Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

# Effective Disease Surveillance By Using Covariate Information

Peihua Qiu | Kai Yang*

Department of Biostatistics, University of Florida, Gainesville, Florida, USA

**Correspondence**
Kai Yang, Department of Biostatistics, University of Florida, Gainesville, FL 32610, USA.
Email: yklmy1994121@ufl.edu

**Abstract**

Effective surveillance of infectious diseases, cancers and other deadly diseases is critically important for public health and safety of our society. Incidence data of such diseases are often collected spatially from different clinics and hospitals through a regional, national or global disease reporting system. In such a system, new batches of data keep being collected over time, and a decision needs to be made immediately after new data are collected regarding whether there is a disease outbreak at the current time point. This is the disease surveillance problem that will be focused in this paper. There are some existing methods for solving this problem, most of which use the disease incidence data only. In practice, however, disease incidence is often associated with some covariates, including the air temperature, humidity, and other weather or environmental conditions. In this paper, we develop a new methodology for disease surveillance which can make use of helpful covariate information to improve its effectiveness. A novelty of this new method is behind the property that only those covariate information that is associated with a true disease outbreak can help trigger a signal. The new method can accommodate seasonality, spatio-temporal data correlation, and nonparametric data distribution. These features make it feasible to use in many real applications.

**KEYWORDS:**
covariates, data correlation, disease surveillance, local smoothing, seasonality, statistical process control

## 1 | INTRODUCTION

Infectious diseases, cancers and other deadly diseases have become a major threat to the public health and safety of our society. Effective disease surveillance is thus a critically important research problem for minimizing the damage of disease outbreaks. To this end, some national and regional disease reporting systems have been developed (Chen et al[1]). However, due to the complexity of the disease surveillance problem, few existing methods can handle it effectively (Shmueli and Burkom[2]). This paper aims to make another effort by developing a new method for disease surveillance.

Our research is motivated by the influenza-like illness (ILI) data that were collected by the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) of the Florida Department of Health. ILI is a respiratory infection caused by a variety of influenza viruses, defined as severe respiratory illness with fever ($> 100^o$ F), cough, sore throat, and difficulty in breathing. It is estimated that 15-40% people in US develop illness from influenza each year, among which about 36,000 people die and about 114,000 people have to be admitted to hospital (Fiore et al[3]). A traditional method to estimate the incidence rate of ILI infection is to carry out repeated seroprevalence surveys that are resource-intensive and slow. Thus, it is unfeasible for early detection of disease outbreaks. To overcome this limitation, Florida Department of Health built ESSENCE, which is a syndromic surveillance system for collecting near real-time pre-diagnostic data from 229 participating emergency departments and 35 urgent care centers that are distributed in all counties of Florida. Currently, the data are updated once a day, and can be accessed by researchers after a proper registration. Such disease incidence data are spatio-temporal (ST) and often have complex data structures. The complexity of the data is usually reflected in: i) complex spatial data variation across different spatial locations, ii) dynamic temporal data variation (e.g., seasonality) even in cases with no disease outbreaks, iii) ST data correlation that is often difficult to describe and estimate, and iv) possible impact of covariates (e.g., weather conditions). Thus, it is challenging to analyze them properly.

In the literature, there have been some existing methods for disease surveillance. Some early disease surveillance methods in the statistical and epidemiological literatures include the Knox, local Knox, and $k$-nearest neighbor methods (e.g., Jacquez,[4] Knox and Bartlett,[5] Kulldorff and Hjalmars[6]). They aim to detect disease outbreaks by identifying irregular space-time patterns in the observed data. However, these methods are all retrospective, and not designed for prospective disease surveillance. Another type of methods are based on the spatial or spatio-temporal scan statistics (e.g., Kulldorff,[7] Takahashi et al,[8] Woodall et al[9]). They try to identify spatial or spatio-temporal disease clusters by testing whether the observed disease counts in different windows of circular or other more flexible shapes are significantly higher than the expected disease counts under the null hypothesis of no disease outbreaks. These methods are based on the assumption that the observed disease incidence data have a parametric distribution, such as the normal, Poisson, or negative binomial distribution, which is rarely valid in practice (Zhang et al[10,11]). Because disease surveillance is a sequential process monitoring problem and statistical process control (SPC) charts provide a major statistical tool for sequential process monitoring, there have been some existing discussions about disease surveillance using the SPC charts (cf., Dassanayake and French,[12] Dong et al,[13] Yang and Qiu[14]). By using a SPC chart, we can sequentially monitor the observed disease incidence data and give a signal each time when a shift from an in-control (IC) status (i.e., the status without any disease outbreaks) to an out-of-control (OC) status (i.e., the status with a disease outbreak) is detected. In that direction, a cumulative sum (CUSUM) chart for monitoring observed counts of disease occurrence in multiple spatial regions was suggested in Dassanayake and French,[12] and an exponentially weighted moving average (EWMA) chart that can accommodate time-varying population sizes was developed in Dong et al.[13] These methods assume that i) the observed disease incidence data have either a normal or Poisson distribution, ii) they are independent of each other at different time points, and iii) their IC distribution does not change over time. All these assumptions could be violated in practice. To overcome these limitations, Yang and Qiu[14] suggested a CUSUM chart for disease surveillance, which can accommodate the dynamic nature of the observed disease incidence rates (e.g., seasonality), spatio-temporal data correlation, and nonparametric data distribution. All these SPC charts are based on the observed disease incidence data only, and they did not use any information in the related covariates. Thus, there should be room for improvement.

Recently, Yang and Qiu[15] suggested an EWMA chart for monitoring univariate sequential processes, which can accommodate helpful information in covariates. A major feature of this method is that the covariate information is used in choosing the weighting parameter of the EWMA chart only: it is chosen large when the covariates tend to have a shift and small otherwise. Because of this feature, the resulting EWMA chart can react to a future shift in the related univariate process quickly in cases when such a shift is mainly due to the covariates. On the other hand, because the covariate information is used in choosing the weighting parameter only and the EWMA charting statistic is a weighted average of observations of the related process performance variable, the chart is sensitive to shifts in the process performance variable only and would not react to any shifts related to the covariates if such shifts do not result in any shifts in the process performance variable. These properties should be relevant to disease surveillance, since disease incidence data are usually associated with air temperature, humidity, and other weather or environmental conditions. Thus, such covariate information should be taken into account during disease surveillance to make the related methods more effective. On the other hand, the related disease surveillance methods should be robust to shifts in the covariates if such shifts do not result in any disease outbreaks.

However, the method in Yang and Qiu[15], called YQ method hereafter, cannot be applied to disease surveillance directly for the following reasons. First, the YQ method is designed for monitoring univariate processes only, while the disease surveillance problem discussed here is for monitoring spatio-temporal data which often involve spatio-temporal data correlation and other complicated data structure, as discussed above. Second, the YQ method is for the conventional process monitoring problem in which the IC process distribution is assumed unchanged over time. In the disease surveillance problem, however, the distribution of the disease incidence data could change over time (e.g., seasonality), even in cases with no disease outbreaks. In that sense, the disease incidence process is dynamic in nature. Third, a regression model between the process performance variable and some covariates needs to be built in order to use the YQ method, which is quite straightforward in cases when all variables are univariate. In the disease surveillance problem, some covariates are spatio-temporal, and it is quite challenging to build a regression relationship between the spatio-temporal disease incidence data and the spatio-temporal covariates. In this paper, we develop a disease surveillance method based on the idea of the YQ method. The new method has all the favorable properties of the YQ method mentioned above, while it can accommodate the complicated structure of the spatio-temporal disease incidence data. Numerical studies will show that it performs well in various different cases.

