Online Sequential Monitoring Of Spatio-Temporal Disease Incidence Rates

Kai Yang and Peihua Qiu
Department of Biostatistics
University of Florida
Gainesville, FL 32611

Abstract

Online sequential monitoring of the incidence rates of chronic or infectious diseases is critically important for public health. Governments have invested a great amount of money in building global, national and regional disease reporting and surveillance systems. In these systems, conventional control charts, such as the cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) charts, are usually included for disease surveillance purposes. However, these charts require many assumptions on the observed data, including the ones that the observed data should be independent at different places and/or times, and they should follow a parametric distribution when no disease outbreaks are present. These assumptions are rarely valid in practice, making the results from the conventional control charts unreliable. Motivated by an application to monitor the Florida influenza-like illness data, we develop a new sequential monitoring approach in this paper, which can accommodate the dynamic nature of the observed disease incidence rates (i.e., the distribution of the observed disease incidence rates can change over time due to seasonality and other reasons), spatio-temporal data correlation, and arbitrary data distribution. It is shown that the new method is more reliable to use in practice than the commonly used conventional charts for sequential monitoring of disease incidence rates. Because of its generality, the proposed method should be useful for many other applications as well, including spatio-temporal monitoring of the air quality in a region or the sea-level pressure data collected in a region of an ocean.

Key Words: Change detection; Correlation; Disease surveillance; Dynamic systems; Early detection; Nonparametric methods; Process control; Sequential monitoring.

1 Introduction

A disease outbreak refers to the occurrence of disease cases in excess of what would normally be expected. In recent years, we experienced the outbreaks of Zika, Ebola, SARS, H1N5, H7N9,

MERS-CoV, chikungunya, and many other damaging infectious diseases. Our society is under a constant threat of bioterrorist attacks and pandemic influenza. It is therefore important to effectively monitor the occurrence of infectious diseases constantly and detect their outbreaks as promptly as possible. Early detection of infectious disease outbreaks can help governments and individuals to take appropriate disease control and prevention measures in a timely manner so that the disease epidemic can be controlled at an early stage and thus its damage can be minimized. This paper aims to develop a new and effective statistical method for sequential monitoring of infectious disease incidence rates and for early detection of their outbreaks. Because of its generality, our proposed method should be useful for other spatio-temporal monitoring problems as well, including the air quality surveillance in environmental research and the sea-level pressure monitoring in oceanography.

Because of the importance to early detect the infectious disease outbreaks, some global, national and regional disease reporting systems have been established to collect/provide data about certain measurements of some important diseases. Commonly used measures of disease frequency include prevalence and incidence (Noordzij et al. 2010). The prevalence reflects the total number of existing disease cases, while the incidence refers to the number of newly diagnosed cases of a disease. Compared to the prevalence, the incidence is more useful in understanding disease etiology and providing guiding principles for targeting interventions. Thus, it is preferred in many disease surveillance systems. Besides, two main types of disease incidence data are available in practice, including (i) number of newly confirmed cases and (ii) incidence rate of a disease. Even though the number of disease cases can provide some useful information about the current burden of a disease, it suffers from the limitation that it cannot distinguish a large population with a low disease rate from a small population with a high disease rate. To overcome this limitation, we can consider the disease incidence rate, defined as the total number of new cases in a region divided by the population of the region in a specific observation time period. This paper is motivated by the incidence rate data of the influenza-like illness (ILI), which is a respiratory infection caused by a variety of influenza viruses. A suspect ILI case is defined as the severe respiratory illness with fever ($> 100^{\circ}$F), cough, sore throat, and difficulty in breathing. It is estimated that 15-40% of the population develop illness from influenza each year in the US. About 36,000 people per year die from influenza infection, and about 114,000 people per year have to be admitted to hospital due to influenza infection (Fiore et al. 2010). A traditional method to estimate the incidence rate of ILI

is to carry out repeated seroprevalence surveys. But, such surveys are resource-intensive and slow. Thus, they are unfeasible for early detection of disease outbreaks. To overcome that difficulty, the Florida Department of Health (FDOH) has built an Electronic Surveillance System for the Early Notification of Community-based Epidemics at Florida (ESSENCE-FL) recently, which is a syndromic surveillance system for collecting *near real-time* pre-diagnostic data from participating hospitals and urgent care centers in Florida. Currently, the system collects data from acute care visits to 229 emergency departments and 35 urgent care centers distributed in all counties of Florida, and the collected data are updated once a day. Figure 1 presents the observed incidence rates of ILI for all 67 counties of Florida on 06/01/2012 (a summer time) and 12/01/2012 (a winter time). From the two plots, we can see that the ILI incidence rates in winter were generally higher than those in summer, and the ILI epidemic in some counties (e.g., the Liberty county in the northwestern Florida) was serious in the winter time.

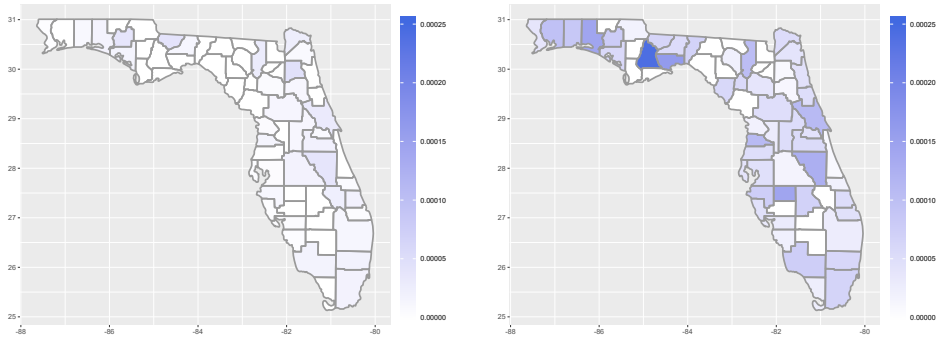

Figure 1: Observed ILI incidence rates in Florida on 06/01/2012 (left) and 12/01/2012 (right). Darker colors denote larger values.

In the surveillance systems like ESSENCE-FL, some conventional process monitoring tools, such as the cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) charts, are usually included for routine disease surveillance (cf., Chen et al. 2010, Kite-Powell et al. 2010). These charts, however, require the assumptions that the observed data are independent and follow a parametric distribution (Qiu 2014). In practice, these assumptions are hardly valid, and it has been well demonstrated in the statistical process control (SPC) literature that results from the conventional control charts would be unreliable or even misleading in cases when their required assumptions are violated (e.g., Capizzi 2015, Chakraborti et al. 2015, Hackl and Ledolter 1991, Qiu and Hawkins 2001). To address this issue, many new SPC methods have been developed recently. For instance, monitoring of correlated data has been discussed in many papers, including Apley

and Lee (2003), Li and Qiu (2019), Psarakis and Papaleonida (2016), and Zhang (1998). Usually, a parametric time series model is used in these methods for describing the data correlation, and then a control chart is applied to the residuals obtained from the estimated time series model. To address the violation of the normality assumption in practice, many nonparametric or distribution-free control charts have been developed in the literature. See, for instance, Bakir and Reynolds (1979), Capizzi and Masarotto (2013), Chakraborti et al. (2001), Chen et al. (2016), Holland and Hawkins (2014), Qiu (2008, 2018), Qiu and Hawkins (2001), and Qiu and Li (2011). To accommodate the dynamic nature of certain processes in practice (i.e., their process distributions would change over time), the so-called dynamic screen system (DySS) has been developed recently (e.g., Qiu and Xiang 2014, 2015, You and Qiu 2019, Zhang et al. 2015). Some versions of the DySS method can accommodate both time-varying process distributions and temporally correlated data (Li and Qiu 2016, 2017). In the statistical/epidemiological literature, there exist some *retrospective* methods for identifying spatial or spatio-temporal disease clusters, including the Knox, local Knox, and scan methods (e.g., Knox and Bartlett 1964, Kulldorff 1997, Marshall et al. 2007). As pointed out by Marshall et al. (2007), Woodall et al. (2008), and Zhou and Lawson (2008), these methods would not be as effective as the CUSUM charts commonly used for process monitoring, because the former cannot be used for *prospective* surveillance of infectious diseases and they could not properly accommodate the spatio-temporal data correlation either. In the literature, there are some discussions about prospective online monitoring of spatial data. For instance, Jiang et al. (2011) suggested a likelihood ratio (LR) based control chart for monitoring spatial data. However, their problem is different from the current problem, in that the underlying in-control (IC) spatial data distribution was assumed to be normal (in order to use the LR framework) and unchanged over time in the former problem, while the current problem does not have these assumed properties. Colosimo et al. (2014), Wang et al. (2014), Zang and Qiu (2018) and some others discussed several different spatial data monitoring problems in which all the IC spatial process distributions were assumed to be unchanged over time, and thus methods in these papers cannot be applied to the current problem either.

Infectious disease data often have complicated structures. For instance, the distribution of disease incidence within a given region and a given time interval often cannot be approximated well by a Poisson, negative binomial, or normal distribution (cf., Zhang et al. 2015), because of the facts that many confounding risk factors could affect the disease incidence in practice,

these risk factors may not be easy to measure, and sometimes it is even difficult for us to notice their existence. Besides, disease incidence often has complex spatio-temporal patterns like spatial clustering, seasonality and day-of-week variation (cf., Zhao et al. 2011), which cannot be described well by a parametric model either. Furthermore, the observed disease incidence data at different time points and different spatial locations are usually correlated: the closer the distance between two observation time points or locations, the stronger the correlation. Such spatio-temporal data correlation is often difficult to describe by a parametric spatial time series model. Therefore, the conventional CUSUM and EWMA charts would not be appropriate to use in the current problem. The more recent control charts described above for monitoring correlated and/or non-Gaussian data or for monitoring dynamic processes are not appropriate here either because of the reasons given below. First, the parametric time series models assumed in the control charts for monitoring correlated data usually cannot be generalized easily for describing the spatial or spatio-temporal data correlation in the current problem. These control charts cannot handle dynamic processes with flexible data correlation structure either. Second, the existing nonparametric control charts are designed mainly for monitoring traditional univariate or multivariate processes whose IC distributions remain unchanged over time. They cannot be used for applications with temporally correlated spatial data and time-varying IC distributions. Third, the existing DySS methods are for monitoring processes with a single or multiple performance variables (cf., Li and Qiu 2016, 2017). They cannot handle the current infectious disease surveillance problem with spatial data.

In this paper, we aim to develop an effective and flexible sequential online monitoring system for infectious disease surveillance and for early detection of disease outbreaks. This method can accommodate the complicated structure of the observed infectious disease data adequately. More specifically, it does not impose any parametric forms on the disease incidence distribution and on the spatio-temporal pattern of the disease incidence trajectory. It can effectively accommodate the spatio-temporal correlation in the observed data and the dynamic nature of the disease incidence rates. *It should be pointed out that most spatio-temporal modeling approaches discussed in the statistical literature are for analyzing spatio-temporal data obtained in a given time interval (e.g., Yang and Qiu 2018, 2019). They are developed for solving the offline modeling problems, in which all observations have been obtained completely at the time when data analysis is performed and no new observations can be added to the existing dataset after data analysis starts. See the related discussion given in Section 2.1 below for a more detailed discussion. Thus, these methods are*

NOT for sequential monitoring of spatio-temporal data (or, the online spatio-temporal monitoring problem), in which new data keep coming over time. As a comparison, the online spatio-temporal monitoring problem is the focus of the current paper.

Our proposed method will be described in detail in Section 2. Its performance will be evaluated numerically in Section 3. The proposed method is applied to the Florida ILI incidence data in Section 4. Several remarks conclude the article in Section 5.

2 Sequential Monitoring Of Disease Incidence Rates

Our proposed method for online monitoring of disease incidence rates is described in this section. It consists of two steps: (i) estimation of a baseline model for describing the regular longitudinal pattern (i.e., seasonality and other data variations over time) of the disease incidence rates when no disease outbreaks are present, and (ii) online monitoring of the observed disease incidence rates and delivery of a signal when a significant shift in the longitudinal pattern of the observed disease incidence rates from the estimated regular longitudinal pattern is detected. These two steps are described in details below.

