s})$, the OC mean function becomes $\mu(t, \boldsymbol{s}) + \delta(t, \boldsymbol{s})$. The three different types of OC mean

function with $\nu = 1$ are shown in the 2nd, 3rd and 4th columns of Figure 1, respectively, at $t = 0.3$ (1st row) or $t = 0.6$ (2nd row). From their expressions and Figure 1, it is clear that Type (i) shifts keep unchanged across both space and time, Type (ii) shifts can change over time but not space, and Type (iii) shifts can change over both space and time. So, the three types of shift functions represent constant, time-varying and space/time-varying shifts in the small region Δ, respectively.

In all simulation examples, it is assumed that the IC mean and covariance functions are unknown, and need to be estimated from an IC dataset $\{z(t_i, \boldsymbol{s}_j), j = 1, \ldots, m, i = 1, \ldots, n\}$. To determine the control limit L, another IC dataset of the same size is generated, and then L is chosen by the block bootstrap procedure discussed at the end of Subsection 2.3 with the bootstrap sample size $B = 1,000$ and the block size τ. The nominal ARL_0 level of the chart EWSL is fixed at 200. In a given case, its actual ARL value is computed based on 1,000 replicated simulations of online monitoring. Because such an ARL value depends on the IC datasets used for estimating the IC model and for determining the control limit L, which are random, the entire simulation described above, from generation of the IC datasets, estimation of the IC model, determination of L, to computation of the actual ARL value, is repeated for 100 times. Then, the average of the 100 computed ARL values is used as the final estimate of the true ARL value. The standard error of this final estimate can be computed accordingly.

3.1 Impact of (n, m) and τ on the performance of EWSL

Performance of the proposed chart EWSL may depend on the IC sample size (n, m) since the IC dataset is used for estimating the IC model and the estimated IC model is used in constructing the chart EWSL. It may depend on the block size τ as well, since the block bootstrap procedure discussed in Subsection 2.3 is used for determining the control limit L and some other quantities used in the charting statistic ST_i in (6). Such dependence is studied in this part. To this end, we consider cases when $\lambda = 0.1$, $\tau = 5$, n changes from 100 to 800, and m changes among 36, 64, 100 and 225. The nominal ARL_0 level of the chart is fixed at 200. Then, the computed actual ARL_0 values of the chart in Cases I-VI are presented in Figure 2, where the shaded region in each plot denotes ARL_0 values that are within 5% of the nominal ARL_0 level. From the figure, it can be seen that (i) the IC performance of the chart gets better when m and/or n increase, (ii) when $n \geq 300$ and $m \geq 64$, the computed ARL_0 values are within 5% of the nominal level in all cases considered, and (iii) by comparing the results in Cases I-III, the results get worse when the spatio-temporal

data correlation becomes stronger from Case I to Case III. So, this example confirms that in Cases I-VI, i) the chart EWSL is reliable to use when $n \geq 300$ and $m \geq 64$, and ii) when one of n and m is larger than the recommended value, the other quantity could be smaller than its recommended value to have a reliable IC performance of EWSL (e.g., cases when $n \geq 500$ and $m \geq 36$).

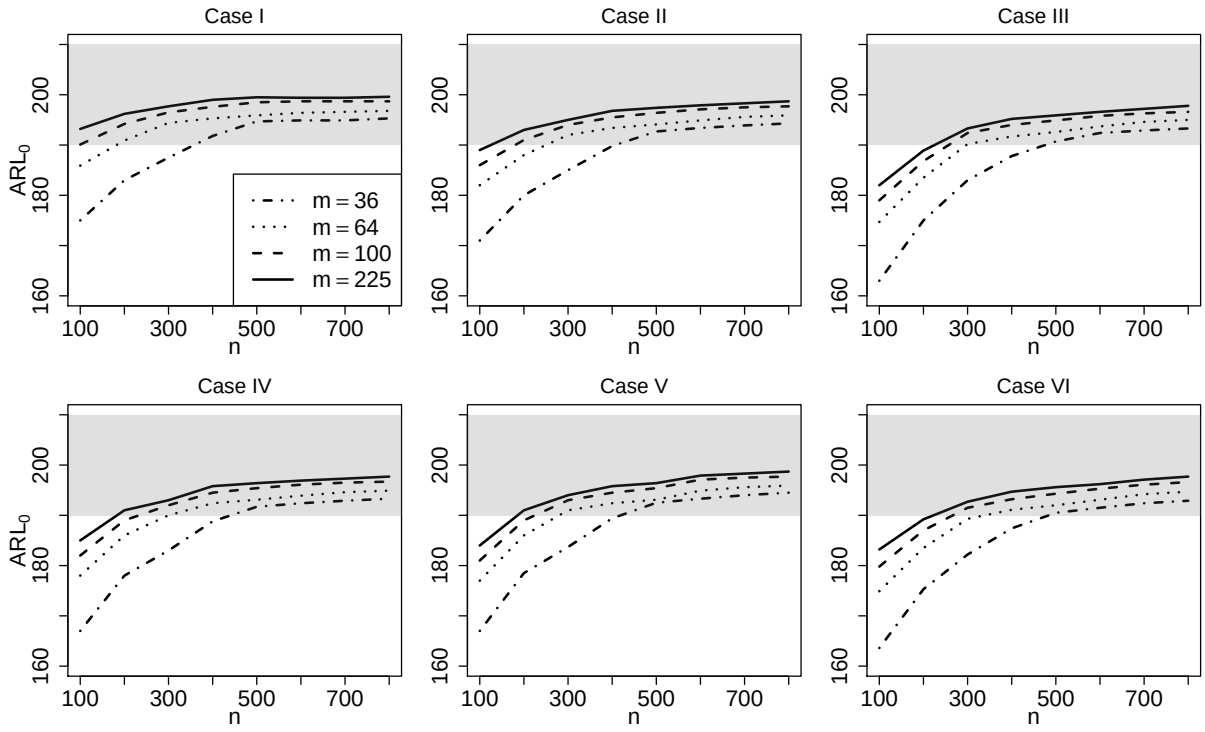

Figure 2: Calculated ARL_0 values of the chart EWSL when $\lambda = 0.1$, $\tau = 5$, $m = 36, 64, 100$ or 225, n changes from 100 to 800, and the nominal ARL_0 level is 200. In each plot, the shaded area denotes ARL_0 values that are within 5% of the nominal ARL_0 level of 200.

To study the dependence of the IC performance of EWSL on the block size τ, we consider cases when $\lambda = 0.1$, $(n, m) = (300, 64)$, τ changes from 1 to 8, and other setups are the same as those in Figure 2. The related results are shown in Figure 3. It is clear from the figure that (i) EWSL has a satisfactory IC performance when $\tau \in [4, 8]$ in all simulation cases, and (ii) by looking at the left panel of the figure, the performance of the chart becomes better when the data correlation gets weaker from Case III to Case I, which is intuitively reasonable. Based on these results, it seems reasonable to choose τ in the range $[4, 8]$.