The remainder of the article is organized as follows. Section 2 will describe the proposed method in detail. Section 3 will study its numerical performance by presenting some simulation results. Section 4 will discuss its application to the Florida ILI data. Finally, Section 5 will provide some concluding remarks.

## 2 | PROPOSED DISEASE SURVEILLANCE METHOD

Our proposed new method consists of several steps that will be described in two subsections below. In Subsection 2.1, a semiparametric spatio-temporal model will be fitted from an IC data to estimate the IC spatio-temporal pattern of the disease incidence rates and build a functional relationship between the disease incidence rates and the related covariates. In Subsection 2.2, a novel EWMA chart will be developed for effective disease surveillance, in which useful covariate information will be accommodated properly.

### 2.1 | Estimation of the IC spatio-temporal pattern

### 2.1.1 | IC spatio-temporal model and its estimation

Let $\{y(t_i, \boldsymbol{s}_{ij}), j = 1, \ldots, m_i, i = 1, \ldots, n\}$ be the observed disease incidence rates in the IC data, where $t_i \in [0, T]$ is the $i$th observation time, $\boldsymbol{s}_{ij} \in \Omega$ is the $j$th observation location at time $t_i$, $m_i$ is the number of observation locations at $t_i$, and $n$ is the number of observation times. These observed disease incidence rates are assumed to follow the model

$$y(t_i, \boldsymbol{s}_{ij}) = \mu(t_i, \boldsymbol{s}_{ij}) + \mathbf{X}_1^T(t_i)\boldsymbol{\beta}_1 + \mathbf{X}_2^T(t_i, \boldsymbol{s}_{ij})\boldsymbol{\beta}_2 + \varepsilon(t_i, \boldsymbol{s}_{ij}), \quad \text{for } j = 1, \ldots, m_i, i = 1, \ldots, n, \tag{1}$$

where $\mathbf{X}_1(t)$ is a vector of $p_1$ time-dependent covariates, $\mathbf{X}_2(t, \boldsymbol{s})$ is a vector of $p_2$ space/time-dependent covariates, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are their regression coefficients, $\mu(t, \boldsymbol{s})$ is the mean of $y(t, \boldsymbol{s})$ after the part explained by $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t, \boldsymbol{s})$ being excluded, and $\varepsilon(t, \boldsymbol{s})$ is a zero-mean random error, for any $(t, \boldsymbol{s}) \in [0, T] \times \Omega$. The covariance function of $y(t, \boldsymbol{s})$ is denoted as

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

For convenience, $V_y(t, t; \boldsymbol{s}, \boldsymbol{s})$ is also denoted as $\sigma_y^2(t, \boldsymbol{s})$, for any $(t, \boldsymbol{s}) \in [0, T] \times \Omega$. In model (1), no parametric forms are imposed on $\mu(t, \boldsymbol{s})$ and $V_y(t, t'; \boldsymbol{s}, \boldsymbol{s}')$. Thus, it is quite general. In practice, besides time-dependent and space/time-dependent covariates, there could be covariates that do not depend on time (they could depend on space) but are associated with the disease incidence rates. Such covariates, however, would not provide any information about the temporal variation of the disease incidence rates. So, they are not included explicitly in model (1).

Model (1) can be regarded as a semiparametric model. To estimate a semiparametric model, it is natural to consider an iterative estimation procedure (e.g., Speckman[16]), in which the nonparametric and parametric parts can be estimated iteratively. To this end, let us first assume that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T = \mathbf{0}$ in model (1). Then, the resulting model becomes the nonparametric spatio-temporal regression model considered in Yang and Qiu[17], and $\mu(t, \boldsymbol{s})$ can be estimated by the following local linear kernel smoothing (LLKS) procedure:

$$\begin{aligned} \underset{\boldsymbol{\theta} \in \mathbb{R}^4}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{m_i} & \left[y(t_i, \boldsymbol{s}_{ij}) - \theta_\mu - \theta_t(t_i - t) - \theta_u(s_{u,ij} - s_u) - \theta_v(s_{v,ij} - s_v)\right]^2 \\ & \times K_1\left((t_i - t)/h_t\right) K_2\left(d_E(\boldsymbol{s}_{ij}, \boldsymbol{s})/h_s\right), \end{aligned} \tag{2}$$

where $\boldsymbol{\theta} = (\theta_\mu, \theta_t, \theta_u, \theta_v)^T$, $\boldsymbol{s} = (s_u, s_v)^T$, $\boldsymbol{s}_{ij} = (s_{u,ij}, s_{v,ij})^T$, $h_t, h_s > 0$ are two bandwidths, $K_1(\cdot)$ and $K_2(\cdot)$ 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 $\boldsymbol{s}$. Let $\mathbf{G}_{ij} = (1, (t_i - t), (s_{u,ij} - s_u), (s_{v,ij} - s_v))^T$ and $w_{ij} = K_1((t_i - t)/h_t)K_2\left(d_E(\boldsymbol{s}_{ij}, \boldsymbol{s})/h_s\right)$, for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$. Then, the solution of (2) to $\theta_\mu$ is defined to be the LLKS estimate of $\mu(t, \boldsymbol{s})$, which has the expression

$$\widehat{\mu}(t, \boldsymbol{s}) = \boldsymbol{e}_1^T \left(\mathbf{G}^T \mathbf{W} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{W} \mathbf{Y}, \tag{3}$$

where $\boldsymbol{e}_1 = (1,0,0,0)^T$, $\mathbf{G} = (\mathbf{G}_{11},\ldots,\mathbf{G}_{nm_n})^T$, $\mathbf{W} = \mathrm{diag}\{w_{11},\ldots,w_{nm_n}\}$, and $\mathbf{Y} = (y(t_1,\boldsymbol{s}_{11}), \ldots, y(t_n,\boldsymbol{s}_{nm_n}))^T$. From (3), it can be seen that the LLKS estimate $\widehat{\mu}(t,\boldsymbol{s})$ is a weighted average of all observations in a neighborhood of $(t,\boldsymbol{s})$, with the weights determined by the two kernel functions and the neighborhood size controlled by the two bandwidths. The entire iterative estimation procedure is described below.

Iterative Algorithm for Estimating Model (1)

1) Set $\boldsymbol{\beta} = \mathbf{0}$ in Model (1) and obtain an initial estimate of $\mu(t,\boldsymbol{s})$ by (3), denoted as $\widehat{\mu}^{(0)}(t,\boldsymbol{s})$.