2.1 Estimation of a baseline model

In the literature, there is much discussion about retrospective modeling of spatio-temporal data obtained within a given spatial region and a given time interval. For instance, some authors suggested estimating the true mean function of the spatio-temporal data using the penalized splines under the generalized additive model or mixed-effects model framework (e.g., Chouldechova and Hastie 2015, Heuvelink and Griffith 2010). Some other existing methods require certain parametric model assumptions (e.g., Diggle 2014). Kafadar (1996) suggested analyzing spatial data using the conventional local constant kernel smoothing procedures. Recent discussions about estimation of the spatio-temporal covariance structure can be found in Shand and Li (2017) and Yang and Qiu (2019). Some overviews on this topic can be found in Banerjee et al. (2004), Cressie and Wikle (2011), Diggle (2014), Gneiting and Guttorp (2010), Gonzalez et al. (2016), Lindström et al. (2015), and Schabenberger and Gotway (2005). Recently, Yang and Qiu (2018) suggested a flexible approach for spatio-temporal data modeling, which was shown more effective than some alternative

approaches in various different cases. This approach is adopted here and briefly described below.

Let Ω and $[0,T]$ be a two-dimensional region and a basic time interval in which the observed incidence rates of a specific disease need to be modeled. In practice, the basic time interval $[0,T]$ usually spans for a whole year from January 1 to December 31, and the disease incidence rate data are collected daily from one year to the next. Denote $\boldsymbol{s} = (s_x, s_y)' \in \Omega$ as a specific location in the region Ω, where s_x and s_y are the x and y coordinates of $\boldsymbol{s}$ (e.g., longitude and latitude). For any $\boldsymbol{s} \in \Omega$ and $t \in [0,T]$, let $N(t, \boldsymbol{s}; dt, O(\boldsymbol{s}))$ be the number of new disease cases observed in a small region $O(\boldsymbol{s})$ around $\boldsymbol{s}$ and in the time interval $[t, t+dt]$, and $M(t, O(\boldsymbol{s}))$ be the population size of the region $O(\boldsymbol{s})$ at time t. Then, $y(t,\boldsymbol{s}) = N(t, \boldsymbol{s}; dt, O(\boldsymbol{s}))/[M(t, O(\boldsymbol{s}))dt]$ is defined to be the disease incidence rate. This definition is commonly used in the epidemiology literature. See, for instance, Last (2001). In cases when no disease outbreaks are present, the observed incidence rates are assumed to follow the model

$$y(t_i, \boldsymbol{s}_{ij}) = \lambda(t_i, \boldsymbol{s}_{ij}) + \varepsilon(t_i, \boldsymbol{s}_{ij}), \qquad \text{for } j = 1, 2, \ldots, m_i,\ i = 1, 2, \ldots, n, \tag{1}$$

where $t_i \in [0,T]$ is the ith time point, $\boldsymbol{s}_{ij} \in \Omega$ is the jth location at time t_i, $\lambda(t_i, \boldsymbol{s}_{ij})$ is the mean of $y(t_i, \boldsymbol{s}_{ij})$, $\varepsilon(t_i, \boldsymbol{s}_{ij})$ is a zero-mean random error, m_i is the number of observation locations at t_i, and n is the number of observation time points. The correlation structure in the observed data can be described by the covariance function

$$\text{Cov}[y(t,\boldsymbol{s}), y(t',\boldsymbol{s}')] = E\left[\varepsilon(t,\boldsymbol{s})\varepsilon(t',\boldsymbol{s}')\right] = V(t,\boldsymbol{s};t',\boldsymbol{s}'), \qquad \text{for } (t,\boldsymbol{s}), (t',\boldsymbol{s}') \in [0,T]\times\Omega. \tag{2}$$

In models (1) and (2), we have not imposed any parametric forms on $\lambda(t,\boldsymbol{s})$, $V(t,\boldsymbol{s};t',\boldsymbol{s}')$, and the error distribution. The only requirements are that (i) $\lambda(t,\boldsymbol{s})$ is a continuous function, and (ii) $V(t,\boldsymbol{s};t',\boldsymbol{s}')$ exists. Thus, these models are flexible.

To estimate the mean function $\lambda(t,\boldsymbol{s})$ in model (1), a natural idea is to use the local kernel estimators. Popular local kernel estimation methods include the Nadaraya-Watson (NW) estimator, the local linear kernel estimator and the local polynomial kernel estimators (cf., Sections 2.3 and 2.4, Qiu 2005). In applications, the most popular local kernel estimator is the local linear kernel estimator, due to its good theoretical and numerical properties, including the so-called automatic boundary carpentry and design adaptation (cf., Section 2.4, Qiu 2005). Compared to the 2nd or higher-order local polynomial kernel estimators, the local linear kernel estimator is more robust to design sparsity (cf., Choi et al. 2000). For these reasons, the following local linear kernel smoothing

procedure is used here:

$$\underset{\boldsymbol{\beta}\in R^4}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \quad \{y(t_i,\boldsymbol{s}_{ij}) - [\beta_0+\beta_1(t_i-t)+\beta_2(s_{x,ij}-s_x)+\beta_3(s_{y,ij}-s_y)]\}^2 \times \quad (3)$$
$$K_1\left(\frac{t_i-t}{h_t}\right) K_2\left(\frac{d_E(\boldsymbol{s}_{ij},\boldsymbol{s})}{h_s}\right),$$

where $\boldsymbol{s}=(s_x,s_y)'$, $\boldsymbol{s}_{ij}=(s_{x,ij},s_{y,ij})'$, K_1 and K_2 are two univariate kernel functions, $h_t, h_s > 0$ are two bandwidths, $d_E(\cdot,\cdot)$ is the Euclidean distance, and $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2,\beta_3)'$. Let $\mathbf{Y} = (y(t_1,\boldsymbol{s}_{11}),\ldots,y(t_1,\boldsymbol{s}_{1m_1}),\ldots,y(t_n,\boldsymbol{s}_{nm_n}))^T$ be the observation vector, and $\mathbf{X}=(\mathbf{X}_{11},\ldots,\mathbf{X}_{1m_1},\ldots,\mathbf{X}_{nm_n})^T$ be the design matrix with $\mathbf{X}_{ij}=(1,t_i-t,(\boldsymbol{s}_{ij}-\boldsymbol{s})^T)^T$, for $j=1,\ldots,m_i$ and $i=1,\ldots,n$. Then, the local linear kernel estimator of $\lambda(t,\boldsymbol{s})$, denoted as $\widehat{\lambda}(t,\boldsymbol{s})$, is defined to be the solution of (3) to β_0, which has the expression

$$\widehat{\lambda}(t,\mathbf{s}) = \mathbf{e}_1'\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}, \quad (4)$$

where $\mathbf{e}_1=(1,0,0,0)'$, $\mathbf{W}=\text{diag}\{w_0(1,1),\ldots,w_0(1,m_1),\ldots,w_0(n,m_n)\}$, $w_0(i,j)=K_1((t_i-t)/h_t)K_2(d_E(\boldsymbol{s}_{ij},\boldsymbol{s})/h_s)$, and $\text{diag}\{\boldsymbol{v}\}$ denotes a diagonal matrix with its diagonal elements given by the vector $\boldsymbol{v}$. Usually, the two kernel functions used in (3) are chosen to have finite supports. Therefore, the estimator $\widehat{\lambda}(t,\mathbf{s})$ in (4) is actually a weighted average of observations in a neighborhood of $(t,\boldsymbol{s})$, where the size of the neighborhood is controlled by the bandwidths and the weights are controlled by the kernel functions. Under some regularity conditions, Yang and Qiu (2018) have shown that $\widehat{\lambda}(t,\boldsymbol{s})$ is statistically consistent.

As in the kernel smoothing literature (cf., Chapter 2, Qiu 2005), the two kernel functions are chosen to be the Epanechnikov kernel function in this paper, due to its good theoretical properties. Namely, $K_1(u)=K_2(u)=[3(1-u^2)/4]I(|u|\le 1)$. Regarding selection of the bandwidths h_t and h_s, it has been well discussed in the literature that the conventional cross-validation (CV) procedure would not perform well, because the CV procedure cannot properly distinguish the data correlation structure from the data mean function in such cases (e.g., Altman 1990, Opsomer et al. 2001). By the modified CV (MCV) procedure suggested in Brabanter et al. (2011) for handling correlated data in the univariate regression setup, h_t and h_s can be chosen by minimizing the following MCV score:

$$\text{MCV}(h_t,h_s) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\left[\widehat{\lambda}_{-(ij)}(t_i,\boldsymbol{s}_{ij}) - y(t_i,\boldsymbol{s}_{ij})\right]^2\right\}, \quad (5)$$

where $\widehat{\lambda}_{-(ij)}(t_i,\boldsymbol{s}_{ij})$ is the leave-one-out estimate of $\lambda(t_i,\boldsymbol{s}_{ij})$ by (3) when the observation $y(t_i,\boldsymbol{s}_{ij})$

is left out in the computation and when the two kernel functions are both chosen to be

$$\widetilde{K}_{\epsilon}(u) = \frac{4}{4-3\epsilon-\epsilon^3} \begin{cases} \frac{3}{4}(1-u^2)I(|u| \leq 1), & \text{if } |u| \geq \epsilon; \\ \frac{3(1-\epsilon^2)}{4\epsilon}|u|, & \text{if } |u| < \epsilon, \end{cases} \tag{6}$$

and $\epsilon \in (0,1)$ is a constant. By using (6), observations around $(t_i, \boldsymbol{s}_{ij})$ are down-weighted when computing $\widehat{\lambda}_{-(ij)}(t_i, \boldsymbol{s}_{ij})$ to reduce the impact of data correlation on bandwidth selection. Regarding ϵ, Brabanter et al. suggested choosing it to be 0.1, based on a large simulation study. This suggestion is adopted here.

So far, we have discussed estimation of the mean function $\lambda(t, \boldsymbol{s})$ in this part. Next, we provide some discussion about the estimation of the variance/covariance function. In the literature, there is some existing research on this topic. See, for instance, Choi et al. (2013), Genton (2007), and Gneiting (2002). However, methods proposed in these papers require some restrictive assumptions, such as the ones that the covariance structure of the spatio-temporal data is stationary and separable in space and time, which are often invalid in practice. Recently, Shand and Li (2017) suggested a method for modelling nonstationary covariance function, with the assumption that the nonstationary process is a projection of a stationary process in a higher-dimension. This method can accommodate nonstationarity in both space and time. But, the estimation relies on correctly specifying a parametric covariance model. To overcome these limitations, Yang and Qiu (2019) suggested a flexible method to model the spatio-temporal covariance structure using the local kernel smoothing approach, which is adopted in this paper and briefly described below.

Let $w_1(i,j,k,l) = K_1\left((t_i - t)/h_t\right) K_1\left((t_k - t')/h_t\right) K_2\left((d_E(\boldsymbol{s}_{ij}, \boldsymbol{s}))/h_s\right) K_2\left((d_E(\boldsymbol{s}_{kl}, \boldsymbol{s}'))/h_s\right)$, and $w_2(i,j) = K_1\left((t_i - t)/h_t\right) K_2\left(d_E(\boldsymbol{s}_{ij}, \boldsymbol{s})/h_s\right)$, for $1 \leq j \leq m_i$, $1 \leq l \leq m_k$, $1 \leq i,k \leq n$, and $(t, \boldsymbol{s}), (t', \boldsymbol{s}') \in [0,T] \times \Omega$. Then, after $\widehat{\lambda}(t_i, \mathbf{s}_{ij})$ is calculated by (4), $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$ can be estimated by the following moment estimate:

$$\widehat{V}(t, \boldsymbol{s}; t', \boldsymbol{s}') = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k} \widehat{\varepsilon}(t_i, \boldsymbol{s}_{ij})\widehat{\varepsilon}(t_k, \boldsymbol{s}_{kl}) w_1(i,j,k,l)}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k} w_1(i,j,k,l)}, \quad \text{when } (t, \boldsymbol{s}) \neq (t', \boldsymbol{s}'), \tag{7}$$

where $\widehat{\varepsilon}(t_i, \boldsymbol{s}_{ij}) = y(t_i, \boldsymbol{s}_{ij}) - \widehat{\lambda}(t_i, \boldsymbol{s}_{ij})$ are the residuals, for $j = 1, 2, ..., m_i$, $i = 1, 2, ..., n$. In cases when $(t, \boldsymbol{s}) = (t', \boldsymbol{s}')$, the variance function $\sigma^2(t, \boldsymbol{s}) = V(t, \boldsymbol{s}; t, \boldsymbol{s})$ can be estimated by

$$\widehat{\sigma}^2(t, \boldsymbol{s}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} \widehat{\varepsilon}^2(t_i, \boldsymbol{s}_{ij}) w_2(i,j)}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_2(i,j)}. \tag{8}$$

2.2 Online monitoring of the spatio-temporal data

As discussed in Section 1, the DySS methods discussed in Li and Qiu (2016, 2017) are for monitoring processes with a single or multiple performance variables. They are not designed for monitoring spatial processes with spatio-temporally correlated data. Also, there are no effective online monitoring schemes in the literature for handling processes with time-varying IC distribution and spatio-temporally correlated data. This paper aims to fill this gap and propose an effective online monitoring scheme for analyzing spatio-temporal disease incidence rates, which is described in details below.