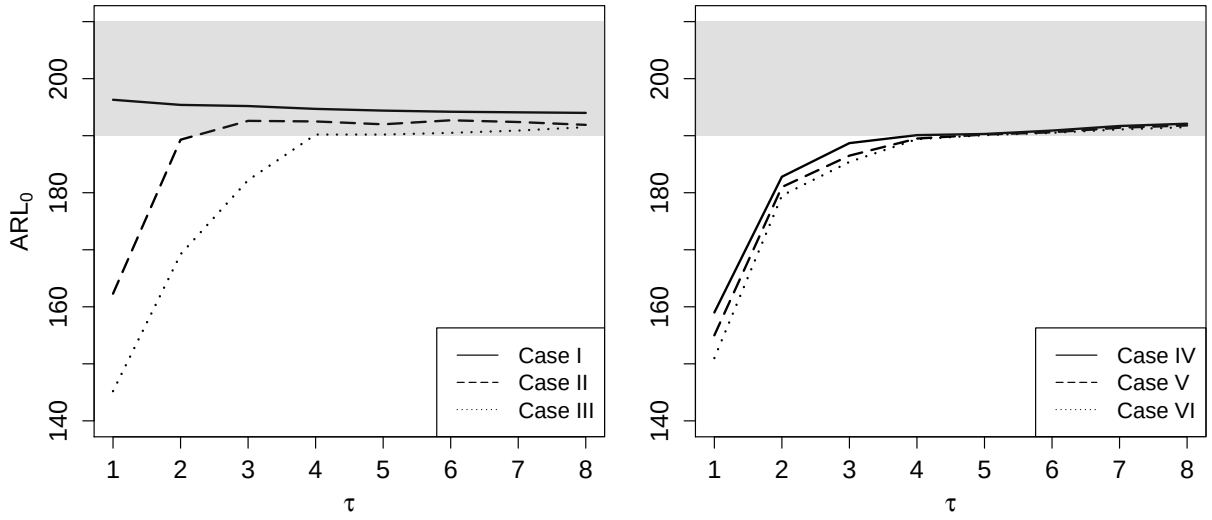

Figure 3: Calculated ARL_0 values of the chart EWSL when $\lambda = 0.1$, $(n, m) = (300, 64)$, τ changes from 1 to 8, and the nominal ARL_0 value is 200. In each plot, the shaded area denotes ARL_0 values that are within 5% of the nominal ARL_0 level of 200.

3.2 Comparison with alternative methods

In this part, we compare the proposed chart EWSL with some representative alternative charts. To this end, the following five alternative charts are considered.

- Yan et al. (2018) suggested a real-time monitoring method for spatio-temporal processes based on the spatio-temporal smooth sparse decomposition (SSD). This method can accommodate space/time-varying mean structure and it has been shown effective for detecting sparse anomalies. However, it is assumed that the random noises are independent and normally distributed. This method is denoted as ST-SSD.

- Zhao et al. (2011) proposed a STPM method by using the kernel smoothing (KS) procedure to estimate the IC space/time-varying mean function. In process monitoring, the observed data are first compared with the IC mean function to compute the raw residuals. Then, at each location, the raw residuals are decorrelated by an AR(2) time series model to obtain the model-based residuals. Finally, this method gives a signal of process shift if a model-based residual exceeds a threshold value c, where c is chosen by a parametric bootstrap procedure under the assumption that the observations at different locations are independent and the

model-based residuals follow a normal distribution. This method is denoted as ST-KS.

- Yang and Qiu (2020) proposed a CUSUM chart for online monitoring of spatio-temporal data, denoted as ST-CUSUM. This method does not require any parametric assumptions on the spatio-temporal data variation, spatio-temporal data correlation, or data distribution. So, this method should be reliable to use in many applications. However, it may not be effective for detecting spatially clustered shifts that start in some small regions, since it treats all observation locations equally in its chart construction.

- Zou and Qiu (2009) suggested a LASSO-based EWMA chart, denoted as LEWMA, for monitoring multivariate processes. It has been shown effective for detecting shifts that occur in a small number of quality variables. This method, however, cannot accommodate spatio-temporal data variation and/or spatio-temporal data correlation. To make the comparison fair, we have applied this chart to the decorrelated and standardized data $\{\widehat{e}(t_i^*, \boldsymbol{s}_j), j = 1, \ldots, m, i = 1, 2, \ldots\}$ in this section. For simplicity in notation, the resulting chart is still denoted as LEWMA. Obviously, this chart is not designed for monitoring spatial data, and thus may not be effective for detecting spatially clustered shifts.

- Consider a modified version of the proposed EWSL chart, described below. First, construct an EWMA chart $\xi(t_i^*, \boldsymbol{s}_j) = \lambda\widehat{e}(t_i^*, \boldsymbol{s}_j) + (1-\lambda)\xi(t_{i-1}^*, \boldsymbol{s}_j)$, where $\xi(t_0^*, \boldsymbol{s}_j) = 0$ and $\lambda > 0$ is a weighting parameter. Then, the mean vector $\boldsymbol{\mu}_{\widehat{e},i} = (\mu_{\widehat{e}}(t_i^*, \boldsymbol{s}_1), \ldots, \mu_{\widehat{e}}(t_i^*, \boldsymbol{s}_m))^T$ is estimated by the following adaptive spatial LASSO (ASL) procedure:

$$\arg\min_{a_1,\ldots,a_m} \sum_{j=1}^{m} (\xi(t_i^*, \boldsymbol{s}_j) - a_j))^2 + \gamma_1 \sum_{j=1}^{m} \varpi_{1j}|a_j| + \gamma_2 \sum_{j=1}^{m} \varpi_{2j} \left| a_j - \frac{\sum_{l=1}^{m} K_s(d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/h) a_l}{\sum_{l=1}^{m} K_s(d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/h)} \right|$$

  where ϖ_{1j} and ϖ_{2j} are the adaptive weights defined below Expression (4), and γ_1 and γ_2 are two tuning parameters chosen by the BIC criterion (5). Other components of the modified chart are the same as those of EWSL. The major difference between this modified chart and EWSL is that the EWMA procedure in the time domain and the ASL procedure in the space domain are separated in the modified chart, while these two components are mixed together in EWSL. To use the modified chart, spatial locations are required to be unchanged over time. This modified chart is denoted as ASL.

Comparison of IC Performance. We first compare the IC performance of the six control charts described above. In the charts LEWMA, ASL and EWSL, the weighting parameter λ is chosen to

be 0.02, 0.05 or 0.1. The allowance constant ϕ in the ST-CUSUM chart is chosen to be 0.1, 0.3 or 0.5. For all the charts, we consider cases when $(n, m) = (300, 64)$ and the nominal ARL_0 value is 200. The control limits of the charts ST-CUSUM, LEWMA, ASL and EWSL are determined by the block bootstrap procedures with the bootstrap sample size $B = 1,000$ and the block size $\tau = 5$. The threshold value c in the ST-KS method is determined by the parametric bootstrap procedure, as discussed in Zhao et al. (2011). The control limit of the chart ST-SSD is determined by the simulation-based procedure discussed in Yan et al. (2018). The calculated actual ARL_0 values of the six charts and the corresponding standard errors are presented in Table 1. From the table, it can be seen that (i) the four charts EWSL, ASL, LEWMA and ST-CUSUM have a satisfactory IC performance in most cases considered, since their calculated ARL_0 values are close to the nominal level of 200, (ii) the two charts ST-KS and ST-SSD have a worse IC performance, compared to the charts EWSL, ASL, LEWMA and ST-CUSUM, in Cases II-VI when there is spatio-temporal data correlation, and (iii) by comparing the results in Cases I-III, the IC performance of all six charts becomes worse from Case I to Case III when the data correlation gets stronger, although the impact of the data correlation on the four charts EWSL, ASL, LEWMA and ST-CUSUM seems quite small.