2) In the $k$th iteration, for $k \geq 1$, implement the following two steps:

   a) Compute the least squares estimate of $\boldsymbol{\beta}$, denoted as $\widehat{\boldsymbol{\beta}}^{(k)}$, from the linear model $Z^{(k)}(t,\boldsymbol{s}) = \mathbf{X}_1^T(t)\boldsymbol{\beta}_1 + \mathbf{X}_2^T(t,\boldsymbol{s})\boldsymbol{\beta}_2 + \varepsilon(t,\boldsymbol{s})$, where $Z^{(k)}(t,\boldsymbol{s}) = y(t,\boldsymbol{s}) - \widehat{\mu}^{(k-1)}(t,\boldsymbol{s})$.

   b) Update the estimate of $\mu(t,\boldsymbol{s})$ by replacing $\mathbf{Y}$ in (3) by $\mathbf{Y}^{(k)} = (y^{(k)}(t_1,\boldsymbol{s}_{11}),\ldots,y^{(k)}(t_n,\boldsymbol{s}_{nm_n}))^T$, where $y^{(k)}(t_i,\boldsymbol{s}_{ij}) = y(t_i,\boldsymbol{s}_{ij}) - \mathbf{X}_1^T(t_i)\widehat{\boldsymbol{\beta}}_1^{(k)} - \mathbf{X}_2^T(t_i,\boldsymbol{s}_{ij})\widehat{\boldsymbol{\beta}}_2^{(k)}$, for $j = 1,\ldots,m_i$ and $i = 1,\ldots,n$. The updated estimate is denoted as $\widehat{\mu}^{(k)}(t,\boldsymbol{s})$.

3) The iterative algorithm stops when $\|\widehat{\boldsymbol{\beta}}^{(k)} - \widehat{\boldsymbol{\beta}}^{(k-1)}\|_1/\|\widehat{\boldsymbol{\beta}}^{(k-1)}\|_1 \leq \varsigma$, where $\varsigma > 0$ is a pre-specified small number and $\|\boldsymbol{a}\|_1$ denotes the summation of the absolute values of all elements in the vector $\boldsymbol{a}$. Then, $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\mu}^{(k)}(t,\boldsymbol{s})$ are the final estimates of $\boldsymbol{\beta}$ and $\mu(t,\boldsymbol{s})$, respectively. These final estimates are also denoted as $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mu}(t,\boldsymbol{s})$.

The covariance function $V_y(t,t';\boldsymbol{s},\boldsymbol{s}')$ and the variance function $\sigma_y^2(t,\boldsymbol{s})$ can be estimated by moment estimation discussed in Yang and Qiu,[18] which is briefly described below. For any $(t,\boldsymbol{s}),(t',\boldsymbol{s}') \in [0,T] \times \Omega$, let $w_\sigma(i,j;t,\boldsymbol{s}) = K_1\left((t_i - t)/h_t\right) K_2\left(d_E(\boldsymbol{s}_{ij},\boldsymbol{s})/h_s\right)$ 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 $1 \leq j \leq m_i$, $1 \leq l \leq m_k$, and $1 \leq i,k \leq n$. Then, when $(t,\boldsymbol{s}) \neq (t',\boldsymbol{s}')$, $V_y(t,t';\boldsymbol{s},\boldsymbol{s}')$ can be estimated by

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

where $\widehat{\varepsilon}_y(t_i,\boldsymbol{s}_{ij}) = y(t_i,\boldsymbol{s}_{ij}) - \widehat{\mu}_y(t_i,\boldsymbol{s}_{ij})$, $\widehat{\mu}_y(t_i,\boldsymbol{s}_{ij}) = \widehat{\mu}(t_i,\boldsymbol{s}_{ij}) + \widehat{\mu}_z(t_i,\boldsymbol{s}_{ij})$ is the estimate of $E\left(y(t_i,\boldsymbol{s}_{ij})\right)$, and $\widehat{\mu}_z(t,\boldsymbol{s})$ is the estimated mean function of $\widehat{z}(t,\boldsymbol{s}) = \mathbf{X}_1^T(t)\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_2^T(t,\boldsymbol{s})\widehat{\boldsymbol{\beta}}_2$ that will be described later in this Subsection. When $(t,\boldsymbol{s}) = (t',\boldsymbol{s}')$, the variance function $\sigma_y^2(t,\boldsymbol{s}) = V_y(t,t;\boldsymbol{s},\boldsymbol{s})$ can be estimated by

$$\widehat{\sigma}_y^2(t,\boldsymbol{s}) = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{\varepsilon}_y^2(t_i,\boldsymbol{s}_{ij})w_\sigma(i,j;t,\boldsymbol{s})}{\sum_{i=1}^n \sum_{j=1}^{m_i} w_\sigma(i,j;t,\boldsymbol{s})}. \tag{5}$$

It should be pointed out that the estimate $\widehat{V}_y(t,t';\boldsymbol{s},\boldsymbol{s}')$ defined in (4) and (5) may not be positive semidefinite to become a legitimate covariance function. To address this issue, we suggest using the projection-based modification procedure discussed in Yang and Qiu[18] to make the estimate positive semidefinite.

The bandwidths used for estimating $\mu(t,\boldsymbol{s})$ and $\boldsymbol{\beta}$ by the iterative algorithm and those for estimating $V_y(t,t';\boldsymbol{s},\boldsymbol{s}')$ and $\sigma_y^2(t,\boldsymbol{s})$ by (4) and (5) do not need to be the same. As a matter of fact, it has been shown in the literature that they should be chosen differently for estimating the mean components and the variance/covariance components (cf., Yang and Qiu[18]). So, in this paper, they are also allowed to be different for the two different purposes. They are denoted as $(h_{t,1},h_{s,1})$ and $(h_{t,2},h_{s,2})$, respectively, for the bandwidths used in the iterative algorithm and the estimation procedure (4)-(5).

In the model estimation procedure described above, the two kernel functions $K_1(u)$ and $K_2(u)$ can both be chosen to be the Epanechnikov function $K_e(u) = 0.75(1-u^2)I(|u| \leq 1)$, because of its good theoretical properties (cf., Epanechnikov[19]). Because the observed data in model (1) are spatio-temporally correlated, the leave-one-out cross-validation (CV) procedure would not perform well for selecting the bandwidths $(h_{t,1},h_{s,1})$ and $(h_{t,2},h_{s,2})$, since it cannot distinguish the mean structure from the data correlation structure properly in such cases (cf., Altman,[20] Opsomer et al[21]). In this paper, we suggest using a modified CV (MCV) score that was adapted from the version by Brabanter et al[22] in the univariate regression setup to select $(h_{t,1},h_{s,1})$, and a spatio-temporal prediction error (PE) score to select $(h_{t,2},h_{s,2})$. Both the MCV score and the PE score are described in detail in Section A of the supplementary file.

### 2.1.2 | Estimation of the IC covariate effect

From model (1), the covariates $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t,\boldsymbol{s})$ affect the disease incidence rate $y(t,\boldsymbol{s})$ through $z(t,\boldsymbol{s}) = \mathbf{X}_1^T(t)\boldsymbol{\beta}_1 + \mathbf{X}_2^T(t,\boldsymbol{s})\boldsymbol{\beta}_2$. Let $\hat{z}(t,\boldsymbol{s}) = \mathbf{X}_1^T(t)\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2^T(t,\boldsymbol{s})\hat{\boldsymbol{\beta}}_2$, where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T)^T$ is obtained from the iterative algorithm described above. Then, $\hat{z}(t,\boldsymbol{s})$ should be a reasonable estimate of $z(t,\boldsymbol{s})$. For any $(t,\boldsymbol{s}), (t',\boldsymbol{s}') \in [0,T] \times \Omega$, define

$$\mu_z(t,\boldsymbol{s}) = E\left(\hat{z}(t,\boldsymbol{s})\right), \quad V_z(t,t';\boldsymbol{s},\boldsymbol{s}') = \text{Cov}\left(\hat{z}(t,\boldsymbol{s}), \hat{z}(t',\boldsymbol{s}')\right)$$

to be the mean and covariance functions of $\hat{z}(t,\boldsymbol{s})$. For simplicity, let $\sigma_z^2(t,\boldsymbol{s}) = V_z(t,t;\boldsymbol{s},\boldsymbol{s})$. Next, we discuss how to estimate $\mu_z(t,\boldsymbol{s})$ and $V_z(t,t';\boldsymbol{s},\boldsymbol{s}')$ from the observed data $\{\hat{z}(t_i,\boldsymbol{s}_{ij}), j = 1,\dots,m_i, i = 1,\dots,n\}$.