After $\lambda(t, \boldsymbol{s})$ and $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$ are estimated by (4), (7) and (8), their estimators can be used for describing the regular longitudinal pattern of the disease incidence rates in cases when no disease outbreaks are present. Then, they can be used for online prospective monitoring of the disease incidence rates. Assume that the incidence rates to monitor are observed at locations $\{\boldsymbol{s}_{ij}^*, j = 1, 2, \ldots, m_i^*\}$ and times t_i^*, for $i = 1, 2, \ldots$. When no disease outbreaks are present, the observed incidence rates are assumed to follow the model (1) in the sense that $y(t_i^*, \boldsymbol{s}_{ij}^*) = \lambda(t_i^*, \boldsymbol{s}_{ij}^*) + \varepsilon(t_i^*, \boldsymbol{s}_{ij}^*)$, for $j = 1, 2, \ldots, m_i^*$ and $i = 1, 2, \ldots$, and the mean function $\lambda(t, \boldsymbol{s})$ is periodic in time with the period T. Namely, $\lambda(t_i^*, \boldsymbol{s}_{ij}^*) = \lambda(t_i^{**}, \boldsymbol{s}_{ij}^*)$, where $t_i^* = t_i^{**} + \ell T$ for all i, $t_i^{**} \in [0, T]$, and $\ell \geq 1$ is an integer. For instance, if the complete cycle of seasonality in the disease incidence rates is a whole year, which is the case for most infectious diseases, then $T = 1$ (year), $\{t_i^*\}$ are observation times in the first and following years, and $\lambda(t, \boldsymbol{s})$ is periodic in different years to reflect the yearly seasonality. In some applications, the period T may be unknown and thus needs to estimated from the IC data. In the statistical literature, there have been some discussions on estimation of the period T. See, for instance, Sun et al. (2012) and Vogt and Linton (2014). Estimation of T is out of the scope of the current paper, and will be discussed in our future research.

Conventional control charts in the literature for process monitoring are designed for cases when process observations are independent and identically distributed. In the current problem, however, all these assumptions could be violated. Therefore, before online process monitoring, we suggest sequentially decorrelating the observations first to improve the process monitoring efficiency, as discussed in Li and Qiu (2016, 2017). However, the data decorrelation procedures considered in Li and Qiu (2016, 2017) cannot be applied to the current problem directly, because the observed data at a given time point is a scalar number or a vector in the problems consid-

ered there while they are spatially distributed in the region Ω here. So, the data decorrelation problem in the current setup is more challenging, and it is discussed in detail below. For each i, let $\boldsymbol{y}(t_i^*) = (y(t_i^*, \boldsymbol{s}_{i1}^*), y(t_i^*, \boldsymbol{s}_{i2}^*), \ldots, y(t_i^*, \boldsymbol{s}_{im_i^*}^*))'$, $\boldsymbol{\lambda}(t_i^*) = (\lambda(t_i^{**}, \boldsymbol{s}_{i1}^*), \lambda(t_i^{**}, \boldsymbol{s}_{i2}^*), \ldots, \lambda(t_i^{**}, \boldsymbol{s}_{im_i^*}^*))'$, and $\boldsymbol{\varepsilon}(t_i^*) = \boldsymbol{y}(t_i^*) - \boldsymbol{\lambda}(t_i^*)$. Assume that t_i^* is the current time point in process monitoring, and we have decorrelated the data at all previous time points t_{i-1}^*, $t_{i-2}^*, \ldots, t_1^*$. The decorrelated data in the past are denoted as $\boldsymbol{e}(t_1^*), \boldsymbol{e}(t_2^*), \ldots, \boldsymbol{e}(t_{i-1}^*)$, and their standardized versions are denoted as $\widetilde{\boldsymbol{e}}(t_1^*), \widetilde{\boldsymbol{e}}(t_2^*), \ldots, \widetilde{\boldsymbol{e}}(t_{i-1}^*)$. After the observation $\boldsymbol{y}(t_i^*)$ at the current time point is obtained, we want to make it uncorrelated with them and its components uncorrelated with each other, which can be accomplished by the algorithm described below. Let the covariance matrix of $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i-1}', \boldsymbol{\varepsilon}(t_i^*)')'$ be denoted as $\boldsymbol{\Sigma}_{ii} = \begin{pmatrix} \boldsymbol{\Sigma}_{i-1,i-1} & \boldsymbol{V}_{i-1,i} \\ \boldsymbol{V}_{i-1,i}' & \boldsymbol{V}_{ii} \end{pmatrix}$, where $\boldsymbol{\varepsilon}_1 = \boldsymbol{\varepsilon}(t_1^*)$, $\boldsymbol{\Sigma}_{i-1,i-1} = \mathrm{Cov}(\boldsymbol{\varepsilon}_{i-1})$, $\boldsymbol{V}_{i-1,i} = \mathrm{Cov}(\boldsymbol{\varepsilon}_{i-1}, \boldsymbol{\varepsilon}(t_i^*))$ and $\boldsymbol{V}_{ii} = \mathrm{Cov}(\boldsymbol{\varepsilon}(t_i^*))$. By the Cholesky decomposition, we have $\boldsymbol{\Phi}_i \boldsymbol{\Sigma}_{ii} \boldsymbol{\Phi}_i' = \boldsymbol{D}_i$, where $\boldsymbol{\Phi}_i = \begin{pmatrix} \boldsymbol{\Phi}_{i-1} & \boldsymbol{0} \\ -\boldsymbol{V}_{i-1,i}' \boldsymbol{\Sigma}_{i-1,i-1}^{-1} & \boldsymbol{I}_{m_i^*} \end{pmatrix}$, $\boldsymbol{D}_i$ =diag$\{\boldsymbol{V}_{11}, \boldsymbol{V}_{22\cdot 1}, ..., \boldsymbol{V}_{ii\cdot i-1}\}$ and $\boldsymbol{V}_{ii\cdot i-1} = \boldsymbol{V}_{ii} - \boldsymbol{V}_{i-1,i}' \boldsymbol{\Sigma}_{i-1,i-1}^{-1} \boldsymbol{V}_{i-1,i}$. Then, we define $\boldsymbol{e}(t_i^*) = \boldsymbol{\varepsilon}(t_i^*) - \boldsymbol{V}_{i-1,i}' \boldsymbol{\Sigma}_{i-1,i-1}^{-1} \boldsymbol{\varepsilon}_{i-1}$. It can be checked that $\mathrm{Cov}(\boldsymbol{e}_i) = \boldsymbol{D}_i$, where $\boldsymbol{e}_i = \left(\boldsymbol{e}_{i-1}', \boldsymbol{e}(t_i^*)'\right)' = \Phi_i \boldsymbol{\varepsilon}_i$. Thus, $\boldsymbol{e}(t_i^*)$ is uncorrelated with $\boldsymbol{e}(t_1^*), \boldsymbol{e}(t_2^*), \ldots, \boldsymbol{e}(t_{i-1}^*)$. Define the standardized version of $\boldsymbol{e}(t_i^*)$ by

$$\widetilde{\boldsymbol{e}}(t_i^*) = V_{ii\cdot i-1}^{-\frac{1}{2}} \boldsymbol{e}(t_i^*) = \boldsymbol{V}_{ii\cdot i-1}^{-\frac{1}{2}} \left(\boldsymbol{\varepsilon}(t_i^*) - \boldsymbol{V}_{i-1,i}' \boldsymbol{\Sigma}_{i-1,i-1}^{-1} \boldsymbol{\varepsilon}_{i-1} \right). \tag{9}$$

Then, $\widetilde{\boldsymbol{e}}(t_1^*), \widetilde{\boldsymbol{e}}(t_2^*), \ldots, \widetilde{\boldsymbol{e}}(t_i^*)$ would be mutually uncorrelated with covariance matrices $\boldsymbol{I}_{m_1^*}, \boldsymbol{I}_{m_2^*}, \ldots,$ $\boldsymbol{I}_{m_i^*}$, respectively. Thus, the decorrelated data are both spatially and temporally uncorrelated. In practice, the quantities $\boldsymbol{\varepsilon}(t_i^*)$, $\boldsymbol{\varepsilon}_{i-1}$, $\boldsymbol{V}_{ii\cdot i-1}$, $\boldsymbol{V}_{i-1,i}$ and $\boldsymbol{\Sigma}_{i-1,i-1}$ on the right-hand-side of (9) are all unobservable. They can be replaced by their estimates, using the estimated mean and covariance functions $\widehat{\lambda}(t, \boldsymbol{s})$ and $\widehat{V}(t, \boldsymbol{s}; t', \boldsymbol{s}')$ defined in (4) and (7)-(8), respectively. The resulting decorrelated data are denoted as $\widehat{\widetilde{\boldsymbol{e}}}(t_1^*), \widehat{\widetilde{\boldsymbol{e}}}(t_2^*), \ldots, \widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$. They should be asymptotically uncorrelated.

Remark 1 *In the above sequential decorrelation procedure, for each i, we need to calculate $\boldsymbol{\Sigma}_{ii}^{-1}$. This can be achieved by the following recursive formula:*

$$\boldsymbol{\Sigma}_{ii}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{i-1,i-1}^{-1} + \boldsymbol{\Sigma}_{i-1,i-1}^{-1} \boldsymbol{V}_{i-1,i} \boldsymbol{V}_{ii\cdot i-1}^{-1} \boldsymbol{V}_{i-1,i}' \boldsymbol{\Sigma}_{i-1,i-1}^{-1} & -\boldsymbol{\Sigma}_{i-1,i-1}^{-1} \boldsymbol{V}_{i-1,i} \boldsymbol{V}_{ii\cdot i-1}^{-1} \\ -\boldsymbol{V}_{ii\cdot i-1}^{-1} \boldsymbol{V}_{i-1,i}' \boldsymbol{\Sigma}_{i-1,i-1}^{-1} & \boldsymbol{V}_{ii\cdot i-1}^{-1} \end{pmatrix}.$$

Remark 2 *When i increases, the computation and storage for the inverse matrix $\boldsymbol{\Sigma}_{ii}^{-1}$ could be demanding because its dimension increases with i. The recursive formula in Remark 1 partially*

overcomes this difficulty. An alternative strategy is described below. First, we notice that

$$\boldsymbol{e}(t_i^*) = \boldsymbol{\varepsilon}(t_i^*) - \sum_{j=1}^{i-1} \boldsymbol{B}_{i,j} \boldsymbol{V}_{jj\cdot j-1}^{-1} \boldsymbol{e}(t_{j-1}^*),$$

where $\boldsymbol{V}_{11\cdot 0} = \boldsymbol{V}_{11}$ *and* $\boldsymbol{B}_{i,j} = Cov(\boldsymbol{\varepsilon}(t_i^*), \boldsymbol{e}(t_j^*))$. *In practice, it is often reasonable to assume that the correlation between* $\boldsymbol{\varepsilon}(t_i^*)$ *and* $\boldsymbol{\varepsilon}(t_j^*)$ *becomes weaker when the two time points* t_i^* *and* t_j^* *are farther apart. Thus, it is often reasonable to assume that* $\boldsymbol{B}_{i,j} = \boldsymbol{0}$ *when* $t_i^* - t_j^* > \tau$, *where* $\tau > 0$ *is a threshold value. This will reduce the computation and data storage greatly without sacrificing much effectiveness of the proposed online process monitoring method.*

After data decorrelation, the original observations $\{\boldsymbol{y}(t_i^*), i = 1, 2, \ldots\}$ have been transformed to the asymptotically uncorrelated ones $\{\widehat{\widetilde{\boldsymbol{e}}}(t_i^*), i = 1, 2, \ldots\}$. Because each $\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$ is a linear combination of the original observations (cf., (9) after $\boldsymbol{\varepsilon}(t_i^*)$ is replaced by $\boldsymbol{y}(t_i^*) - \widehat{\boldsymbol{\lambda}}(t_i^*)$ and other quantities are replaced by their corresponding estimates as well), its distribution would be close to $N(\boldsymbol{0}, \boldsymbol{I}_{m_i^*})$ if $\boldsymbol{y}(t_i^*)$ is correlated with a substantial number of previous spatial observations. Consequently, the distribution of $\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)'\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$ would be close to $\chi^2_{m_i^*}$. Then, we suggest the following CUSUM chart to sequentially detect the upward shifts in the disease incidence rates:

$$C_i^+ = \max\left(0, C_{i-1}^+ + \frac{\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)'\widehat{\widetilde{\boldsymbol{e}}}(t_i^*) - m_i^*}{\sqrt{2m_i^*}} - k\right), \qquad \text{for } i \geq 1, \tag{10}$$

where $C_0^+ = 0$, $k > 0$ is an allowance constant, and $[\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)'\widehat{\widetilde{\boldsymbol{e}}}(t_i^*) - m_i^*]/\sqrt{2m_i^*}$ is the standardized version of $\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)'\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$. The chart gives a signal of an upward mean shift in the disease incidence rates when

$$C_i^+ > \gamma, \tag{11}$$

where $\gamma > 0$ is a control limit. The charting statistic C_i^+ has made use of all history data by using the cumulative information in the observed data. It is reset to 0 each time when the cumulative sum $C_{i-1}^+ + \left[\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)'\widehat{\widetilde{\boldsymbol{e}}}(t_i^*) - m_i^*\right]/\sqrt{2m_i^*}$ is smaller than k. This re-starting mechanism makes it possess some good theoretical properties (cf., Moustakides 1986).