Comparison of OC Performance. We then study the OC performance of the six control charts in Cases I-VI for detecting the three types of shifts described earlier. In each case, let $(n, m) = (300, 64)$, $d = 0.2$, and the nominal ARL_0 level be 200. The other setups are the same as those in Table 1. To make the comparison fair, the control limit or the threshold value of each chart has been adjusted so that its actual ARL_0 value equals the nominal level of 200. Also, procedure parameters in each chart (e.g., λ in LEWMA, ASL and EWSL) are chosen such that the ARL_1 value reaches the minimum for detecting a given shift. Namely, the optimal ARL_1 values of different charts are compared here, because it has been pointed out in the literature that the OC performance of different charts with pre-specified parameter values may not be comparable (e.g., Qiu et al. 2020). The optimal ARL_1 values of the six charts are presented in Figure 4. From the figure, we can have the following conclusions. (i) The charts ST-CUSUM, LEWMA, ASL and EWSL perform better than the charts ST-SSD and ST-KS in most cases considered, which is reasonable because some assumptions in ST-SSD and ST-KS are violated here. (ii) LEWMA has a better performance than ST-CUSUSM, which confirms the benefit to build a LASSO-based variable selection procedure into a control chart. (iii) the charts ASL and EWSL perform better

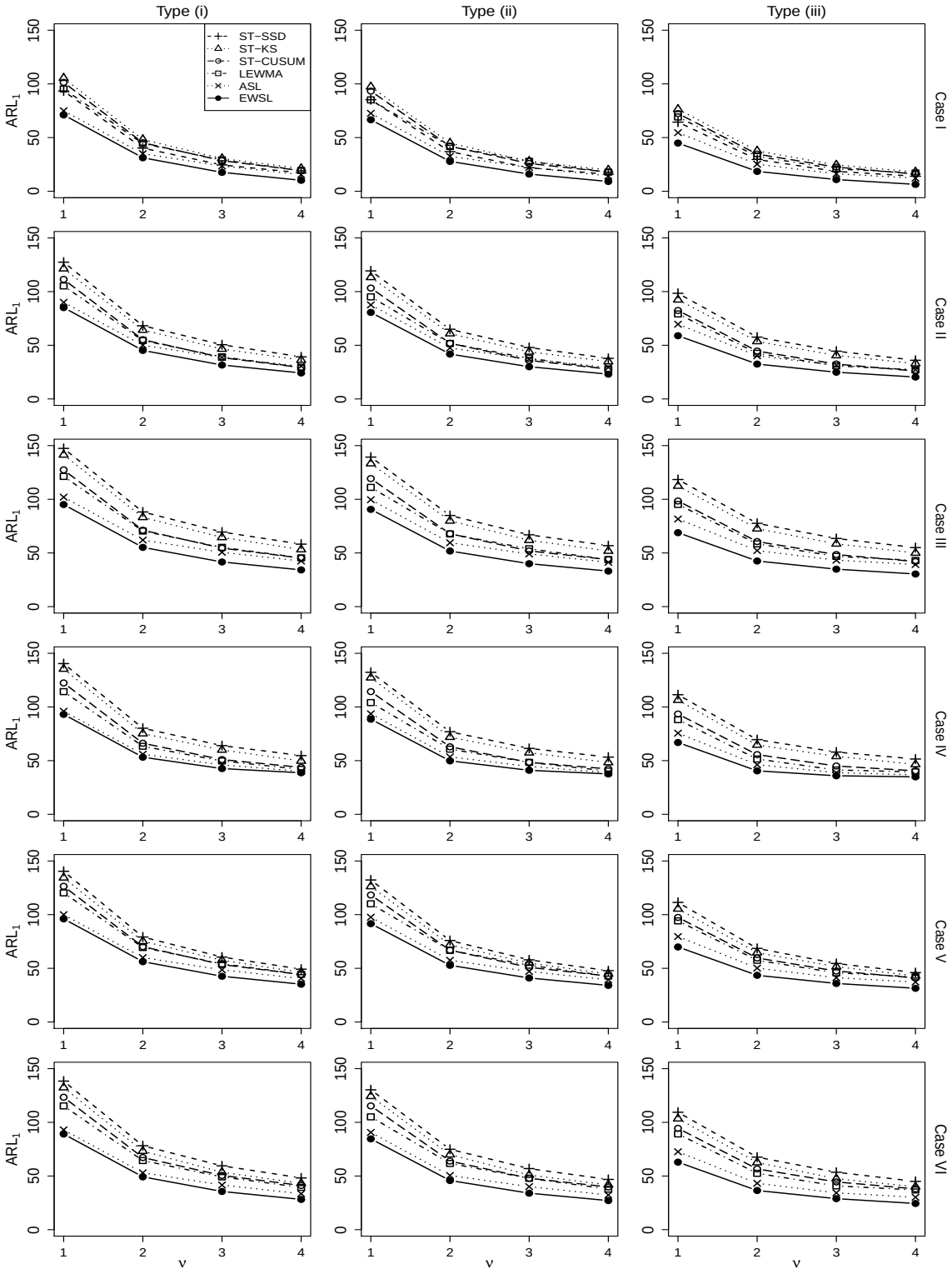

Figure 4: Calculated optimal ARL_1 values of the six charts for detecting the the shifts of Types (i)-(iii) in Cases I-VI when the nominal ARL_0 is 200, $(n, m) = (300, 64)$ and $d = 0.2$.

Table 1: Calculated ARL_0 values and their standard errors (in parentheses) of the six control charts in different cases when the nominal ARL_0 value is 200. In the table, each λ value is used for all three charts LEWMA, ASL and EWSL.

Case	ST-SSD	ST-KS	ϕ	ST-CUSUM	λ	LEWMA	ASL	EWSL
I			0.1	191 (10.1)	0.02	195 (10.9)	204 (10.8)	198 (10.5)
	197 (10.6)	193 (10.6)	0.3	191 (10.3)	0.05	194 (10.8)	198 (10.6)	199 (10.5)
			0.5	195 (10.4)	0.10	196 (10.8)	200 (10.7)	200 (10.6)
II			0.1	186 (10.0)	0.02	186 (10.4)	194 (10.3)	195 (10.2)
	171 (9.7)	182 (10.2)	0.3	188 (10.1)	0.05	195 (10.5)	197 (10.5)	197 (10.7)
			0.5	192 (10.2)	0.10	195 (10.9)	198 (10.8)	199 (10.6)
III			0.1	185 (10.0)	0.02	186 (10.2)	191 (10.2)	190 (10.0)
	150 (8.1)	169 (9.6)	0.3	186 (10.1)	0.05	189 (10.4)	192 (10.5)	193 (10.3)
			0.5	190 (10.3)	0.10	191 (10.8)	196 (10.7)	195 (10.4)
IV			0.1	187 (10.8)	0.02	190 (11.3)	192 (11.3)	193 (11.1)
	164 (9.7)	182 (9.9)	0.3	189 (11.2)	0.05	193 (11.5)	195 (11.4)	196 (11.2)
			0.5	193 (11.3)	0.10	196 (11.7)	198 (11.5)	198 (11.3)
V			0.1	182 (10.6)	0.02	183 (11.1)	191 (10.8)	192 (10.9)
	144 (7.8)	165 (7.6)	0.3	183 (11.0)	0.05	188 (11.2)	194 (11.1)	193 (11.0)
			0.5	188 (11.2)	0.10	191 (11.4)	195 (11.2)	195 (11.1)
VI			0.1	177 (10.4)	0.02	179 (11.2)	190 (10.7)	191 (10.6)
	138 (7.6)	156 (8.4)	0.3	181 (10.7)	0.05	183 (11.3)	192 (11.2)	192 (11.0)
			0.5	184 (10.7)	0.10	189 (11.3)	193 (11.3)	194 (11.1)

than LEWMA in this example, which confirms that the spatial information used in ASL and EWSL is helpful for spatio-temporal process monitoring. (iv) The proposed chart EWSL has the best performance in all cases considered. Thus, compared to ASL in which the EWMA procedure in the time domain and the adaptive spatial LASSO procedure in the space domain are separated, EWSL with the two procedures mixed together in the space-time domain would be more effective in detecting spatially clustered process shifts that start in small collected regions.