The mean function $\mu_z(t,\boldsymbol{s})$ can be estimated by the LLKS estimate, denoted as $\hat{\mu}_z(t,\boldsymbol{s})$, obtained from (3) after $\mathbf{Y}$ is replaced by $\hat{\mathbf{Z}} = (\hat{z}(t_1,\boldsymbol{s}_{11}),\dots,\hat{z}(t_n,\boldsymbol{s}_{nm_n}))^T$. The bandwidths used here, denoted as $(h_{t,3}, h_{s,3})$, could be different from $(h_{t,1}, h_{s,1})$ that are used when estimating $\boldsymbol{\beta}$ and $\mu(t,\boldsymbol{s})$. But, $(h_{t,3}, h_{s,3})$ can still be chosen by minimizing the MCV score defined in (A.1) in the supplementary file, after $y(t_i,\boldsymbol{s}_{ij})$ and $\hat{\mu}_{-(ij)}(t_i,\boldsymbol{s}_{ij})$ are replaced by $\hat{z}(t_i,\boldsymbol{s}_{ij})$ and $\hat{\mu}_{z,-(ij)}(t_i,\boldsymbol{s}_{ij})$, respectively, where $\hat{\mu}_{z,-(ij)}(t_i,\boldsymbol{s}_{ij})$ denotes the LLKS estimate of $\mu_z(t_i,\boldsymbol{s}_{ij})$ without using the $(i,j)$th observation $\hat{z}(t_i,\boldsymbol{s}_{ij})$. The covariance and variance functions $V_z(t,t';\boldsymbol{s},\boldsymbol{s}')$ and $\sigma_z^2(t,\boldsymbol{s})$ can still be estimated by (4) and (5), except that the quantities $\{\hat{\varepsilon}_y(t_i,\boldsymbol{s}_{ij})\}$ need to be replaced by $\{\hat{\varepsilon}_z(t_i,\boldsymbol{s}_{ij}) = \hat{z}(t_i,\boldsymbol{s}_{ij}) - \hat{\mu}_z(t_i,\boldsymbol{s}_{ij})\}$. Of course, the bandwidths used here, denoted as $(h_{t,4}, h_{s,4})$, can also be different from $(h_{t,2}, h_{s,2})$ used in (4) and (5). But, they can still be selected by minimizing the PE score defined in (A.3) of the supplementary file, after $y(t_i,\boldsymbol{s}_{ij})$ and $\hat{y}_{-(ij)}(t_i,\boldsymbol{s}_{ij})$ are replaced by $\hat{z}(t_i,\boldsymbol{s}_{ij})$ and $\hat{z}_{-(ij)}(t_i,\boldsymbol{s}_{ij})$, respectively, where $\hat{z}_{-(ij)}(t_i,\boldsymbol{s}_{ij})$ denotes the predicted value of $\hat{z}(t_i,\boldsymbol{s}_{ij})$ by the kriging method, defined similarly to that in (A.4) of the supplementary file.

### 2.1.3 | Statistical properties of the estimates

So far, we have discussed estimation of the IC spatio-temporal pattern. Under some mild conditions, the uniform convergence of the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\mu}(t,\boldsymbol{s})$, $\hat{\mu}_z(t,\boldsymbol{s})$, $\hat{\mu}_y(t,\boldsymbol{s}) = \hat{\mu}(t,\boldsymbol{s}) + \hat{\mu}_z(t,\boldsymbol{s})$, $\hat{V}_y(t,t';\boldsymbol{s},\boldsymbol{s}')$, and $\hat{V}_z(t,t';\boldsymbol{s},\boldsymbol{s}')$ can be established, which is presented in Theorem 1 in the appendix.

## 2.2 | Spatio-temporal disease surveillance by using covariate information

### 2.2.1 | Construction of the proposed method for spatio-temporal disease surveillance

After the IC spatio-temporal pattern of the disease incidence rates is estimated from an IC data, we are ready to describe our proposed method for disease surveillance. A main feature of the new method is that it makes use of helpful covariate information when detecting disease outbreaks, but its signal can only be triggered by unusual spatio-temporal pattern of the observed disease incidence rates. To be more specific, assume that the disease incidence rates to monitor at time $t_i^* \in (T,\infty)$, for $i = 1,2,\dots$, are observed at spatial locations $\{\boldsymbol{s}_{ij}^* \in \Omega, j = 1,\dots,m_i^*\}$. These observations are denoted as $\{y(t_i^*,\boldsymbol{s}_{ij}^*)\}$, and the related observations of the time-dependent and space/time-dependent covariates are denoted as $\{\mathbf{X}_1(t_i^*)\}$, and $\{\mathbf{X}_2(t_i^*,\boldsymbol{s}_{ij}^*)\}$, respectively. To detect disease outbreaks, let us first consider the following standardized residuals:

$$\hat{e}_y(t_i^*,\boldsymbol{s}_{ij}^*) = \frac{y(t_i^*,\boldsymbol{s}_{ij}^*) - \hat{\mu}_y(t_i^*,\boldsymbol{s}_{ij}^*)}{\hat{\sigma}_y(t_i^*,\boldsymbol{s}_{ij}^*)}, \quad \text{for } j = 1,\dots,m_i^*, \ \ i = 1,2,\dots, \tag{6}$$

where $\hat{\sigma}_y(t_i^*,\boldsymbol{s}_{ij}^*) = \sqrt{\hat{V}_y(t_i^*,t_i^*;\boldsymbol{s}_{ij}^*,\boldsymbol{s}_{ij}^*)}$, and $\hat{\mu}_y(t_i^*,\boldsymbol{s}_{ij}^*)$ and $\hat{V}_y(t_i^*,t_i^*;\boldsymbol{s}_{ij}^*,\boldsymbol{s}_{ij}^*)$ are obtained from the IC data (cf., Subsection 2.1). It should be noticed that the original estimates $\hat{\mu}_y(t,\boldsymbol{s})$ and $\hat{V}_y(t,t;\boldsymbol{s},\boldsymbol{s})$ obtained from the IC data are defined in the time internal $[0,T]$ and the spatial domain $\Omega$. However, the observation times $\{t_i^*, i = 1,2,\dots\}$ in Expression (6) are in the time interval $(T,\infty)$. To make this expression well defined, the estimates $\hat{\mu}_y(t,\boldsymbol{s})$ and $\hat{V}_y(t,t;\boldsymbol{s},\boldsymbol{s})$ should be extended in the time domain periodically from $[0,T]$ to $[0,\infty)$ with the period of $T$ in advance. For instance, if $t_i^* = \tilde{t}_i^* + lT$, where $\tilde{t}_i^* \in [0,T]$ and $l \geq 1$ is an integer, then we define $\hat{\mu}_y(t_i^*,\boldsymbol{s}_{ij}^*) = \hat{\mu}_y(\tilde{t}_i^*,\boldsymbol{s}_{ij}^*)$, for any $\boldsymbol{s}_{ij}^* \in \Omega$. In (6), the observed spatio-temporal pattern of the disease incidence rates has been compared to the estimated IC spatio-temporal pattern described by $\hat{\mu}_y(t_i^*,\boldsymbol{s}_{ij}^*)$ and $\hat{\sigma}_y(t_i^*,\boldsymbol{s}_{ij}^*)$. So, $\{\hat{e}_y(t_i^*,\boldsymbol{s}_{ij}^*)\}$ can be used for detecting disease outbreaks: the larger their values, the more likely a disease outbreak.