To evaluate the performance of the proposed CUSUM chart (10)-(11), in the literature we usually use the IC average run length (ARL), denoted as ARL_0, which is the average number of time points from the beginning of process monitoring to the signal time under the condition that the process is IC, and the out-of-control (OC) ARL, denoted as ARL_1, which is the the average number of time points from the process shift time to the signal time after the process becomes OC.

Usually, the ARL_0 value is specified at a given level, and the chart performs better if its ARL_1 value is smaller when detecting a shift of a given size.

In the CUSUM chart (10)-(11), there are two parameters k and γ to choose. Usually, k is specified beforehand, and then γ is chosen to achieve a given ARL_0 level. It has been well demonstrated that a large k value is good for detecting large shifts, and a small k value is good for detecting small shifts (cf., Chapter 4, Qiu 2014). Commonly used k values include 0.1, 0.2, 0.3, 0.5 and 1.0. If the distribution of $\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$ is exactly $N(\mathbf{0}, \boldsymbol{I}_{m_i^*})$, for each i, then the control limit γ can be easily computed numerically. However, when i is small or when $\boldsymbol{y}(t_i^*)$ is correlated with a small number of previous spatial observations, the true distribution of $\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$ could be quite different from $N(\mathbf{0}, \boldsymbol{I}_{m_i^*})$. To make the proposed chart more robust to the normality of $\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)$, we suggest using the block bootstrap procedure (cf., Lahiri 2003) for determining γ, which is described below.

For the IC dataset $\{y(t_i, \boldsymbol{s}_{ij}), j = 1, 2, \ldots, m_i, i = 1, 2, \ldots, n\}$, let us divide it into two parts: $\{y(t_i^{(1)}, \boldsymbol{s}_{ij}^{(1)}), j = 1, 2, \ldots, m_i^{(1)}, i = 1, 2, \ldots, n_1\}$ and $\{y(t_i^{(2)}, \boldsymbol{s}_{ij}^{(2)}), j = 1, 2, \ldots, m_i^{(2)}, i = 1, 2, \ldots, n_2\}$, where $n_1 + n_2 = n$. The first part is used for obtaining the estimates of $\lambda(t, \boldsymbol{s})$ and $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$, as described in Subsection 2.1, and the second part is used for determining γ as follows.

(1) Compute the decorrelated data $\{\widehat{\widetilde{\boldsymbol{e}}}(t_i^{(2)}), i = 1, 2, ..., n_2\}$ from the original data $\{y(t_i^{(2)}, \boldsymbol{s}_{ij}^{(2)}), j = 1, 2, \ldots, m_i^{(2)}, i = 1, 2, \ldots, n_2\}$, as discussed above.

(2) Randomly choose a sequence of integers from $\{1, 2, ..., n_2 - b\}$ with replacement, where b is a pre-specified block size. The selected integers are denoted as $\{i_1, i_2, \ldots\}$. For a given control limit γ, calculate the run length (RL) by $RL_0(\gamma) = \min\{i \geq 1, C_i^+ > \gamma\}$, where C_i^+ is calculated by (10) after the decorrelated data $\{\widehat{\widetilde{\boldsymbol{e}}}(t_i^*), i = 1, 2, \ldots\}$ are replaced by the bootstrap sample $\{\widehat{\widetilde{\boldsymbol{e}}}(t_{i_1}^{(2)}), \widehat{\widetilde{\boldsymbol{e}}}(t_{i_1+1}^{(2)}), \ldots, \widehat{\widetilde{\boldsymbol{e}}}(t_{i_1+b}^{(2)}), \widehat{\widetilde{\boldsymbol{e}}}(t_{i_2}^{(2)}), \widehat{\widetilde{\boldsymbol{e}}}(t_{i_2+1}^{(2)}), \ldots, \widehat{\widetilde{\boldsymbol{e}}}(t_{i_2+b}^{(2)}), \ldots\}$. This process is then repeated for B times, and the average of the B values of $RL_0(\gamma)$ is denoted as $ARL_0(\gamma)$.

(3) Search for the γ value by the bi-section or other alternative numerical procedures (cf., Capizzi and Masarotto 2016) such that $ARL_0(\gamma)$ equals the pre-specified value of ARL_0.

The entire proposed method for online spatio-temporal data monitoring can be summarized below.

Summary of the Proposed Spatio-Temporal Data Monitoring Procedure

- Split the IC dataset $\{y(t_i, \boldsymbol{s}_{ij}), j = 1, 2, \ldots, m_i, i = 1, 2, \ldots, n\}$ into two parts: $\{y(t_i^{(1)}, \boldsymbol{s}_{ij}^{(1)}), j = 1, 2, \ldots, m_i^{(1)}, i = 1, 2, \ldots, n_1\}$ and $\{y(t_i^{(2)}, \boldsymbol{s}_{ij}^{(2)}), j = 1, 2, \ldots, m_i^{(2)}, i = 1, 2, \ldots, n_2\}$.

- Estimate $\lambda(t, \boldsymbol{s})$, $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$ and $\sigma^2(t, \boldsymbol{s})$ from the first part of the IC dataset $\{y(t_i^{(1)}, \boldsymbol{s}_{ij}^{(1)}), j = 1, 2, \ldots, m_i^{(1)}, i = 1, 2, \ldots, n_1\}$ by (4), (7) and (8).

- Determine the control limit γ used in (11) from the second part of the IC dataset $\{y(t_i^{(2)}, \boldsymbol{s}_{ij}^{(2)}), j = 1, 2, \ldots, m_i^{(2)}, i = 1, 2, \ldots, n_2\}$. To this end, the IC data $\{y(t_i^{(2)}, \boldsymbol{s}_{ij}^{(2)}), j = 1, 2, \ldots, m_i^{(2)}, i = 1, 2, \ldots, n_2\}$ need to be decorrelated first (cf., (9)), and then γ is determined by the block bootstrap procedure from the decorrelated data, as discussed above.

- For online monitoring of the process observations $\{y(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, 2, \ldots, m_i^*, i = 1, 2, \ldots\}$, the process observations need to be decorrelated first, and the decorrelated observations are denoted as $\{\widehat{\widetilde{\boldsymbol{e}}}(t_i^*), j = 1, 2, \ldots, m_i^*, i = 1, 2, \ldots\}$.

- Apply the proposed chart (10)-(11) to the decorrelated observations $\{\widehat{\widetilde{\boldsymbol{e}}}(t_i^*), j = 1, 2, \ldots, m_i^*, i = 1, 2, \ldots\}$. It gives a signal when (11) is true.

3 Simulation Studies

In this section, we evaluate the numerical performance of our proposed method described in the previous section by reporting some simulation results. In the simulation studies, the true IC mean function $\lambda(t, \boldsymbol{s})$ is assumed to be

$$\lambda(t, \boldsymbol{s}) = 0.02 + 0.01e^{-(s_x + s_y)} + 0.01\cos(t),$$

where $\boldsymbol{s} = (s_x, s_y)'$ as in expression (3). The basic time interval is assumed to be $[0, 1]$, and the spatial domain is assumed to be $\Omega = [0, 1] \times [0, 1]$. When $t = 0.25$ or 0.75, $\lambda(t, \boldsymbol{s})$ is shown by the two plots in the first column of Figure 2. For simplicity, it is assumed that the observation times $\{t_i, i = 1, 2, \ldots, n\}$ are equally spaced in $[0, 1]$, and the spatial locations $\{\boldsymbol{s}_{ij}, j = 1, 2, \ldots, m\}$ keep unchanged over time and they are equally spaced in Ω as well. The random errors $\{\varepsilon(t_i, \boldsymbol{s}_{ij})\}$ in the model (1) are generated by using the R function **spatialnoise()** in the package neuRosim, where the parameter ρ that controls both the spatial and temporal data correlation is chosen to be 0.1 (representing a case with a relatively weak data correlation), 0.3, or 0.5 (representing a case with a

relatively strong data correlation), and the variance parameter σ^2 is chosen to be 0.01^2. By using this R function, all random errors $\{\varepsilon(t_i, \boldsymbol{s}_{ij})\}$ are normally distributed.

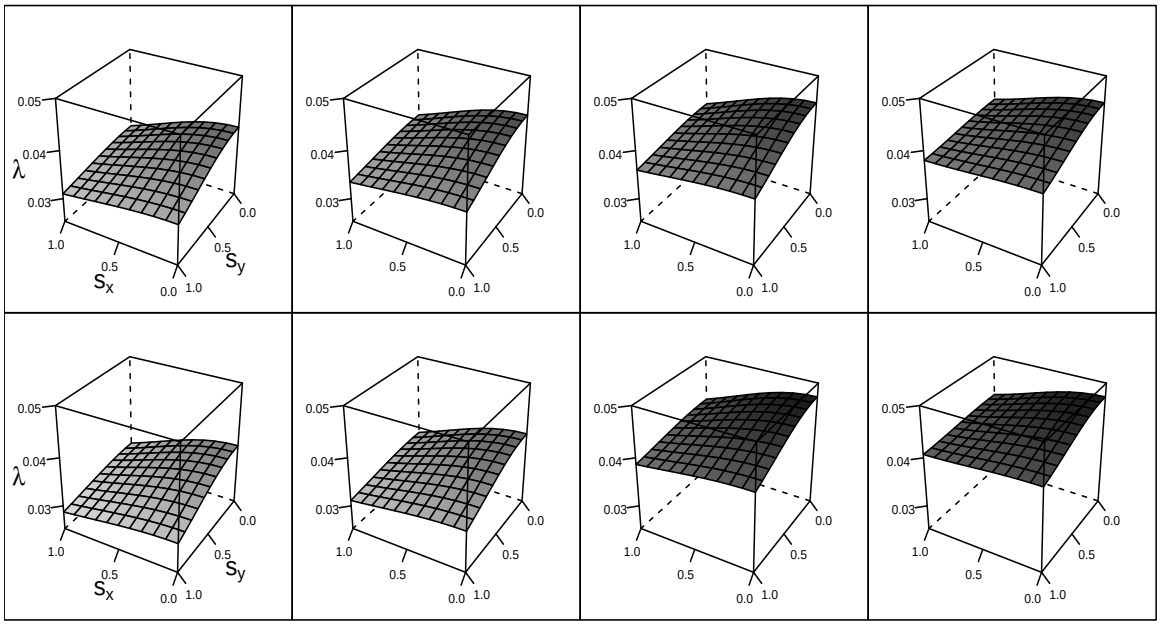

Figure 2: IC mean function $\lambda(t, \boldsymbol{s})$ (1st column) and three OC mean functions (2nd, 3rd and 4th columns) when $t = 0.25$ (1st row) and $t = 0.75$ (2nd row).