In the previous example, the parameter d that controls the size of the shift region Δ is fixed at 0.2. Next, we consider cases when $d = 0.1$, 0.3, 0.4 or 0.5, and other setups are the same as those in Figure 4. The optimal ARL_1 values of the six charts for detecting the shifts of Type (iii) in Case II are shown in Figure 5. From the figure, we can see that EWSL still performs the best among the six charts in all cases considered. Results for detecting the other two types of shifts have similar patterns, and thus are omitted here.

In all above examples, it is assumed that the process mean shift occurs in a single connected

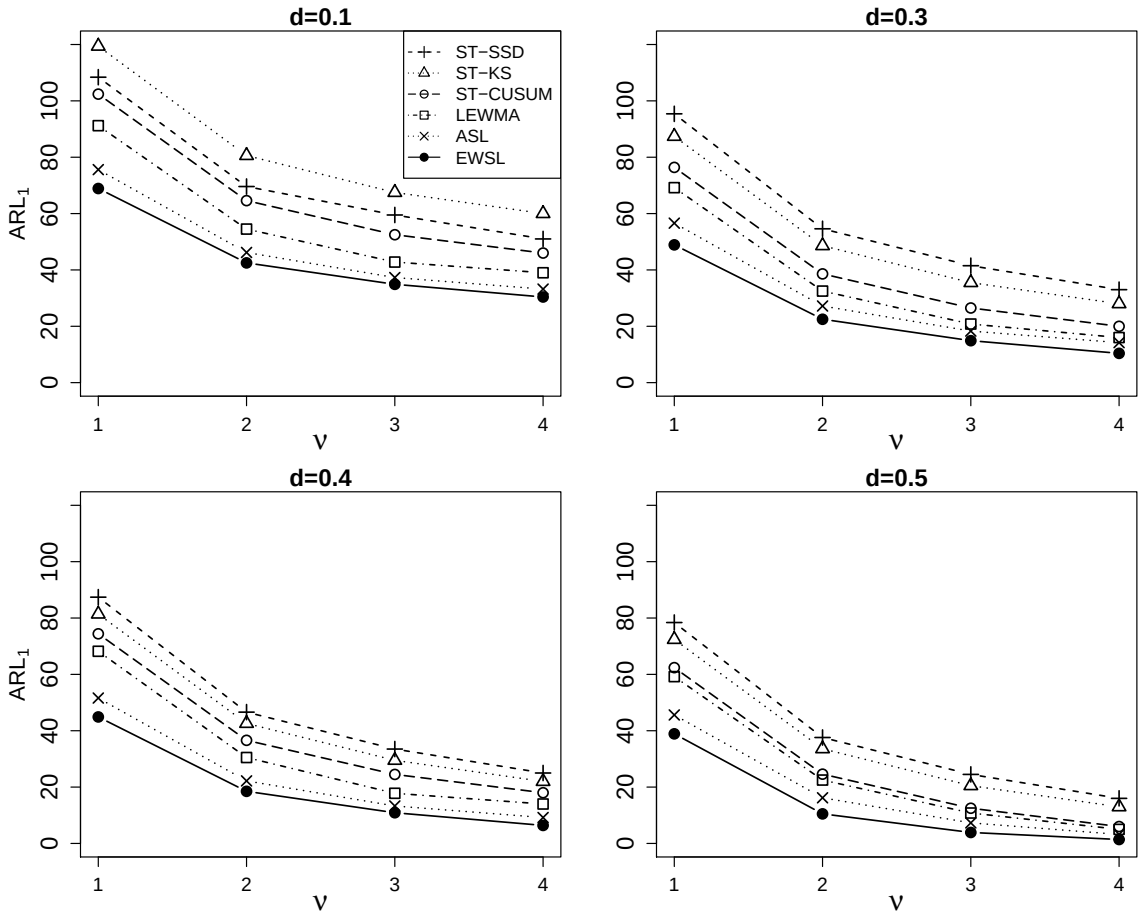

Figure 5: Calculated optimal ARL_1 values of the six charts for detecting the shifts of Type (iii) in Case II when the nominal ARL_0 is 200, $(n, m) = (300, 64)$, and $d = 0.1$, 0.3, 0.4 or 0.5.

region Δ. In the next example, we consider two scenarios in which the process mean shift occurs in multiple regions. In scenario 1, the mean shift occurs in $\Delta_1 = \{\boldsymbol{s}, d_E(\boldsymbol{s}, \boldsymbol{s}_0) \le 0.1,$ or $d_E(\boldsymbol{s}, \boldsymbol{s}_{01}) \le 0.1$ or $d_E(\boldsymbol{s}, \boldsymbol{s}_{02}) \le 0.1\}$ which consists of three connected regions, where $\boldsymbol{s}_{01} = (0.25, 0.25)^T$ and $\boldsymbol{s}_{02} = (0.75, 0.75)^T$. In scenario 2, the mean shift occurs in Δ_2 which consists of 10 randomly selected sites from $\{\boldsymbol{s}_j, j = 1, \ldots, m\}$. Both Δ_1 and Δ_2 are shown in Figure S.1 of the supplementary file. In the two scenarios, process observations are generated as before in Case II for detecting Type (iii) shifts, except that Δ should be replaced by Δ_1 and Δ_2, respectively. Then, the optimal ARL_1 values of the six charts in the same setups as those in Figure 4 are shown in Figure S.2 of the supplementary file. From the figure, it can be seen that i) the proposed chart EWSL performs the best among all six charts in scenario 1, ii) it performs better than the charts ST-SSD, ST-KS, ST-CUSUM and ASL in scenario 2, and iii) it performs slightly worse than the chart LEWMA in scenario 2. The

third conclusion is expected because the chart LEWMA is based on the adaptive LASSO procedure, which is good in detecting shifts at scattered spatial locations, while the proposed chart is based on the spatial LASSO procedure, which is good in detecting shifts in small clustered regions but would lose some effectiveness in detecting shifts at scattered spatial locations.