From model (1), the covariates $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t,\boldsymbol{s})$ affect the disease incidence rate $y(t,\boldsymbol{s})$ through $z(t,\boldsymbol{s}) = \mathbf{X}_1(t)^T\boldsymbol{\beta}_1 + \mathbf{X}_2(t,\boldsymbol{s})^T\boldsymbol{\beta}_2$. A shift in $z(t,\boldsymbol{s})$ could result in a shift in $y(t,\boldsymbol{s})$. Similar to (6), the following standardized residuals could be useful for detecting a shift in $z(t, \boldsymbol{s})$:

$$\hat{e}_z(t_i^*, \boldsymbol{s}_{ij}^*) = \frac{\hat{z}(t_i^*, \boldsymbol{s}_{ij}^*) - \hat{\mu}_z(t_i^*, \boldsymbol{s}_{ij}^*)}{\hat{\sigma}_z(t_i^*, \boldsymbol{s}_{ij}^*)}, \quad \text{for } j = 1, \ldots, m_i^*, \quad i = 1, 2, \ldots, \tag{7}$$

where $\hat{z}(t_i^*, \boldsymbol{s}_{ij}^*) = \mathbf{X}_1(t_i^*)^T \hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2(t_i^*, \boldsymbol{s}_{ij}^*)^T \hat{\boldsymbol{\beta}}_2$, $\hat{\sigma}_z(t_i^*, \boldsymbol{s}_{ij}^*) = \sqrt{\hat{V}_z(t_i^*, t_i^*; \boldsymbol{s}_{ij}^*, \boldsymbol{s}_{ij}^*)}$, and $\hat{\mu}_z(t_i^*, \boldsymbol{s}_{ij}^*)$ and $\hat{V}_z(t_i^*, t_i^*; \boldsymbol{s}_{ij}^*, \boldsymbol{s}_{ij}^*)$ are computed from the IC data and have been extended in the time domain periodically from $[0, T]$ to $[0, \infty)$ with the period of $T$. Next, the standardized residuals at different observation locations are combined at each observation time, so that a univariate control chart can be used for detecting shifts in $z(t, \boldsymbol{s})$. To this end, because of the spatial data correlation among $\{\hat{e}_z(t_i^*, \boldsymbol{s}_{ij}^*), j = 1, \ldots, m_i^*\}$ defined in (7), we first decorrelate them by defining $\tilde{\mathbf{e}}_z(t_i^*) = \hat{C}_z(t_i^*)^{-1/2} \hat{\mathbf{e}}_z(t_i^*)$, where $\hat{\mathbf{e}}_z(t_i^*) = (\hat{e}_z(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \hat{e}_z(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T$, and $\hat{C}_z(t_i^*)$ is the estimated correlation matrix of $\hat{\mathbf{e}}_z(t_i^*)$ computed from $\hat{V}_z(t_i^*, t_i^*; \boldsymbol{s}, \boldsymbol{s}')$, for $\boldsymbol{s}, \boldsymbol{s}' \in \Omega$. It can be checked that the elements of $\tilde{\mathbf{e}}_z(t_i^*)$, denoted as $\tilde{e}_z(t_i^*, \boldsymbol{s}_{ij}^*)$, for $j = 1, \ldots, m_i^*$, are asymptotically uncorrelated with the asymptotic mean of 0 and the asymptotic variance of 1 when there are no disease outbreaks by the time $t_i^*$. Then, the following EWMA charting statistic (Roberts[23]) is considered:

$$E_{z,i} = \lambda \check{e}_z(t_i^*) + (1 - \lambda) E_{z,i-1}, \quad \text{for } i \geq 1, \tag{8}$$

where $E_{z,0} = 0$, $\lambda \in (0, 1]$ is a weighting parameter, and $\check{e}_z(t_i^*) = \sum_{j=1}^{m_i^*} \tilde{e}_z(t_i^*, \boldsymbol{s}_{ij}^*) / \sqrt{m_i^*}$, for each $i$. If there is an upward mean shift in $z(t, \boldsymbol{s})$ at or before the time $t_i^*$, then the value of $E_{z,i}$ would be relatively large because of the shift (cf., Chapter 5, Qiu[24]). Therefore, the EWMA charting statistic $E_{z,i}$ provides a measure of the likelihood of an upward mean shift in $z(t, \boldsymbol{s})$.

In the current spatio-temporal disease surveillance problem, our ultimate goal is to detect shifts in the disease incidence rate $y(t, \boldsymbol{s})$, which may or may not be caused by shifts in $z(t, \boldsymbol{s})$. In addition, shifts in $z(t, \boldsymbol{s})$ are not our major concern in the disease surveillance problem, although any helpful information in $z(t, \boldsymbol{s})$ should be used in disease surveillance. By these considerations and the idea in Yang and Qiu[15] to use covariate information during online process monitoring, the following EWMA charting statistic is suggested for disease surveillance: for $i \geq 1$,

$$E_{y,i} = \mathrm{W}(E_{z,i}; \lambda, \kappa)\, \check{e}_y(t_i^*) + \left[1 - \mathrm{W}(E_{z,i}; \lambda, \kappa)\right] E_{y,i-1}, \tag{9}$$

where $E_{y,0} = 0$, and $\mathrm{W}(E_{z,i}; \lambda, \kappa) \in (0, 1]$ is a weighting parameter for $\check{e}_y(t_i^*)$ that depends on the covariate charting statistic $E_{z,i}$ and two parameters $\lambda \in (0, 1]$ and $\kappa > 0$. In (9), the quantity $\check{e}_y(t_i^*)$ is defined similarly to $\check{e}_z(t_i^*)$, as

$$\check{e}_y(t_i^*) = \sum_{j=1}^{m_i^*} \tilde{e}_y(t_i^*, \boldsymbol{s}_{ij}^*) / \sqrt{m_i^*},$$

where $\tilde{\mathbf{e}}_y(t_i^*) = (\tilde{e}_y(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \tilde{e}_y(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T = \hat{C}_y(t_i^*)^{-1/2} \hat{\mathbf{e}}_y(t_i^*)$, $\hat{\mathbf{e}}_y(t_i^*) = (\hat{e}_y(t_i^*, \boldsymbol{s}_{i1}^*), \ldots, \hat{e}_y(t_i^*, \boldsymbol{s}_{im_i^*}^*))^T$, and $\hat{C}_y(t_i^*)$ is the estimated correlation matrix of $\hat{\mathbf{e}}_y(t_i^*)$ obtained from the estimate $\hat{V}_y(t_i^*, t_i^*; \boldsymbol{s}, \boldsymbol{s}')$. Then, the chart gives a signal of disease outbreak at time $t_i^*$ if $E_{y,i} > L$, where $L > 0$ is a control limit. This chart is denoted as NEW hereafter.