Performance of the proposed chart (10)-(11). We first study the IC performance of the proposed method in cases when the IC mean function $\lambda(t, \boldsymbol{s})$ and the IC covariance function $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$ are both assumed unknown. In such cases, they should be estimated from an IC dataset first. Here, the IC data are generated at n_1 time points in $[0, 1]$ and m equally spaced positions in the region $[0, 1] \times [0, 1]$, as discussed above. Then, the control limit γ of the CUSUM chart (10)-(11) is searched from another independently generated IC dataset of the same size by the block bootstrap procedure with $B = 10,000$ and $b = 5$, as discussed at the end of Section 2. Thus, we choose $n_2 = n_1$ here for simplicity. In the simulation studies below, (m, n_1) is chosen to be $(64, 200)$ or $(100, 300)$ to investigate the effect of the IC sample size on the performance of the proposed chart. For the chart, its ARL_0 value is chosen to be 25 or 50, and its allowance constant k is chosen to be 0.1, 0.3, or 0.5. Then, the chart is applied to the sequential observations generated from the IC model, and an IC run length value can be computed. This sequential monitoring process is repeated for 500 times, and the average of the 500 IC run length values is computed to estimate the actual ARL_0 value of our proposed method. Because this ARL_0 value depends on

the initial IC dataset used for estimating $\lambda(t, \boldsymbol{s})$ and $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$ and for determining the control limit γ, we repeat the entire simulation, from estimation of $\lambda(t, \boldsymbol{s})$ and $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$, determination of γ, to computation of ARL_0, for 100 times to reduce the randomness of the results. The average of the 100 ARL_0 values is used as the simulated actual ARL_0 value of the method. The results in different cases, together with the corresponding standard errors, are presented in Table 1. From the table, we can have the following conclusions. (i) The simulated actual ARL_0 values are close to the nominal ARL_0 values in all cases considered, implying that the proposed method is reliable to use in practice. (ii) The simulated actual ARL_0 values are closest to the nominal ARL_0 values in cases when $\rho = 0.1$, compared to cases when $\rho = 0.3$ or 0.5, which is intuitively reasonable. (iii) When (m, n_1) are larger, the difference between the simulated actual ARL_0 value and the corresponding nominal ARL_0 value gets smaller and the related standard error becomes smaller, which indicates that the control chart (10)-(11) constructed based on a larger IC dataset would have a more reliable IC performance.

Table 1: Simulated actual ARL_0 values and their standard errors (in parentheses) of the proposed CUSUM chart (10)-(11) in different cases.

		$\rho = 0.1$		$\rho = 0.3$		$\rho = 0.5$	
(m, n_1)	k	ARL_0=25	$ARL_0 = 50$	ARL_0=25	$ARL_0 = 50$	ARL_0=25	$ARL_0 = 50$
(64,200)	0.1	24.46(0.81)	48.75(2.03)	23.83(0.72)	47.95(1.90)	23.26(0.69)	47.28(1.85)
	0.3	24.48(0.87)	49.21(2.22)	23.96(0.81)	48.52(2.00)	23.66(0.80)	47.38(2.07)
	0.5	24.70(0.87)	49.41(2.23)	24.04(0.90)	48.92(2.23)	23.78(0.79)	47.54(2.28)
(100,300)	0.1	24.81(0.42)	49.29(1.05)	24.11(0.41)	48.19(1.05)	23.54(0.47)	47.83(1.11)
	0.3	24.92(0.47)	49.38(1.19)	24.33(0.43)	48.89(1.07)	23.92(0.47)	48.02(1.08)
	0.5	24.97(0.49)	49.43(1.21)	24.45(0.42)	49.12(1.08)	24.09(0.51)	48.34(1.08)

Next, we study the OC performance of the proposed method. To this end, the following three different types of shift functions are considered and each type has four different shift magnitudes:

(i) $\delta_1(t, \boldsymbol{s}) = 0.0025\nu$, for $\nu = 1, 2, 3, 4$;

(ii) $\delta_2(t, \boldsymbol{s}) = 0.0025 + 0.01t\nu$, for $\nu = 1, 2, 3, 4$;

(iii) $\delta_3(t, \boldsymbol{s}) = 0.0025 + 0.01t + 0.001(s_x^2 + s_y^2)\nu$, for $\nu = 1, 2, 3, 4$.

So, after the shift, the OC mean function becomes $0.02 + 0.01e^{-(s_x+s_y)} + 0.01\cos(t) + \delta_l(t, \boldsymbol{s})$, for $l = 1, 2$, or 3, which are shown in the 2nd, 3rd and 4th columns of Figure 2, when $\nu = 1$ and $t = 0.25$ (1st row) or 0.75 (2nd row). To compute the ARL_1 value of the chart, the CUSUM chart (10)-(11) is first applied to the sequential observations generated from model (1) with one of the above three shifts, and the related run length value is recorded. Then, the above step is repeated for 500 times to compute the ARL_1 value. Finally, the entire simulation, from estimation of $\lambda(t, \boldsymbol{s})$ and $V(t, \boldsymbol{s}; t', \boldsymbol{s}')$, determination of γ, to computation of ARL_1, is repeated for 100 times to compute the averaged value of the 100 ARL_1 values. The results when $ARL_0 = 50$, together with the corresponding standard errors, are presented in Table 2. From the table, it can be seen that i) the proposed method has a good shift detection power, and the method with a larger k performs better for detecting relatively large shifts, as expected, and ii) the control chart based on a larger IC dataset has a better shift detection performance.

Comparison with alternative methods. Next, we compare our proposed method, denoted as NEW, with four representative existing control charts: (i) the distribution-free multivariate EWMA chart by Chen et al. (2016), denoted as DFEWMA, (ii) the multivariate CUSUM chart proposed by Crosier (1988), denoted as MCUSUM, (iii) the multivariate EWMA chart (MEWMA) originally discussed by Lowry et al. (1992), and (iv) the Shewhart chart based on the Hotelling's T^2 statistic that was discussed in Tracy et al. (1992), denoted as T^2. When using these four conventional charts, the spatial data at each observation time are treated as a multivariate observation, which is appropriate because the spatial locations in the simulation studies considered in this section are unchanged over time. These charts all assume that the IC distribution is unchanged over time. Besides, it is assumed that the multivariate observations are normally distributed in the charts MCUSUM, MEWMA and T^2. These assumptions are often violated in practice, as discussed in Section 1, although the related control charts are routinely used in different applications. To compare the five related control charts, we set their actual ARL_0 values at 50, and the results when $ARL_0 = 25$ would result in similar conclusions. For the DFEWMA chart, its window size parameter w needs to be properly specified beforehand. By the suggestion in Chen et al. (2016), it can be chosen as the smallest integer that satisfies $(1 - w_p)^w \leq 0.05$, where w_p is the weighting parameter of the DFEWMA chart. This strategy is used throughout this paper. In the charts MCUSUM, MEWMA and T^2, the mean and covariance matrix of the IC distribution are estimated from the

Table 2: Calculated ARL_1 values and their standard errors (in parentheses) of the proposed procedure for detecting different shifts when $ARL_0 = 50$, $(m, n_1) = (64, 200)$ or $(100, 300)$, $\rho = 0.1$, 0.3 or 0.5, and $k = 0.1, 0.3$ or 0.5. In each row, bold numbers denote cases with the smallest ARL_1 values when comparing different k values with a fixed value of ρ.

		$\rho = 0.1$			$\rho = 0.3$			$\rho = 0.5$		
Type	ν	$k = 0.1$	$k = 0.3$	$k = 0.5$	$k = 0.1$	$k = 0.3$	$k = 0.5$	$k = 0.1$	$k = 0.3$	$k = 0.5$
$(m, n_1) = (64, 200)$										
$\delta_1(t, \mathbf{s})$	1	**17.68(0.59)**	18.20(0.66)	18.77(0.75)	**19.40(0.62)**	19.82(0.69)	20.78(0.78)	23.40(0.87)	**23.23(0.93)**	23.94(1.04)
	2	4.85(0.10)	4.23(0.10)	**4.03(0.12)**	5.77(0.16)	5.22(0.16)	**5.04(0.16)**	8.35(0.33)	7.83(0.33)	**7.74(0.34)**
	3	2.27(0.04)	1.86(0.03)	**1.66(0.02)**	2.63(0.06)	2.21(0.04)	**1.98(0.04)**	3.64(0.10)	3.19(0.09)	**2.96(0.09)**
	4	1.40(0.02)	1.17(0.01)	**1.09(0.01)**	1.59(0.03)	1.33(0.01)	**1.21(0.01)**	2.09(0.04)	1.78(0.03)	**1.61(0.03)**
$\delta_2(t, \mathbf{s})$	1	12.92(0.30)	**12.61(0.33)**	12.84(0.38)	14.18(0.36)	**13.87(0.37)**	14.10(0.39)	17.03(0.49)	**16.58(0.51)**	16.72(0.55)
	2	10.82(0.20)	10.34(0.22)	**10.28(0.25)**	11.80(0.27)	**11.34(0.26)**	11.35(0.27)	14.07(0.35)	**13.52(0.37)**	13.54(0.39)
	3	9.51(0.16)	8.98(0.17)	**8.86(0.19)**	10.33(0.22)	9.83(0.21)	**9.74(0.21)**	12.26(0.29)	11.69(0.30)	**11.63(0.31)**
	4	8.60(0.13)	8.03(0.14)	**7.87(0.16)**	9.32(0.18)	8.79(0.17)	**8.64(0.17)**	10.98(0.26)	10.42(0.26)	**10.28(0.26)**
$\delta_3(t, \mathbf{s})$	1	8.95(0.21)	8.40(0.23)	**8.38(0.28)**	10.10(0.26)	9.65(0.28)	**9.62(0.29)**	12.94(0.41)	**12.37(0.42)**	12.42(0.45)
	2	6.14(0.13)	5.50(0.13)	**5.35(0.12)**	7.12(0.18)	6.56(0.18)	**6.42(0.19)**	9.62(0.33)	9.05(0.33)	**8.99(0.34)**
	3	4.38(0.08)	3.79(0.08)	**3.54(0.08)**	5.10(0.12)	4.54(0.11)	**4.32(0.11)**	7.05(0.23)	6.50(0.23)	**6.32(0.23)**
	4	3.28(0.06)	2.78(0.05)	**2.51(0.05)**	3.80(0.08)	3.30(0.07)	**3.05(0.07)**	5.23(0.16)	4.71(0.15)	**4.50(0.15)**
$(m, n_1) = (100, 300)$										
$\delta_1(t, \mathbf{s})$	1	**13.87(0.28)**	13.98(0.32)	14.57(0.36)	15.15(0.38)	**14.88(0.40)**	15.44(0.44)	18.97(0.62)	**18.66(0.63)**	19.16(0.65)
	2	3.75(0.04)	3.17(0.04)	**2.90(0.04)**	4.29(0.06)	3.70(0.06)	**3.45(0.06)**	5.92(0.17)	5.36(0.16)	**5.16(0.18)**
	3	1.84(0.01)	1.49(0.01)	**1.32(0.01)**	2.08(0.02)	1.70(0.02)	**1.51(0.01)**	2.73(0.04)	2.33(0.04)	**2.10(0.04)**
	4	1.13(0.00)	1.02(0.00)	**1.01(0.00)**	1.27(0.01)	1.09(0.00)	**1.05(0.00)**	1.62(0.02)	1.36(0.01)	**1.23(0.01)**
$\delta_2(t, \mathbf{s})$	1	11.42(0.18)	**11.06(0.19)**	11.23(0.22)	12.42(0.23)	**11.81(0.23)**	11.95(0.25)	15.19(0.39)	**14.75(0.40)**	14.86(0.42)
	2	10.03(0.13)	**9.12(0.14)**	9.15(0.16)	10.84(0.17)	10.21(0.17)	**10.20(0.18)**	13.11(0.29)	12.58(0.29)	**12.55(0.30)**
	3	8.66(0.11)	**7.71(0.11)**	7.75(0.12)	9.57(0.14)	8.64(0.14)	**8.42(0.15)**	11.71(0.23)	11.14(0.23)	**11.05(0.24)**
	4	7.38(0.09)	7.06(0.09)	**7.00(0.11)**	8.69(0.12)	7.86(0.12)	**7.71(0.13)**	10.70(0.19)	10.14(0.19)	**9.99(0.19)**
$\delta_3(t, \mathbf{s})$	1	7.47(0.10)	6.87(0.11)	**6.79(0.12)**	8.34(0.15)	7.69(0.15)	**7.60(0.16)**	10.85(0.29)	10.24(0.29)	**10.12(0.30)**
	2	5.03(0.06)	4.41(0.06)	**4.20(0.06)**	5.72(0.08)	5.05(0.08)	**4.85(0.09)**	7.61(0.19)	7.03(0.18)	**6.83(0.19)**
	3	3.60(0.04)	3.05(0.03)	**2.78(0.03)**	4.11(0.05)	3.53(0.04)	**3.28(0.05)**	5.47(0.12)	4.93(0.11)	**4.68(0.12)**
	4	2.71(0.02)	2.24(0.02)	**2.00(0.02)**	3.09(0.03)	2.59(0.03)	**2.34(0.03)**	4.09(0.07)	3.60(0.07)	**3.34(0.07)**

IC data. To make the comparison as fair as possible, the allowance constant k in the CUSUM charts MCUSUM and NEW and the weighting parameter in the EWMA charts DFEWMA and MEWMA are chosen such that their ARL_1 values reach the minimum when detecting a given shift (i.e., their optimal performance is compared here). The results are presented in Figure 3. From the figure, it can be seen that (i) our proposed method outperforms all four existing methods in all cases considered, and (ii) in most cases our method outperforms them in quite large margins.