4 Case Study: Monitoring PM2.5 Concentrations in China

In this section, we demonstrate the use of the proposed chart EWSL in a real application by considering online monitoring of the PM2.5 concentrations in China. Because of its fast expansion in economic activities in the past decades, China faces a challenging air polution problem. PM2.5 particles in the air, with diameters less than 2.5 micrometers, can carry toxic substances that are harmful to human health. Numerous studies have shown that long-term exposure to high PM2.5 concentrations is associated with the incidence of many diseases, including cancers, respiratory and cardiovascular diseases (e.g., Berger et al. 2018, Fu et al. 2019, Sahu et al. 2019). It is therefore extremely important to establish an air polution data collection and monitoring system to collect air polution data and monitor the collected data carefully. To this end, the air polution data collection and monitoring system has been developed by the China National Environmental Monitoring Centre (CNEMC), and a dataset containing the daily PM2.5 concentration levels at 183 major cities in China from 2014 to 2016 can be downloaded from its official web page (http://www.cnemc.cn/en/). This dataset is used here to demonstrate the proposed chart EWSL.

To show the data, the daily PM2.5 concentrations at four representative cities in China, including Beijing, Harbin, Chengdu and Guangzhou, are presented in Figure 6. From the figure, it can be seen that the data in years 2014 and 2015 are quite stable. Thus, observed data in these two years are used as the IC dataset for designing the chart EWSL. To this end, the IC dataset splits into two parts: the data in the year 2014 are used for estimating the IC model (1), and the data in the year 2015 are used for determining the control limit L using the block bootstrap procedure discussed in Subsection 2.3. Then, online monitoring of the PM2.5 concentration levels starts from the beginning of the year 2016.

Besides the chart EWSL, the five alternative charts ST-SSD, ST-KS, ST-CUSUM, LEWMA and ASL discussed in Section 3 will also be considered, for a comparison purpose. For all charts, the nominal ARL_0 value is set to be 200. In the charts ASL and EWSL, λ is chosen to be 0.1. The

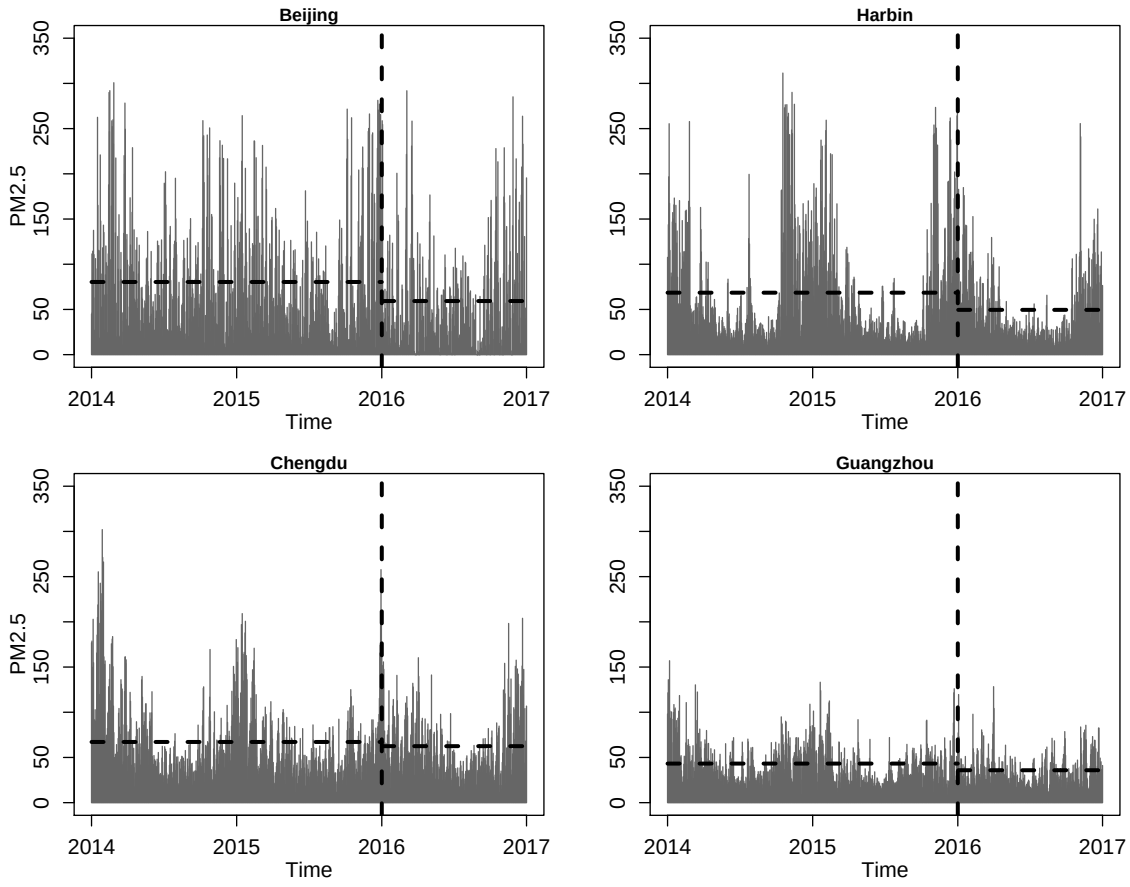

Figure 6: Observed PM2.5 concentrations at four representative cities in China from the year 2014 to the year 2016. In each plot. the vertical dashed line separates the IC data and the data for online process monitoring, and the horizontal dashed lines denote the sample means of the two data subsets.

weighting parameter in the LEWMA chart and the allowance constant ϕ in the ST-CUSUM chart are also chosen to be 0.1. The control limits of the charts ST-CUSUM, LEWMA, ASL and EWSL are determined by the block bootstrap procedure discussed in Subsection 2.3 with the bootstrap sample size to be $B = 1,000$ and the block size to be $\tau = 5$. For the ST-SSD chart, its control limit is determined by the simulation-based procedure, as discussed in Yan et al. (2018). The threshold value c in the ST-KS method is chosen by the parametric bootstrap procedure, as in Zhao et al. (2011). The charting statistics of the six methods are shown in Figure 7. It can be seen from the figure that (i) ST-SSD gives a signal on 04/11/2016 which is two weeks earlier than the signal of ST-KS, (ii) the first signals of ST-CUSUM and LEWMA are on 04/07/2016 and 03/25/2016, respectively, which are earlier than the signal times of ST-SSD and ST-KS, and (iii) the first signal

of EWSL is on 03/17/2016 which is two days earlier than the first signal of ASL. This example confirms the benefit to use the proposed EWSL chart for STPM.