In the SPC literature (cf., Qiu[24]), the performance of a chart like (9) is usually evaluated by the IC average run length (ARL), defined as the average number of observation time points from the beginning of process monitoring to a signal from the chart when the process is IC, and the OC ARL, defined as the average number of observation time points from the occurrence of a shift in the process distribution to a signal time of the chart. The IC and OC ARL values are denoted as $\text{ARL}_0$ and $\text{ARL}_1$, respectively. Regarding the design of the chart, its $\text{ARL}_0$ value is usually fixed at a given level, and its parameters are chosen such that the specified $\text{ARL}_0$ value is reached. Then, the chart performs better if its $\text{ARL}_1$ value is smaller for detecting a shift of a given size.

From its construction, it can be seen that the EWMA charting statistic $E_{y,i}$ in (9) is a weighted average of the observed disease incidence rates. Thus, only shifts in the disease incidence rates can trigger a signal of the chart. The covariate information is used in the weighting parameter $\mathrm{W}(E_{z,i}; \lambda, \kappa)$ only, which will be chosen to be an increasing function of $E_{z,i}$. So, when $E_{z,i}$ is larger, or when there is more evidence of a shift in the covariate combination $z(t, \boldsymbol{s})$, the weight $\mathrm{W}(E_{z,i}; \lambda, \kappa)$ will be chosen larger so that the current and several most recent observations will receive more weights. In such cases, the possible shift in the disease incidence rates can be detected more effectively. Because the covariate information is used in the weighting parameter only, a shift in $z(t, \boldsymbol{s})$ would not trigger a signal from the chart (9) if that shift does not result in a shift in the disease incidence rates. Thus, the chart (9) can accomplish the research goal of the paper stated in Section 1 and at the beginning of this subsection.

### 2.2.2 | Determination of the weighting function $\mathrm{W}(u;\lambda,\kappa)$ and the control limit $L$

To use the EWMA chart (9), we need to properly specify the weighting function $\mathrm{W}(u;\lambda,\kappa)$ (as a function of $u$) in advance. As mentioned earlier, this function should be chosen to be an increasing function of $u$. Since it is a weight used in an EWMA chart, its value needs to be in the interval (0, 1]. By taking into account all these considerations, we suggest using the following weighting function that has been recommended in Yang and Qiu[15]:

$$\mathrm{W}(u;\lambda,\kappa) = \begin{cases} \min\{1, \lambda + (u/\kappa - 1)\}, & \text{if } u > \kappa, \\ \lambda, & \text{otherwise.} \end{cases} \tag{10}$$

From (10), $\mathrm{W}(u;\lambda,\kappa)$ is a linear function of $u$ with a lower bound of $\lambda > 0$ and an upper bound of 1. The lower bound $\lambda$ is reached when $u \le \kappa$. So, by using this weighting function in the EWMA chart (9), when $E_{z,i} \le \kappa$, $\mathrm{W}(u;\lambda,\kappa)$ becomes the regular weighting parameter $\lambda$. The parameter $\kappa$ is similar to a control limit for the EWMA charting statistic $E_{z,i}$. So, when $E_{z,i} \le \kappa$, it is unlikely that there is an upward mean shift in $z(t,\boldsymbol{s})$ by the time $t_i^*$. In such cases, it is reasonable to use the regular weighting parameter $\lambda$ in (9). When $E_{z,i} > \kappa$, it is likely that an upward mean shift in $z(t,\boldsymbol{s})$ has occurred at or before the time $t_i^*$. In such cases, the weight $\mathrm{W}(E_{z,i};\lambda,\kappa)$ defined in (10) will be larger than $\lambda$, implying that the observations at $t_i^*$ and a few previous time points will receive more weights, which is intuitively reasonable, as explained before.

In the weighting function $\mathrm{W}(u;\lambda,\kappa)$ defined in (10), there are two parameters $\lambda$ and $\kappa$ involved. As explained above, $\lambda$ is a regular weighting parameter used in an EWMA chart. It can be chosen to be the same as the one used in defining the EWMA charting statistic $E_{z,i}$ in (8). In the SPC literature, commonly used values for $\lambda$ include 0.05, 0.1, 0.2, 0.5 and 1.0. Regarding $\kappa$, it can be selected similarly to a control limit of an EWMA chart. Let the IC ARL value of the EWMA chart with the charting statistic $E_{z,i}$ be $\mathrm{ARL}_{0,z}$. Then, $\kappa$ can be chosen from an IC data $\{(\mathbf{X}_1(t_i'), \mathbf{X}_2(t_i', \boldsymbol{s}_{ij}'), y(t_i', \boldsymbol{s}_{ij}')), j = 1,\ldots,m_i', i = 1,\ldots,n'\}$ by a block bootstrap procedure consisting of the following five steps, where the IC data used here could be different from the one used for estimating the IC model (1).

1) The standardized residuals $\{\check{e}_z(t_i'), i = 1,\ldots,n'\}$ are first computed by (7), and $n' - l + 1$ blocks $\{Q_k, k = 1,\ldots,n'-l+1\}$ can be formed, where $Q_k = \{\check{e}_z(t_i'), i = k,\ldots,k+l-1\}$ and $l$ is a block size.

2) A sequence of blocks can be randomly selected with replacement from $\{Q_k, k = 1,\ldots,n'-l+1\}$, and the selected blocks are placed one after another to form a bootstrap sample, denoted as $\{\check{e}_{z,i}^{(b)}, i \ge 1\}$.

3) For $i \ge 1$, we calculate the EWMA charting statistic from the bootstrap sample, defined as $E_{z,i}^{(b)} = \lambda\check{e}_{z,i}^{(b)} + (1-\lambda)E_{z,i-1}^{(b)}$, and record the IC run length as $\mathrm{RL}_0(\kappa) = \min\{i, E_{z,i}^{(b)} > \kappa\}$, for a given value of $\kappa$.

4) The second and third steps are then repeated for $B$ times, and the average of the $B$ values of $\mathrm{RL}_0(\kappa)$ is used for approximating $\mathrm{ARL}_0(\kappa)$.

5) The bisection search algorithm is used to search for a value of $\kappa$ such that $\mathrm{ARL}_0(\kappa)$ reaches the pre-specified value of $\mathrm{ARL}_{0,z}$.

For the EWMA chart (9), its control limit $L$ can be determined in the same way by the above block bootstrap procedure, once its IC ARL value, denoted as $\mathrm{ARL}_0$, is given and the weighting function $\mathrm{W}(u;\lambda,\kappa)$ is determined.