Why is data decorrelation needed? In the proposed CUSUM chart (10)-(11), the original process observations $\{\boldsymbol{y}(t_i^*), i = 1, 2, \ldots\}$ need to be decorrelated (cf., (9)) before the chart is actually applied. That is because the conventional CUSUM chart has a good performance only when process observations are serially independent (cf., Chapter 4, Qiu 2014). As explained before Expression (10), the decorrelated data would be asymptotically normally distributed and thus asymptotically independent under some regularity conditions. Therefore, data decorrelation should be able to improve the performance of the proposed chart NEW. To confirm this, we present an example to compare NEW with the version that is applied to the standardized process observations $\{\widetilde{\boldsymbol{\varepsilon}}(t_i^*)\}$ directly, where $\widetilde{\boldsymbol{\varepsilon}}(t_i^*) = (\widetilde{\varepsilon}(t_i^*, \boldsymbol{s}_{i1}^*), \widetilde{\varepsilon}(t_i^*, \boldsymbol{s}_{i2}^*), \ldots, \widetilde{\varepsilon}(t_i^*, \boldsymbol{s}_{im_i^*}^*))'$, and $\widetilde{\varepsilon}(t_i^*, \boldsymbol{s}_{ij}^*) = (y(t_i^*, \boldsymbol{s}_{ij}^*) - \widehat{\lambda}(t_i^*, \boldsymbol{s}_{ij}^*))/\widehat{\sigma}(t_i^*, \boldsymbol{s}_{ij}^*)$, for $j = 1, 2, \ldots, m_i^*$ and $i = 1, 2, \ldots$. The latter chart is denoted as NEWWOD, representing NEW without data decorrelation. More specifically, NEWWOD is same as NEW, except that the decorrelated data $\{\widehat{\widetilde{\boldsymbol{e}}}(t_i^*)\}$ used in (10) need to be replaced by $\{\widetilde{\boldsymbol{\varepsilon}}(t_i^*)\}$. In the example of Figure 3, let us consider cases when $(m, n_1) = (64, 200)$, $ARL_0 = 50$, $\rho = 0.1, 0.3$, or 0.5, and the shift function is $\delta_3(t, \boldsymbol{s})$. In such cases, the calculated optimal ARL_1 values of NEW and NEWWOD are presented in Figure 4. From the figure, it is clear that (i) the performance of NEW and NEWWOD is similar in cases when the data correlation is small (i.e., cases when $\rho = 0.1$), and (ii) NEW has a better performance than NEWWOD in cases when the data correlation gets larger (i.e., cases when $\rho = 0.3$ or 0.5). Thus, data decorrelation indeed improves the performance of NEW.

4 Application to the Florida Influenza-Like Illness Dataset

In this section, we apply our proposed method to the Florida Influenza-Like Illness (ILI) dataset described in Section 1. In this dataset, numbers of new ILI patients in all 67 counties (i.e. $m_i = 67$ for all i in model (1)) were reported daily from 2012 to 2015. Then, the ILI incidence rate on a

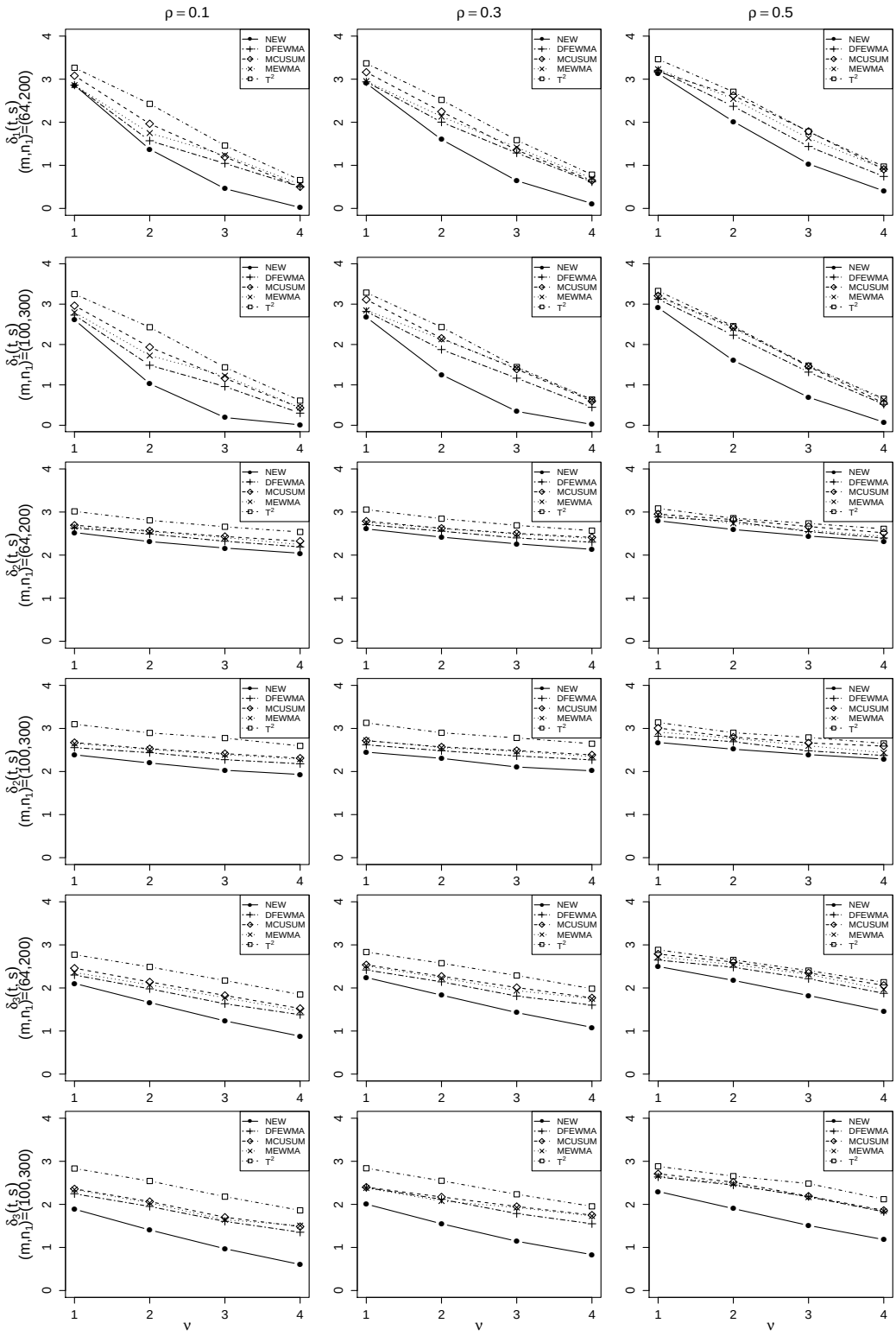

Figure 3: Calculated optimal ARL_1 values of the five control charts for different (m, n_1) and ρ values when the nominal ARL_0 is fixed at 50. In each plot, the y-axis denotes the calculated optimal ARL_1 values in natural log scale.

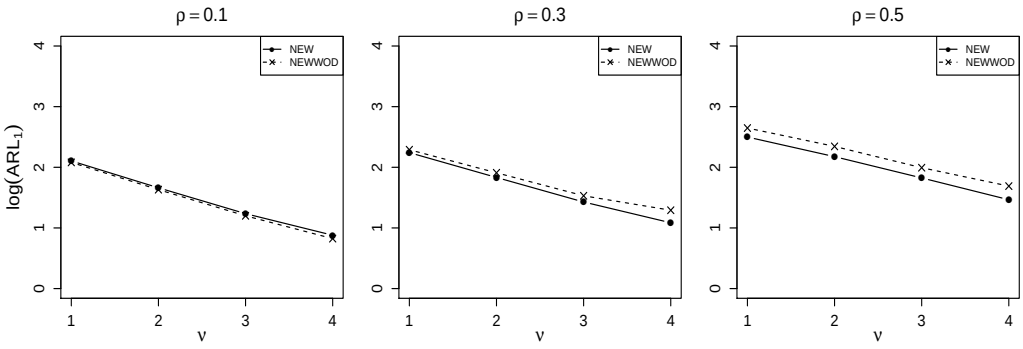

Figure 4: Calculated optimal ARL_1 values of NEW and NEWWOD in cases when $(m, n_1) = (64, 200)$, $ARL_0 = 50$, $\rho = 0.1, 0.3$, or 0.5, and the shift function is $\delta_3(t, \boldsymbol{s})$. In each plot, the y-axis denotes the calculated optimal ARL_1 values in natural log scale.

specific day for each county can be calculated as the ratio of the number of ILI patients on that day and the county population number that can be obtained from the Florida Office of Economic and Demographic Research (http://edr.state.fl.us/Content/population-demographics/data/). The calculated ILI incidence rates in four representative counties Collier, Nassau, Pinellas and Santa Rosa are shown in Figure 5. From the plots of the figure, it can be seen that the incidence rates have seasonal patterns with winter peaks and summer troughs, and that it seems reasonable to assume that the pattern of the incidence rates is periodic from one year to the next. We also checked the spatial pattern of the data (cf., Figure 1), and found that there were some obvious spatial clusters of the disease incidence rates. The seasonality and spatial pattern of the data can be accommodated by our proposed method. Therefore, our method should be appropriate to use in this application.

From the data, it seems that there were no major disease outbreaks during the years 2012 and 2013. Therefore, that part of the data is used as the IC data in this example for setting up the proposed method. For the IC data, the Durbin-Watson test for checking the temporal data correlation in the four representative counties Collier, Nassau, Pinellas and Santa Rosa gives the p-values of $2.2 \times 10^{-16}, 0.018, 2.2 \times 10^{-16}$, and 2.2×10^{-16}, respectively, which imply significant temporal data correlation in the four counties. The Moran's I test for checking spatial correlation for the data on 06/01/2012 and 12/01/2012 (cf., Figure 1) gives the p-values of 0.471 and 0.004, respectively. So, the spatial correlation on 12/01/2012 (a winter time) is significant, although it is insignificant on 06/01/2012 (a summer time). The IC data are then divided into two parts, as discussed in Section 2. The IC data in year 2013 are used for estimating the mean and covariance function of

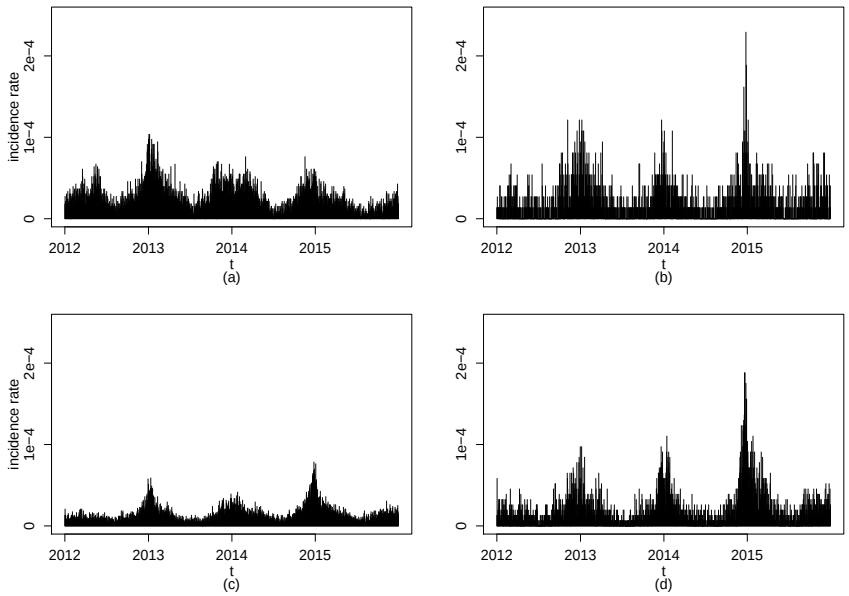

Figure 5: Observed ILI incidence rates in four representative counties of Florida: Collier (plot (a)), Nassau (plot (b)), Pinellas (plot (c)), and Santa Rosa (plot (d)).

the baseline model (cf., Section 2.1), and the IC data in year 2012 are used for determining the control limit γ of the CUSUM chart (10)-(11) by the block bootstrap procedure with $B = 10,000$ and $b = 5$ (cf., Section 2.2). So, n_1 and n_2 are 365 and 366, respectively, in this example. In the chart, we choose $k = 0.1$ and $ARL_0 = 200$, which are commonly used in the SPC literature. Then, the chart starts to sequentially monitor the disease incidence rates on January 1, 2014. This chart is presented in Figure 6(a). From the plot, the chart gives the first signal of shift on October 16, 2014. As a comparison, the four alternative charts DFEMWA, MCUSUM, MEWMA and T^2 are shown in Figure 6(b)-(e), respectively, where the weighting parameters or the allowance constants of the first three charts are all chosen to be 0.1 and the ARL_0 values of the four charts are all chosen to be 200. For the chart DFEWMA, its time-varying control limits are determined by the permutation procedure suggested in Chen et al. (2016). For the charts MCUSUM, MEWMA and T^2, their control limits are chosen under the assumptions that the spatial data at different time points follow a same multivariate normal distribution and they are temporally independent. Because these assumptions are obviously invalid, it has been well demonstrated in the literature that the performance of the related charts would be unreliable (cf., Qiu 2018). From the plots (b)-(e), we can see that (i) both DFEWMA and MEWMA give signals almost every day, (ii) MCUSUM gives the first signal in the middle of November which is about one month later than the first signal by NEW, and (iii) T^2 gives brief signals in the late January and middle October 2014 and

more convincing signals after mid-November 2014. We tried several other values of k for NEW and MCUSUM and several other values of the weighting parameter for DFEWMA and MEWMA. The results are similar. For instance, the CUSUM charts NEW and MCUSUM with $k = 0.5$ and the EWMA charts DFEWMA and MEWMA with the weighting parameter being 0.5 are presented in Figure 7. Because the chart T^2 does not have such parameters, it is omitted here. From the figure, we can see that NEW and MCUSUM give their first signals on October 14 and November 10 of 2014, respectively, while DFEWMA and MEWMA give signals many times starting in early January. Because some model assumptions (e.g., the IC process distribution does not change over time) of the charts DFEMWA, MCUSUM, MEWMA and T^2 are obviously violated in this example, their results may not be reliable.