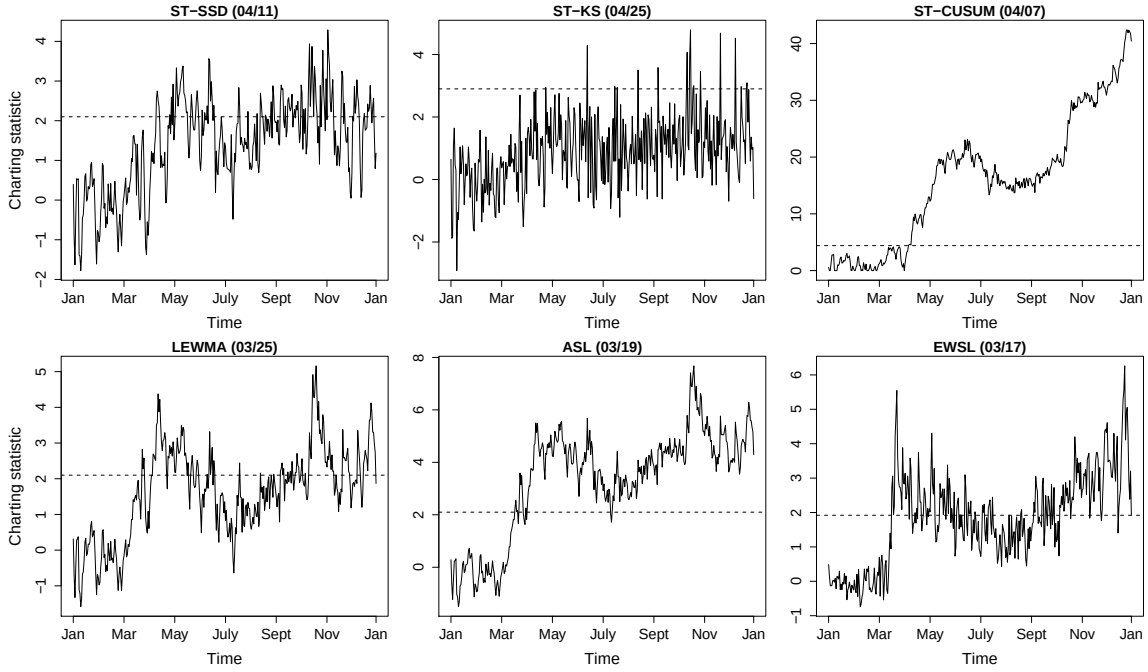

Figure 7: Six control charts ST-SSD, ST-KS, ST-CUSUM, LEWMA, ASL and EWSL. The dashed horizontal line in each plot denotes the control limit or the threshold value of the related chart. The date besides each method label is the first signal time of the method.

To verify whether the detected shift by EWSL is real, we first present in Figure 8 the residual maps at the capital cities of all provinces in China on March 17 in years 2014, 2015 and 2016. Here, the residual at $(t, \boldsymbol{s})$ is defined to be $\widehat{\varepsilon}(t, \boldsymbol{s}) = z(t, \boldsymbol{s}) - \widehat{\mu}(t, \boldsymbol{s})$, where $z(t, \boldsymbol{s})$ is the observation at $(t, \boldsymbol{s})$ and $\widehat{\mu}(t, \boldsymbol{s})$ is the estimated IC mean by the LLKS procedure (A.1) given in Appendix from the IC dataset. From the residual maps, it can be seen that the map for the year 2016 is generally darker compared to the other two maps, implying that the residual values in the year 2016 are generally smaller compared to those in the years 2014 and 2015. Thus, these residual maps confirm a downward mean shift that occurs on or before 03/17/2016. Next, we present the mean estimates of the decorrelated and standardized data (i.e., $\widehat{\mu}_{\widehat{\varepsilon}}(t, \boldsymbol{s})$) obtained by the penalized EWKS procedure (4) at the capital cities of all provinces in China on March 17 of the years 2014, 2015 and 2016 in Figure 9. From the figure, it can be seen that (i) the estimated mean values are much smaller in the year 2016, compared to those in the years 2014 and 2015, and (ii) the cities with the negative estimated values in the year 2016 are indeed spatially clustered. According to

Xiao et al. (2020), the Chinese government started to implement the Air Pollution Prevention and Control Action Plan in 2013 to try to reduce air pollution in China. Our analysis confirms that this policy is quite effective, and the PM2.5 concentrations in the country have a downward mean shift occurred on or before 03/17/2016.

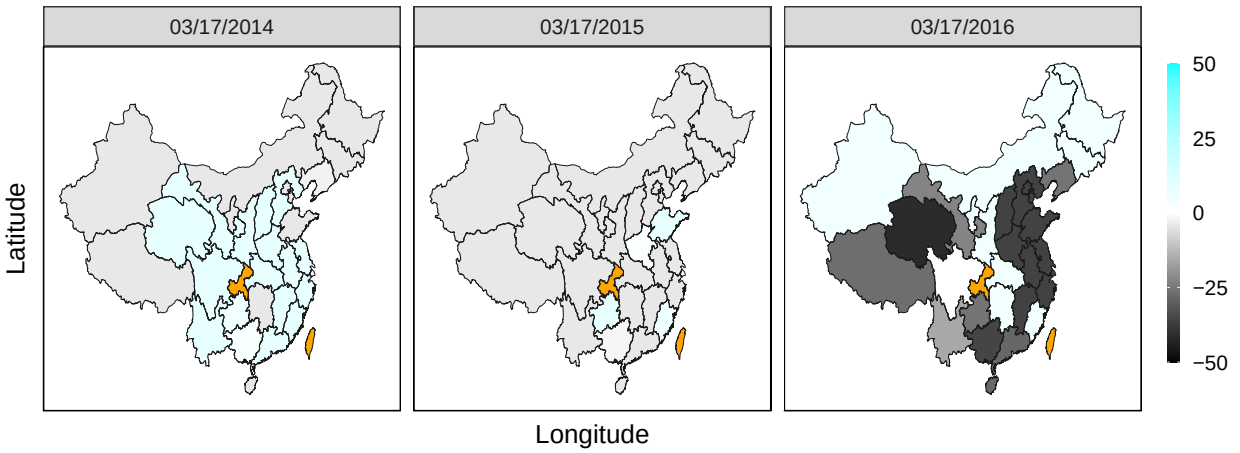

Figure 8: Residual maps at the capital cities of all provinces in China on March 17 of the years 2014, 2015 and 2016. In the maps, regions with the orange color denote those without any reported PM2.5 concentration data.

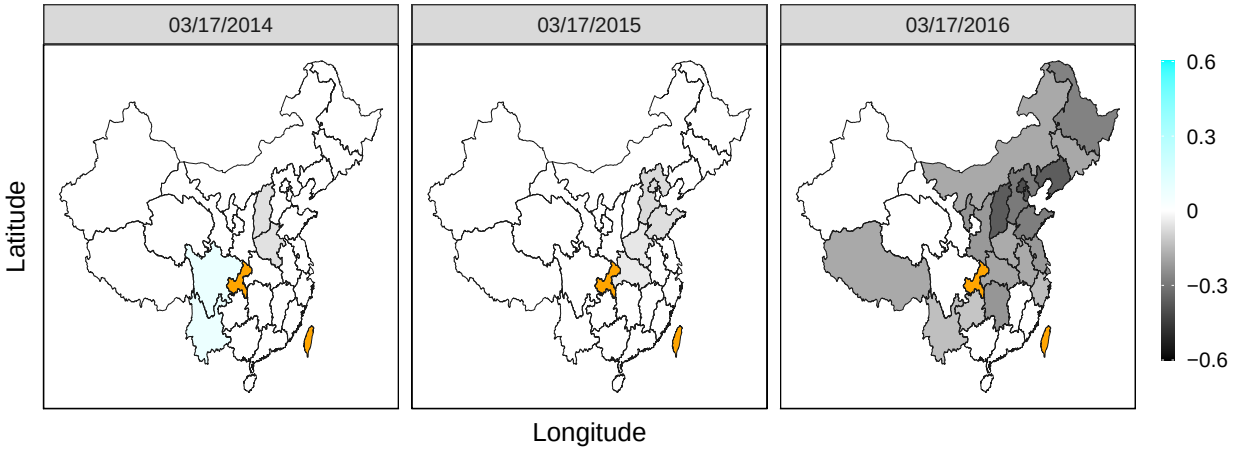

Figure 9: Maps of the mean estimates of the decorrelated and standardized data $\{\widehat{\mu}_{\widehat{e}}(t_i, \boldsymbol{s}_j)\}$ obtained by the penalized EWKS procedure (4) at the capital cities of all provinces in China on March 17 of the years 2014, 2015 and 2016. In the maps, regions with the orange color denote those without any reported PM2.5 concentration data.