### 2.2.3 | A modification for detecting upward mean shifts

It should be pointed out that the proposed chart NEW (cf., (9)) may not be effective for detecting an upward mean shift in the disease incidence rate $y(t,\boldsymbol{s})$ in cases when $y(t,\boldsymbol{s})$ has actually downward shifts at some spatial locations. The reason is that the decorrelated and standardized residuals have been averaged across different spatial locations at each time point when the charting statistic $E_{y,i}$ is computed. So, positive and negative residuals at different spatial locations will be canceled out, making the resulting chart ineffective. To overcome this limitation, the following modified version of NEW, denoted as MNEW, is suggested. Let

$$\hat{e}_{y,+}(t_i^*, \boldsymbol{s}_{ij}^*) = \max(\hat{e}_y(t_i^*, \boldsymbol{s}_{ij}^*), 0), \quad \hat{e}_{z,+}(t_i^*, \boldsymbol{s}_{ij}^*) = \max(\hat{e}_z(t_i^*, \boldsymbol{s}_{ij}^*), 0),$$

for $j = 1,\ldots,m_i^*$ and $i = 1,2,\ldots$. Then, the means of these quantities would be non-zero, and their standardized values are defined to be

$$\hat{e}_{y,0}(t_i^*, \boldsymbol{s}_{ij}^*) = \frac{\hat{e}_{y,+}(t_i^*, \boldsymbol{s}_{ij}^*) - \hat{\mu}_{y,+}(t_i^*, \boldsymbol{s}_{ij}^*)}{\hat{\sigma}_{y,+}(t_i^*, \boldsymbol{s}_{ij}^*)}, \quad \hat{e}_{z,0}(t_i^*, \boldsymbol{s}_{ij}^*) = \frac{\hat{e}_{z,+}(t_i^*, \boldsymbol{s}_{ij}^*) - \hat{\mu}_{z,+}(t_i^*, \boldsymbol{s}_{ij}^*)}{\hat{\sigma}_{z,+}(t_i^*, \boldsymbol{s}_{ij}^*)},$$

where $\widehat{\sigma}_{y,+}(t_i^*, \boldsymbol{s}_{ij}^*) = \sqrt{\widehat{V}_{y,+}(t_i^*, t_i^*; \boldsymbol{s}_{ij}^*, \boldsymbol{s}_{ij}^*)}$, $\widehat{\sigma}_{z,+}(t_i^*, \boldsymbol{s}_{ij}^*) = \sqrt{\widehat{V}_{z,+}(t_i^*, t_i^*; \boldsymbol{s}_{ij}^*, \boldsymbol{s}_{ij}^*)}$, the estimated mean functions $\widehat{\mu}_{y,+}(t, \boldsymbol{s})$ and $\widehat{\mu}_{z,+}(t, \boldsymbol{s})$ are both obtained by the LLKS procedure (3) after $\{y(t_i, \boldsymbol{s}_{ij})\}$ are replaced respectively by $\{\widehat{e}_{y,+}(t_i, \boldsymbol{s}_{ij})\}$ and $\{\widehat{e}_{z,+}(t_i, \boldsymbol{s}_{ij})\}$, and the covariance functions $\widehat{V}_{y,+}(t, t'; \boldsymbol{s}, \boldsymbol{s}')$ and $\widehat{V}_{z,+}(t, t'; \boldsymbol{s}, \boldsymbol{s}')$ are obtained by the estimation procedure (4)-(5) after $\{\widehat{\varepsilon}_y(t_i, \boldsymbol{s}_{ij})\}$ are replaced respectively by $\{\widehat{\varepsilon}_{y,+}(t_i, \boldsymbol{s}_{ij}) = \widehat{e}_{y,+}(t_i, \boldsymbol{s}_{ij}) - \widehat{\mu}_{y,+}(t_i, \boldsymbol{s}_{ij})\}$ and $\{\widehat{\varepsilon}_{z,+}(t_i, \boldsymbol{s}_{ij}) = \widehat{e}_{z,+}(t_i, \boldsymbol{s}_{ij}) - \widehat{\mu}_{z,+}(t_i, \boldsymbol{s}_{ij})\}$. The bandwidths in obtaining $\widehat{\mu}_{y,+}(t, \boldsymbol{s})$ and $\widehat{\mu}_{z,+}(t, \boldsymbol{s})$ can be selected by the MCV score defined in (A.1) of the supplementary file, while the bandwidths for obtaining $\widehat{V}_{y,+}(t, t'; \boldsymbol{s}, \boldsymbol{s}')$ and $\widehat{V}_{z,+}(t, t'; \boldsymbol{s}, \boldsymbol{s}')$ can be chosen by the PE score defined in (A.3) of that file. Then, the charting statistic of MNEW is defined by (9), after the quantities $\{\check{e}_y(t_i^*)\}$ computed from $\{\widehat{e}_y(t_i^*, \boldsymbol{s}_{ij}^*)\}$ are replaced by the corresponding ones computed from $\{\widehat{e}_{y,0}(t_i^*, \boldsymbol{s}_{ij}^*)\}$.

## 3 | SIMULATION STUDIES

In this section, we investigate the numerical performance of the proposed method described in Section 2 using Monte Carlo simulations. Our discussion is organized in four parts. The first three parts focus on the proposed chart MNEW. More specifically, Subsection 3.1 is about its performance in estimating the IC model, Subsection 3.2 is about the impact of the IC sample size and block size on its performance, and Subsection 3.3 is about the impact of the parameters $(\lambda, \mathrm{ARL}_{0,z})$ on its performance. Then, the performance of MNEW is compared to that of NEW and several other competing methods in Subsection 3.4.

Before presenting the simulation results, let us provide a detailed description about the simulation setup. For simplicity, let us assume that $[0, T] = [0, 1]$, the observation times are $\{t_i = i/n, i = 1, \ldots, n\}$, and the observation locations are unchanged over time and equally spaced in $\Omega = [0, 1] \times [0, 1]$. In such cases, the observation locations can simply be denoted as $\{\boldsymbol{s}_j, j = 1, \ldots, m\}$, where $m$ is square of an integer. In this section, $(n, m)$ are chosen to be (200, 64) or (400, 100), unless stated otherwise. The IC model is assumed to be the following one:

$$y(t, \boldsymbol{s}) = \mu(t, \boldsymbol{s}) + \beta_1 X_1(t) + \beta_2 X_2(t, \boldsymbol{s}) + \varepsilon(t, \boldsymbol{s}), \quad \text{for} \quad (t, \boldsymbol{s}) \in [0, 1] \times \Omega,$$

where $X_1(t) = \mu_1(t) + \varepsilon_1(t)$, $X_2(t, \boldsymbol{s}) = \mu_2(t, \boldsymbol{s}) + \varepsilon_2(t, \boldsymbol{s})$, and $\varepsilon_1(t)$, $\varepsilon_2(t, \boldsymbol{s})$ and $\varepsilon(t, \boldsymbol{s})$ are mutually independent zero-mean random errors. In the above model, it is assumed that $\beta_1 = \beta_2 = 0.3$, $\mu_1(t) = 0.01(t - 0.5)^2$, $\mu_2(t, \boldsymbol{s}) = 0.01(t - 0.5)^2 + 0.01\left[(s_u - 0.5)^2 + (s_v - 0.5)^2\right]$, and $\mu(t, \boldsymbol{s}) = 0.01\cos(2\pi t) + 0.01\exp\{-(s_u + s_v)/2\} + 0.02$, where $\boldsymbol{s} = (s_u, s_v)^T$. The space/time-varying mean functions $\mu(t, \boldsymbol{s})$ and $\mu_2(t, \boldsymbol{s})$ are presented in the 1st and 2nd rows of Figure 1(a), respectively, when $t = 0$ (1st column), $t = 0.5$ (2nd column) and $t = 1$ (3rd column), and the time-varying mean function $\mu_1(t)$ is shown in Figure 1(b). The random errors $\{\varepsilon(t_i, \boldsymbol{s}_j)\}$, $\{\varepsilon_1(t_i)\}$ and $\{\varepsilon_2(t_i, \boldsymbol{s}_j)\}$ are generated as follows:

- The quantities $\{\varepsilon_1(t_i)\}$ are generated from the AR(1) model $\varepsilon_1(t_i) = \rho_t \varepsilon_1(t_{i-1}) + (1 - \rho_t^2)^{1/2}\eta_1(t_i)$, where $|\rho_t| < 1$ is a constant and $\{\eta_1(t_i)\}$ are i.i.d. with the common distribution $N(0, 0.006^2)$.