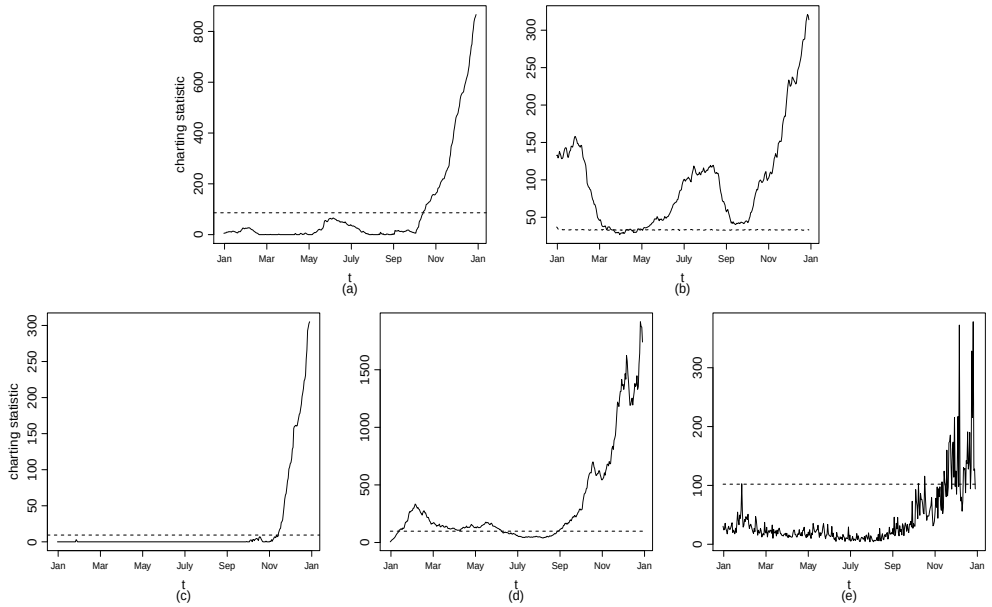

Figure 6: Five control charts for monitoring the ILI incidence rate data in cases when their weighting parameters or allowance constants are chosen to be 0.1: NEW (plot (a)), DFEWMA (plot (b)), MCUSUM (plot (c)), MEWMA (plot (d)), and T^2 (plot (e)). The dashed horizontal line in each plot denotes the control limit.

To check whether the signal from our method NEW is valid, the observed disease incidence rates in the entire Florida state during September 1 and December 31 in years 2012, 2013 and 2014 are shown in Figure 8. It can be seen that the disease incidence rates are quite similar in the month of September among the three years, and the rates in 2014 start to deviate upward from those in 2012 and 2013 beginning in October. This plot shows that a real disease outbreak occurs in the first half of October in 2014. Our method NEW detects such an outbreak shortly after it occurs.

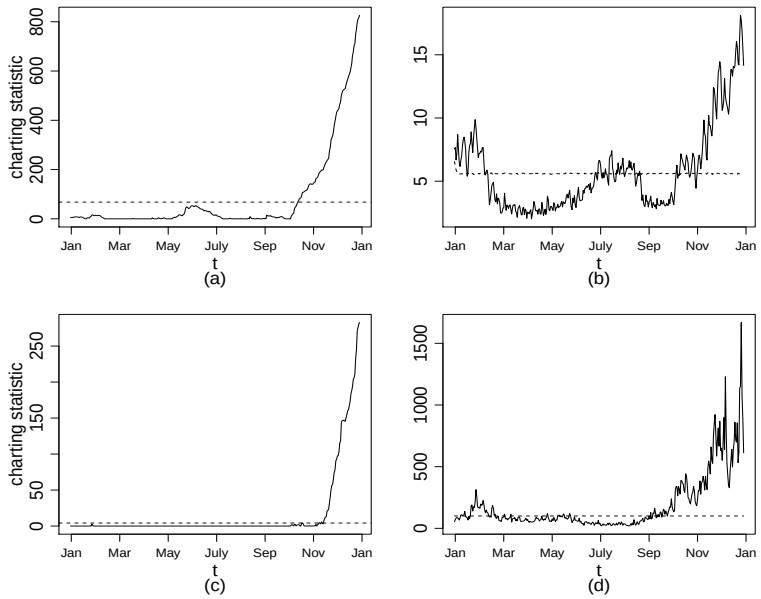

Figure 7: Four control charts for monitoring the ILI incidence rate data in cases when their weighting parameters or allowance constants are chosen to be 0.5: NEW (plot (a)), DFEWMA (plot (b)), MCUSUM (plot (c)), and MEWMA (plot (d)). The dashed horizontal line in each plot denotes the control limit.

To further investigate this, we present the maps of the residuals, defined as $\widehat{\boldsymbol{\varepsilon}}(t_i^*) = \boldsymbol{y}(t_i^*) - \widehat{\boldsymbol{\lambda}}(t_i^{**})$, on October 16th of the years 2012, 2013, and 2014 in the three maps of Figure 9, where $\widehat{\boldsymbol{\lambda}}(t_i^{**})$ is the estimated values of $\boldsymbol{\lambda}(t_i^{**})$ from the IC data and t_i^{**} is the time in $[0, T]$ that corresponds to t_i^* (cf., the related discussion in Section 2). It can be seen that the residual map for the year 2014 has stronger colors, implying that the residual values are larger in that year, compared to years 2012 and 2013. As an example, the Jackson county on the northern border has a strong color in year 2014, and its color is light in the previous two years. The observed disease incidence rates of this county during years 2012-2015 are shown in Figure 10, where the vertical line denotes the signal time of NEW. It can be seen that the incidence rates start to increase a bit earlier than the signal time and our method NEW successfully detects the increasing trend before the major disease outbreak in December 2014.

5 Concluding Remarks

We have described a sequential monitoring method for spatio-temporal disease surveillance and early detection, which was motivated by the Florida Influenza-Like Illness data. This method

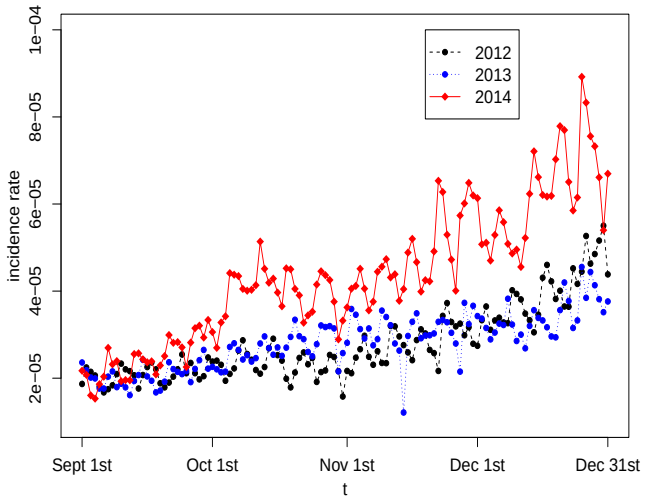

Figure 8: Disease incidence rates of the entire Florida state during September 1 and December 31 in years 2012, 2013 and 2014.

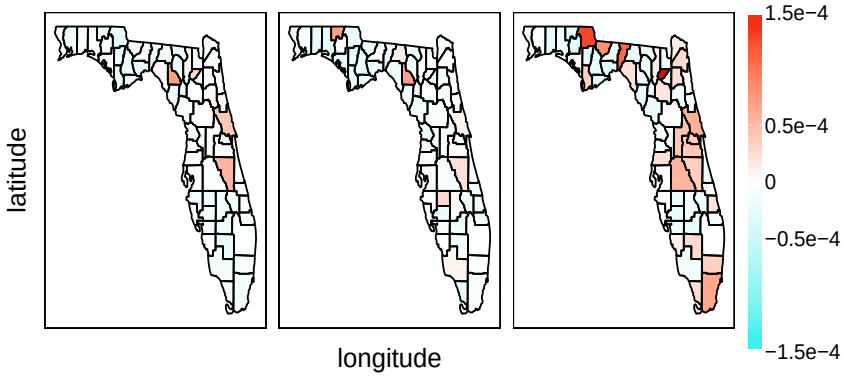

Figure 9: Maps of the residuals on October 16th of the years 2012 (left), 2013 (middle) and 2014 (right).

can accommodate time-varying IC distribution of the disease incidence rates. It does not impose any parametric assumptions on the disease incidence rate distribution, the disease incidence rate trajectory over both space and time, and the spatio-temporal data correlation. Both simulation studies and the application to the Florida Influenza-Like Illness data show that this method provides a reliable tool for disease surveillance applications. To use our proposed method, it is required to have an IC dataset available beforehand, from which a baseline model is estimated to describe the longitudinal pattern of the disease incidence rates in cases when no disease outbreaks are present. In practice, such IC data are only available for some diseases about which we have a reasonably good knowledge, such as the influenza and some other respiratory diseases. For some other diseases, especially those relatively new, it is often challenging to know whether disease outbreaks are present or not in a given period of time. It is also challenging to collect a suitable amount of IC data in such

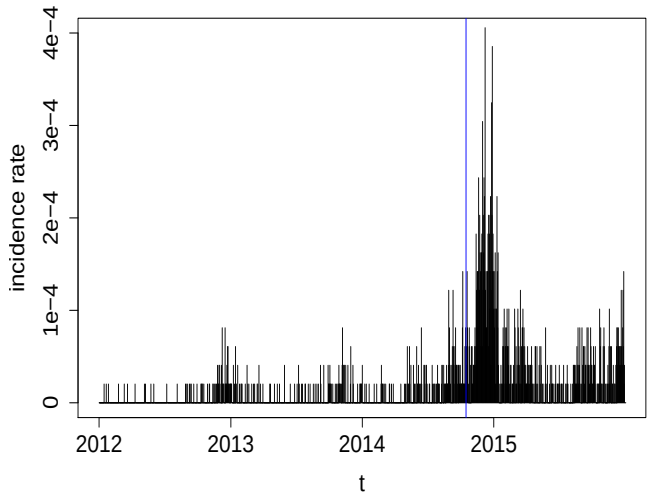

Figure 10: Observed disease incidence rates of the Jackson county in years 2012-2015. The vertical line indicates the signal time of our method NEW.

cases. This issue is related to the so-called phase-I SPC problem (cf., Chapter 3, Qiu 2014). But, conventional phase-I SPC charts cannot be applied to the current problem because they require the assumptions that the IC distribution of the disease incidence rates remains unchanged over time and observations are independent of each other, which are obviously invalid here. These limitations might be possible to overcome, using similar ideas to those discussed in Section 2. Another possible method to handle the challenges mentioned above is to use the so-called self-starting control charts (Hawkins 1987). The major idea of such charts is to keep expanding the IC data by combining the current IC data with the new observations after they are confirmed to be IC. Again, the conventional self-starting control charts cannot be applied to the current problem directly, and many issues need to be addressed properly. In addition, after the signal obtained from the proposed monitoring procedure, it is important to know when and where the detected disease outbreak occurs. This is related to the post-signal diagnosis discussed in the SPC literature (cf., Li et al. 2017, Xian et al. 2019), and has not been discussed in the current paper yet. All these topics require much future research effort.