5 Concluding Remarks

In this article, we have presented a new online monitoring procedure for STPM, which combines the ideas of exponentially weighted smoothing in the time domain and the spatial LASSO variable selection in the space domain. It has been shown that this method is effective for detecting process shifts that start in small clustered regions. This method can well accommodate space/time-varying mean and covariance structures, spatio-temporal data correlation, and nonparametric data distribution. Thus, it can provide a powerful tool for solving the STPM problem. However, there are still some issues that need to be addressed in the future research. For instance, after a signal of shift is given by the EWSL chart, it is important to figure out when and where the detected shift occurs. To address this issue, a reliable post-signal diagnostic technique should be developed. Although the mean estimate $\widehat{\mu}_{\widehat{e}}(t, \boldsymbol{s})$ by the penalized EWKS procedure (4) can provide some useful information for localizing the detected shift, development of such a diagnostic tool is not straightforward and requires some future research effort. Also, the proposed chart EWLS is based on the L_1 adaptive spatial LASSO penalty. In the statistical literature, there are some alternative penalized approaches for variable selection, such as the elastic net (Zou and Hastie 2005) and the smoothly clipped absolute deviation procedure (Fan and Li 2001). It should be interesting to study these alternative variable selection procedures in the context of STPM and then select a more effective variable selection approach. Furthermore, performance of the proposed chart EWSL depends on the weighting parameter λ. Since this parameter is similar to the weighting parameter in an EWMA chart, it is expected that the optimal selection of λ depends on the shift size which is often unknown in practice. To overcome the difficulty to select a proper value for λ, one strategy is to combine a sequence of potential values of λ to achieve a robust OC performance (e.g., Qiu et al. 2018). Another strategy is to consider an adaptive EWMA chart in which the shift size is estimated sequentially and the value of λ is adjusted accordingly. These strategies to choose λ should also be studied in the future research.

Acknowledgments: The authors thank the editor, the associate editor, and two referees for many insightful comments and suggestions, which improved the quality of the paper greatly.

Appendix: Estimation of $\mu(t, \boldsymbol{s})$, $\sigma^2(t, \boldsymbol{s})$ and $V(t, t'; \boldsymbol{s}, \boldsymbol{s}')$

To estimate the mean function $\mu(t, \boldsymbol{s})$ in Model (1), we can consider using the following local

linear kernel smoothing (LLKS) procedure, as suggested by Yang and Qiu (2018):

$$\begin{aligned}\underset{\boldsymbol{\theta}\in\mathbb{R}^4}{\arg\min}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[z(t_i,\boldsymbol{s}_{ij})-\ \theta_\mu-\theta_t(t_i-t)-\theta_x(s_{x,ij}-s_x)-\theta_y(s_{y,ij}-s_y)\right]^2\\ \times K_t\left((t_i-t)/h_t\right)K_s\left(d_E(\boldsymbol{s}_{ij},\boldsymbol{s})/h_s\right),\end{aligned} \tag{A.1}$$

where $\boldsymbol{\theta}=(\theta_\mu,\theta_t,\theta_x,\theta_y)^T$, $\boldsymbol{s}=(s_x,s_y)^T$, $h_t,h_s>0$ are two bandwidths, $K_t(\cdot)$ and $K_s(\cdot)$ are two kernel functions, and $d_E(\boldsymbol{s}_{ij},\boldsymbol{s})$ is the Euclidean distance between the two spatial locations $\boldsymbol{s}_{ij}$ and $\boldsymbol{s}$. Then, the solution of (A.1) to θ_μ is defined to be the LLKS estimate of $\mu(t,\boldsymbol{s})$, which has the expression

$$\widehat{\mu}(t,\boldsymbol{s})=\boldsymbol{\zeta}_1^T\left(\mathbf{G}^T\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{W}\mathbf{Z}, \tag{A.2}$$

where $\boldsymbol{\zeta}_1=(1,0,0,0)^T$, $\mathbf{G}=(\mathbf{G}_{11},\ldots,\mathbf{G}_{nm_n})^T$, $\mathbf{G}_{ij}=(1,(t_i-t),(s_{x,ij}-s_x),(s_{y,ij}-s_y))^T$, $\mathbf{W}=\text{diag}\{w_{11},\ldots,w_{nm_n}\}$, $w_{ij}=K_t((t_i-t)/h_t)K_s\left(d_E(\boldsymbol{s}_{ij},\boldsymbol{s})/h_s\right)$ and $\mathbf{Z}=(z(t_1,\boldsymbol{s}_{11}),\ldots,z(t_n,\boldsymbol{s}_{nm_n}))^T$, for $j=1,\ldots,m_i$ and $i=1,\ldots,n$. From (A.2), it can be seen that $\widehat{\mu}(t,\boldsymbol{s})$ is actually a weighted average of all observations in a spatio-temporal neighborhood of $(t,\boldsymbol{s})$, with the weights determined by the two kernel functions and with the neighborhood size controlled by the two bandwidths.

After the mean function $\mu(t,\boldsymbol{s})$ is estimated by (A.2), the residuals $\{\widehat{\varepsilon}(t_i,\boldsymbol{s}_{ij})=z(t_i,\boldsymbol{s}_{ij})-\widehat{\mu}(t_i,\boldsymbol{s}_{ij}),j=1,\ldots,m_i,i=1,\ldots,n\}$ can be computed. These residuals can be used for estimating the covariance and variance functions $V(t,t';\boldsymbol{s},\boldsymbol{s}')$ and $\sigma^2(t,\boldsymbol{s})$ by the weighted moment estimation procedure, as discussed in Yang and Qiu (2019). More specifically, the variance function $\sigma^2(t,\boldsymbol{s})=V(t,t;\boldsymbol{s},\boldsymbol{s})$ can be estimated by

$$\widehat{\sigma}^2(t,\boldsymbol{s})=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widehat{\varepsilon}^2(t_i,\boldsymbol{s}_{ij})w_\sigma(i,j;t,\boldsymbol{s})}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}w_\sigma(i,j;t,\boldsymbol{s})},\quad\text{for }(t,\boldsymbol{s})\in[0,T]\times\Omega, \tag{A.3}$$

where $w_\sigma(i,j;t,\boldsymbol{s})=K_t\left((t_i-t)/g_t\right)K_s\left(d_E(\boldsymbol{s}_{ij},\boldsymbol{s})/g_s\right)$, for $1\le j\le m_i$ and $1\le i\le n$, and $g_t,g_s>0$ are two bandwidths that can be different from the bandwidths h_t and h_s used for estimating $\mu(t,\boldsymbol{s})$. The covariance function $V(t,t';\boldsymbol{s},\boldsymbol{s}')$ can be estimated by

$$\widehat{V}(t,t';\boldsymbol{s},\boldsymbol{s}')=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\widehat{\varepsilon}(t_i,\boldsymbol{s}_{ij})\widehat{\varepsilon}(t_k,\boldsymbol{s}_{kl})w_v(i,j,k,l;t,t',\boldsymbol{s},\boldsymbol{s}')}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}w_v(i,j,k,l;t,t',\boldsymbol{s},\boldsymbol{s}')}, \tag{A.4}$$

where $(t,\boldsymbol{s})\neq(t',\boldsymbol{s}')$, and $w_v(i,j,k,l;t,t',\boldsymbol{s},\boldsymbol{s}')=w_\sigma(i,j;t,\boldsymbol{s})w_\sigma(k,l;t',\boldsymbol{s}')$, for all i,j,k,l. Note that the estimates $\widehat{\sigma}^2(t,\boldsymbol{s})$ and $\widehat{V}(t,t';\boldsymbol{s},\boldsymbol{s}')$ obtained from (A.3) and (A.4) may not be positive semidefinite, and thus may not be legitimate variance and covariance functions. To overcome this limitation, a necessary modification is needed to make them positive semidefinite. In this paper, the projection-based modification procedure discussed in Yang and Qiu (2019) is used.

To use the estimation procedures (A.1)-(A.4), the two kernel functions $K_t(\cdot)$ and $K_s(\cdot)$ and the four bandwidths (h_t, h_s) and (g_t, g_s) need to be chosen in advance. In this paper, both kernel functions are chosen to be the Epanechnikov function $K_e(u) = 0.75(1-u^2)I(|u| \leq 1)$, because of its good theoretical properties (Epanechnikov 1969). The bandwidths (h_t, h_s) are chosen by the modified cross-validation criterion discussed in Section 2.3 of Yang and Qiu (2018), and the bandwidths (g_t, g_s) are chosen by minimizing the mean squared predication error as discussed in Yang and Qiu (2019).

References

Apley, D.W., and Tsung, F. (2002), "The autoregressive T^2 chart for monitoring univariate autocorrelated processes," *Journal of Quality Technology*, **34**, 80–96.

Berger, K., Malig, B.J., Hasheminassab, S., Pearson, D.L., Sioutas, C., Ostro, B., and Basu, R. (2018), "Associations of source-apportioned fine particles with cause-specific mortality in California," *Epidemiology*, **29**, 639–648.

Epanechnikov, V.A. (1969), "Non-parametric estimation of a multivariate probability density," *Theory of Probability and its Applications*, **14**, 153–158.

Fan, J., and Li, R. (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, **96**, 1348–1360.

Fu, P., Guo, X., Cheung, F.M.H., and Yung, K.K.L. (2019), "The association between PM2.5 exposure and neurological disorders: a systematic review and meta-analysis," *Science of the Total Environment*, **655**, 1240–1248.

Huang, H.C., Hsu, N.J., Theobald, D.M., and Breidt, F.J. (2010), "Spatial lasso with applications to GIS model selection," *Journal of Computational and Graphical Statistics*, **19**, 963–998.

Knox, E., and Bartlett, M. (1964), "The detection of space-time interactions," *Journal of the Royal Statistical Society (Series C)*, **13**, 25–30.

Kulldorff, M. (1997), "A spatial scan statistic," *Communications in Statistics–Theory and Methods*, **26**, 1481–1496.

Kulldorff, M., and Hjalmars, U. (1999), "The Knox method and other tests for space-time interactions," *Biometrics*, **55**, 544–552.

Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, New York: Wiley.

Qiu, P. (2014), *Introduction to Statistical Process Control*, Boca Raton, FL: CRC Press.

Qiu, P., Li, W., and Li, J. (2020), "A new process control chart for monitoring short-range serially correlated data," *Technometrics*, **62**, 71–83.

Qiu, P., Zi, X., and Zou, C. (2018), "Nonparametric dynamic curve monitoring," *Technometrics*, **60**, 386–397.

Qiu, P., Zou, C., and Wang, Z. (2010), "Nonparametric profile monitoring by mixed effects modeling," *Technometrics*, **52**, 265–277.

Reynolds, M.R., Amin, R.W, and Arnold, J.C. (1990), "CUSUM charts with variable sampling intervals," *Technometrics*, **32**, 371–384.

Sahu, S.K., Zhang, H., Guo, H., Hu, J., Ying, Q., and Kota, S.H. (2019), "Health risk associated with potential source regions of pm 2.5 in indian cities," *Air Quality Atmosphere & Health*, **12**, 327–340.

Samarov, D.V., Hwang, J., and Litorja, M. (2015), "The spatial LASSO with applications to unmixing hyperspectral biomedical images," *Technometrics*, **57**, 503–513.

Schwarz, G. (1978),"Estimating the dimension of a model," *The Annals of Statistics*, **6**, 461–464.

Takahashi, K., Kulldorff, M., Tango, T., and Yih, K. (2008), "A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring," *International Journal of Health Geographics*, **7**, article 14.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the LASSO," *Journal of the Royal Statistical Society (Series B)*, **58**, 267–288.

Tibshirani, R., and Taylor, J. (2011), "The solution path of the generalized LASSO," *The Annals of Statistics*, **39**, 1335–1371.

Wang, K., and Jiang, W. (2009), "High-dimensional process monitoring and fault isolation via variable selection," *Journal of Quality Technology*, **41**, 247–258.

Wang, A., Xian, X., Tsung, F., and Liu, K. (2018), "A spatial-adaptive sampling procedure for online monitoring of big data streams," *Journal of Quality Technology*, **50**, 329–343.

Woodall, W.H., Marshall, J.B., Joner, M.D., Jr, Fraker, S.E., and Abdel-Salam, A.S.G. (2008), "On the use and evaluation of prospective scan methods for health-related surveillance," *Journal of the Royal Statistical Society (Series A)*, **171**, 223–237.

Xian, X., Wang, A., and Liu, K. (2018), "A nonparametric adaptive sampling strategy for online monitoring of big data streams," *Technometrics*, **60**, 14–25.

Xiao, Q., Geng, G., Liang, F., Wang, X., Lv, Z., Lei, Y., Huang, X., Zhang, Q., Liu, Y., and He, K. (2020), "Changes in spatial patterns of PM2.5 pollution in China 2000–2018: Impact of clean air policies," *Environmental International*, **141**, 105776.

Yan, H., Paynabar, K., and Shi, J. (2018), "Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition," *Technometrics*, **60**, 181–197.

Yang, K., and Qiu, P. (2018), "Spatio-temporal incidence rate data analysis by nonparametric regression," *Statistics in Medicine*, **37**, 2094–2107.

Yang, K., and Qiu, P. (2019), "Nonparametric estimation of the spatio-temporal covariance structure," *Statistics in Medicine*, **38**, 4555–4565.

Yang, K., and Qiu, P. (2020), "Online sequential monitoring of spatio-temporal disease incidence rates," *IISE Transactions*, **52**, 1218–1233.

Yang, Y. (2005), "Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation," *Biometrika*, **92**, 937–950.

Zhao, Y., Zeng, D., Herring, A.H., Ising, A., Waller, A., Richardson, D., and Kosorok, M.R. (2011), "Detecting disease outbreaks using local spatiotemporal methods," *Biometrics*, **67**, 1508–1517.

Zou, C., Ning, X., and Tsung, F. (2012), "LASSO-based multivariate linear profile monitoring," *Annals of Operations Research*, **192**, 3–19.

Zou, C., and Qiu, P. (2009), "Multivariate statistical process control using LASSO," *Journal of the American Statistical Association*, **104**, 1586–1596.

Zou, H. (2006), "The adaptive LASSO and its oracle properties," *Journal of the American Statistical Association*, **101**, 1418–1429.

Zou, H., and Hastie, T. (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society (Series B)*, **67**, 301–320.