Acknowledgments: The authors thank the editors and three referees for many constructive comments and suggestions which improved the quality of the paper greatly. This research is supported in part by an NSF grant.

Author Biosketches

Kai Yang is currently a PhD student of the Department of Biostatistics at the University of Florida. His thesis research is mainly on spatio-temporal data modeling and monitoring under the supervision of Professor Peihua Qiu. He has published two papers on nonparametric estimation of the mean and variance/covariance structures of spatial data, and has several other papers currently under review. Besides that topic, his thesis research also discusses effective process monitoring by using covariate information.

Peihua Qiu received his Ph.D. in statistics from the Department of Statistics at the University of Wisconsin - Madison in 1996. He worked as a senior research consulting statistician of the Biostatistics Center at the Ohio State University during 1996-1998. He then worked as an assistant professor (1998-2002), an associate professor (2002-2007), and a full professor (2007-2013) at the School of Statistics of the University of Minnesota. He is an elected fellow of the American Statistical Association, an elected fellow of the Institute of Mathematical Statistics, an elected member of the International Statistical Institute, a senior member of the American Society for Quality, and a lifetime member of the International Chinese Statistical Association. He served as associate editor for *Journal of the American Statistical Association*, *Biometrics*, *Technometrics*, *Surgery*, and *Statistical Papers*, and guest co-editor for *Multimedia Tools and Applications*, and *Quality and Reliability Engineering International*. He was the editor-elect (2013) and editor (2014-2016) of *Technometrics*. He is currently an associate editor of *Quality Engineering*, and the Professor and Founding Chair of the Department of Biostatistics at the University of Florida.

Peihua Qiu has made substantial contributions in the areas of jump regression analysis, image processing, statistical process control, survival analysis, and disease screening and surveillance. So far, he has published about 120 research papers in referred journals, many of which appeared in top journals, including *Technometrics, Journal of the American Statistical Association, Annals of Statistics, Annals of Applied Statistics, Journal of the Royal Statistical Society (Series B), Biometrika, Biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence*, and *IISE Transactions*. His research monograph titled *Image Processing and Jump Regression Analysis* (2005, Wiley) won the inaugural Ziegel prize in 2007 for its contribution in bridging the gap between jump regression analysis in statistics and image processing in computer science. His second book titled *Introduction to Statistical Process Control* was published in 2014 by Chapman & Hall/CRC.

References

Apley, D.W., and Lee, H.C. (2003), "Design of exponentially weighted moving average control charts for autocorrelated processes with model uncertainty," *Technometrics*, **45**, 187–198.

Altman, N.S. (1990), "Kernel smoothing of data with correlated errors," *Journal of the American Statistical Association*, **85**, 749–758.

Bakir, S.T., and Reynolds, M.R., Jr. (1979), "A nonparametric procedure for process control based on within group ranking," *Technometrics*, **21**, 175–183.

Banerjee, S., Carlin, C.P., Gelfand, A.E. (2004), *Hierarchical Modeling and Analysis for Spatial Data,* Chapman Hall/CRC: Boca Raton, FL.

Brabanter, K.D., Brabanter, J.D., Suykens, J.A.K., and De Moor, B. (2011), "Kernel regression in the presence of correlated errors," *Journal of Machine Learning Research*, **12**, 1955–1976.

Capizzi, G. (2015), "Recent advances in process monitoring: nonparametric and variable-selection methods for phase I and phase II," *Quality Engineering*, **27**, 44-67.

Capizzi, G., and Masarotto, G. (2013), "Phase I distribution-free analysis of univariate data," *Journal of Quality Technology*, **45**, 273–284.

Capizzi, G., and Masarotto, G. (2016), "Efficient control chart calibration by simulated stochastic approximation," *IIE Transactions*, **48**, 57–65.

Chakraborti, S., van der Laan, P. and Bakir, S.T. (2001), "Nonparametric control charts: an overview and some results," *Journal of Quality Technology*, **33**, 304–315.

Chakraborti, S., Qiu, P., and Mukherjee, A. (2015), "Editorial to the special issue: Nonparametric statistical process control charts," *Quality and Reliability Engineering International*, **31**, 1–2.

Chen, H., Zeng, D., and Yan, P. (2010), *Infectious Disease Informatics*, Springer: New York.

Chen, N., Zi, X., and Zou, C. (2016), "A distribution-free multivariate control chart," *Technometrics*, **58**, 448–459.

Choi, E., Hall, P., and Rousson, V. (2000), "Data sharpening methods for bias reduction in nonparametric regression," *The Annals of Statistics*, **28**, 1339–1355.

Choi, I., Li, B., and Wang, X. (2013), "Nonparametric estimation of spatial and space-time covariance function," *Journal of Agricultural, Biological, and Environmental Statistics*, **18**, 611–630.

Chouldechova, A., and Hastie, T. (2015), "Generalized additive model selection," arXiv:1506.03850.

Colosimo, B. M., Cicorella, P., Pacella, M., and Blaco, M. (2014), "From profile to surface monitoring: SPC for cylindrical surfaces via Gaussian processes," *Journal of Quality Technology*, **46**, 95–113.

Cressie, N., Wikle, C. (2011), *Statistics for Spatio-Temporal Data,* John Wiley & Sons: New York.

Crosier, R.B. (1988), "Multivariate generalizations of cumulative sum quality-control schemes," *Technometrics*, **30**, 291–303.

Diggle, P.J. (2014), *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns (3rd ed.)*, Chapman & Hall/CRC: London.

Fiore, A.E., Uyeki, T.M., Broder, K., Finelli, L., Euler, G.L., Singleton, J.A., Iskander, J.K., Wortley, P.M., Shay, D.K., Bresee, J.S., and Cox, N.J. (2010), "Prevention and control of influenza with vaccines: recommendations of the advisory committee on immunization practices (acip)," *MMWR Recommendations and Reports*, **59**, 1–62.

Genton, M.G. (2007), "Separable approximations of space-time covariance matrices," *Environmetrics*, **18**, 681–695.

Gneiting, T. (2002), "Nonseparable stationary covariance functions for space-time data," *Journal of the American Statistical Association*, **97**, 590–600.

Gneiting, T., and Guttorp, P. (2010), "Continuous parameter spatio-temporal processes," In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 427–436, CRC Press: Boca Raton, FL.

Gonzalez, J.A., Rodriguez-Cortes, F.J., Cronie, O., and Mateu, J. (2016), "Spatio-temporal point process statistics: a review," *Spatial Statistics*, **18**, 505–544.

Hackl, P., and Ledolter, J. (1991), "A control chart based on ranks," *Journal of Quality Technology*, **23**, 117–124.

Hawkins, D.M. (1987), "Self-starting cusum charts for location and scale," *The Statistician*, **36**, 299–316.

Heuvelink, G.B.M., and Griffith, D.A. (2010), "Space time geostatistics for geography: a case study of radiation monitoring across parts of Germany," *Geographical Analysis*, **42**, 161–179.

Holland, M.D., and Hawkins, D.M. (2014), "A control chart based on a nonparametric multivariate change-point model," *Journal of Quality Technology*, **46**, 63–77.

Jiang, W., Han, S.W., Tsui, K.L., and Woodall, W.H. (2011), "Spatiotemporal surveillance methods in the presence of spatial correlation," *Statistics in Medicine*, **30**, 569-583.

Kafadar, K. (1996), "Smoothing geographical data, particularly rates of disease," *Statistics in Medicine*, **15**, 2539–2560.

Kite-Powell, A., Ofori-Addo, A., and Hamilton, J. (2010), *ESSENCE User Guide (Version 1.0)*, Florida Department of Health, Bureau of Epidemiology.

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.

Lahiri, S.N. (2003), *Resampling Methods for Dependent Data,* New York: Springer.

Last, J.M. (2001), *A Dictionary of Epidemiology (4th edition)*, Oxford University Press.

Li, J., and Qiu, P. (2016), "Nonparametric dynamic screening system for monitoring correlated longitudinal data," *IIE Transactions*, **48**, 772–786.

Li, J., and Qiu, P. (2017), "Construction of an efficient multivariate dynamic screening system," *Quality and Reliability Engineering International*, **33**, 1969–1981.

Li, W., and Qiu, P. (2019), "A general charting scheme for monitoring serially correlated data with short-memory dependence and nonparametric distributions," *IISE Transactions*, **52**, 61–74.

Li, J., Liu, K., and Xian, X. (2017), "Causation-based process monitoring and diagnosis for multivariate categorical processes," *IISE Transactions*, **49**, 332–343.

Lindström, J., Szpiro, A., Sampson, P.D., Bergen, S., Sheppard, L. (2015), "Spatiotemporal: an R package for spatio-temporal modelling of air-pollution," https://cran.rproject.org/web/packages/SpatioTemporal/index.html.

Lowry, C.A., Woodall, W.H., Champ, C.W., and Rigdon, S.E. (1992), "A multivariate exponentially weighted moving average control chart," *Technometrics*, **34**, 46–53.

Marshall, J.B., Spitzner, D.J., and Woodall, W.H. (2007), "Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time," *Statistics in Medicine*, **26**, 1579–1593.

Moustakides, G.V. (1986), "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, **14**, 1379–1387.

Noordzij, M., Dekker, F.W., Zoccali, C., and Jager, K.J. (2010), "Measures of disease frequency: prevalence and incidence." *Nephron Clinical Practice*, **115**, c17–c20.

Opsomer, J., Wang, Y., and Yang, Y. (2001), "Nonparametric regressin with correlated errors," *Statistical Science*, **16**, 134–153.

Psarakis, S., and Papaleonida, G.E.A. (2016), "SPC procedure for monitoring autocorrelated processe," *Quality Technology & Quantitative Management*, **4**, 501–540.

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

Qiu, P. (2008), "Distribution-free multivariate process control based on log-linear modeling," *IIE Transactions*, **40**, 664–677.

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

Qiu, P. (2018), "Some perspectives on nonparametric statistical process control," *Journal of Quality Technology*, **50**, 49–65.

Qiu, P., and Hawkins, D.M. (2001), "A rank based multivariate CUSUM procedure," *Technometrics*, **43**, 120–132.

Qiu, P., and Li, Z. (2011), "On nonparametric statistical process control of univariate processes," *Technometrics*, **53**, 390–405.

Qiu, P., and Xiang, D. (2014), "Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior," *Technometrics*, **56**, 248–260.

Qiu, P., and Xiang, D. (2015), "Surveillance of cardiovascular diseases using a multivariate dynamic screening system," *Statistics in Medicine*, **34**, 2204–2221.

Schabenberger, O., and Gotway, C.A. (2005), *Statistical Methods for Spatial Data Analysis,* New York: Chapman and Hall.

Shand, L., and Li B. (2017), "Modeling nonstationarity in space and time," *Biometrics*, **73**, 759–768.

Sun, Y., Hart, J.D., and Genton, M.G. (2012), "Nonparametric inference for periodic sequences," *Technometrics*, **54**, 83–96.

Tracy, N.D., Young, J.C. , and Mason, R.L. (1992), "Multivariate control charts for individual observations," *Journal of Quality Technology*, **24**, 88–95.

Vogt, M., and Linton, O. (2014), "Nonparametric estimation of a periodic sequence in the presence of a smooth trend," *Biometrika*, **101**, 121–140.

Wang, A., Wang, K., and Tsung, F. (2014), "Statistical surface monitoring by spatial-structure modeling," *Journal of Quality Technology*, **46**, 359–376.

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., Li, J., and Liu, K. (2019), "Causation-based monitoring and diagnosis for multivariate categorical processes with ordinal information," *IEEE Transactions on Automation Science and Engineering*, **16**, 886–897.

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.

You, L. and Qiu, P. (2019), "An effective method for online disease risk monitoring," *Technometrics*, DOI: 10.1080/00401706.2019.1625813.

Zang, Y., and Qiu, P. (2018), "Phase I monitoring of spatial surface data from 3D printing," *Technometrics*, **60**, 169–180.

Zhang, J., Kang, Y., Yang, Y., and Qiu, P. (2015), "Statistical monitoring of the hand, foot, and mouth disease in China," *Biometrics*, **71**, 841–850.

Zhang, N.F. (1998), "A statistical control chart for stationary process data," *Technometrics*, **40**, 24–38.

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.

Zhou, H., and Lawson, A. B. (2008), "EWMA smoothing and Bayesian spatial modeling for health surveillance," *Statistics in Medicine*, **27**, 5907–5928.