- Let $\boldsymbol{\varepsilon}_2(t_i) = (\varepsilon_2(t_i, \boldsymbol{s}_1), \ldots, \varepsilon_2(t_i, \boldsymbol{s}_m))^T$. Then, $\boldsymbol{\varepsilon}_2(t_i)$ is generated from the $m$-dimensional AR(1) model $\boldsymbol{\varepsilon}_2(t_i) = \rho_t \boldsymbol{\varepsilon}_2(t_{i-1}) + (1 - \rho_t^2)^{1/2}\boldsymbol{\eta}_2(t_i)$, where $\boldsymbol{\eta}_2(t_i) = (\eta_2(t_i, \boldsymbol{s}_1), \ldots, \eta_2(t_i, \boldsymbol{s}_m))^T$ are temporally independent Gaussian spatial processes whose spatial correlation is described by the covariance function $\mathrm{Cov}\left(\eta_2(t_i, \boldsymbol{s}_j), \eta_2(t_i, \boldsymbol{s}_l)\right) = 0.006^2 \exp\{-d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/\rho_s\}$, and $\rho_s > 0$ is a constant.

- 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 in the same way as that for $\boldsymbol{\varepsilon}_2(t_i)$, except that its spatial covariance at each time point is assumed to be $\mathrm{Cov}(\eta(t_i, \boldsymbol{s}_j), \eta(t_i, \boldsymbol{s}_l)) = 0.003^2 \exp\{-d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/\rho_s\}$.

In the above setup, it can be checked that, for any $1 \le i, k \le n$ and $1 \le j, l \le m$, the covariance between $y(t_i, \boldsymbol{s}_j)$ and $y(t_k, \boldsymbol{s}_l)$ is

$$V_y(t_i, t_k; \boldsymbol{s}_j, \boldsymbol{s}_l) = \rho_t^{|k-i|}\left[0.006^2\beta_1^2 + \left(0.006^2\beta_2^2 + 0.003^2\right)\exp\{-d_E(\boldsymbol{s}_j, \boldsymbol{s}_l)/\rho_s\}\right].$$

Thus, the parameters $\rho_t$ and $\rho_s$ control the data correlation in time and space, respectively; the larger their values, the stronger the correlation. To consider cases with different spatio-temporal data correlation, $(\rho_t, \rho_s)$ are chosen to be (0.2, 0.1), (0.4, 0.2) or (0.6, 0.3).

[Figure 1 about here.]

After a disease outbreak, the OC model for the disease incidence rates is assumed to be

$$y^{(\delta)}(t, \boldsymbol{s}) = \mu^{(\delta)}(t, \boldsymbol{s}) + z^{(\delta)}(t, \boldsymbol{s}) + \varepsilon(t, \boldsymbol{s}),$$

where $\mu^{(\delta)}(t,s) = \mu(t,s) + \sigma_y\delta_\mu(t,s)$, $z^{(\delta)}(t,s) = z(t,s) + \sigma_y\delta_z(t,s)$, $\sigma_y = \{0.006^2 \times 2 \times 0.3^2 + 0.003^2\}^{1/2} = 0.0039$ is the IC standard deviation of $y(t,s)$, $z(t,s) = \beta_1 X_1(t) + \beta_2 X_2(t,s)$, $\delta_z(t,s)$ and $\delta_\mu(t,s)$ describe the shift sizes in $y(t,s)$ due to the covariates and other factors that are not included in the model, respectively, and $\mu(t,s)$, $X_1(t)$, $X_2(t,s)$ and $\varepsilon(t,s)$ are the same as those in the IC model. By some simple calculation, it can be checked that the OC mean function of $y(t,s)$ in the above setup is $\mu_y^{(\delta)}(t,s) = \mu_y(t,s) + \sigma_y\{\delta_\mu(t,s) + \delta_z(t,s)\}$, where $\mu_y(t,s)$ is the IC mean function of $y(t,s)$. The following four scenarios are considered about the shift sizes $\delta_\mu(t,s)$ and $\delta_z(t,s)$: for $\nu = 1,2,3,4$,

**(I)** $\delta_\mu(t,s) = 0.20\nu \times \Delta_1(t,s)$, $\delta_z(t,s) = 0$,

**(II)** $\delta_\mu(t,s) = 0.04\nu \times \Delta_1(t,s)$, $\delta_z(t,s) = 0.16\nu \times \Delta_1(t,s)$,

**(III)** $\delta_\mu(t,s) = 0.20\nu \times \Delta_2(t,s)$, $\delta_z(t,s) = 0$, and

**(IV)** $\delta_\mu(t,s) = 0.04\nu \times \Delta_2(t,s)$, $\delta_z(t,s) = 0.16\nu \times \Delta_2(t,s)$,

where $\Delta_1(t,s) = 2(t-0.5)^2 + \exp\left\{-[(s_u-0.5)^2 + (s_v-0.5)^2]\right\}$ which is always positive, $\Delta_2(t,s) = 2(t-0.5)^2 + \exp\{-[(s_u - 0.5)^2 + (s_v-0.5)^2]\}\text{sign}\{|s_u-0.5| + |s_v-0.5| - 0.5\}$ which could be negative in regions close to the center (0.5,0.5) of $\Omega$, and $\text{sign}(\cdot)$ is the sign function. From their construction, it can be seen that the shifts in types (I) and (III) are not due to covariates at all, while those in types (II) and (IV) are due to both covariates and other factors. By comparing the shifts in types (I) and (III), those in type (I) are always positive at all observation times and locations, but those in type (III) could be negative in spatial regions close to the center (0.5,0.5) of $\Omega$. Similarly, the shifts in type (II) are always positive, but those in type (IV) could be negative for both components $\delta_\mu(t,s)$ and $\delta_z(t,s)$.

In all simulation examples in this section, the regression coefficients $(\beta_1, \beta_2)$, and the mean and covariance functions are all assumed unknown, and they are estimated from an IC dataset of size $(n,m)$ generated from the IC model. To determine the control limits $\kappa$ and $L$ of the charts (8) and (9), another IC dataset of the same size is generated, and the control limits $\kappa$ and $L$ are then determined by the block bootstrap procedure with the sample size $B = 10,000$ and the block size $l$, as discussed in Subsection 2.2. The actual $\text{ARL}_0$ and $\text{ARL}_1$ values are then computed based on 1,000 replicated simulations of online monitoring. Note that these $\text{ARL}_0$ and $\text{ARL}_1$ values depend on the randomly generated IC dataset used for estimating the semiparametric spatio-temporal model (1) and for determining the control limits $\kappa$ and $L$. To reduce the randomness due to the IC dataset, the entire simulation process described above, from generation of the IC datasets, estimation of the IC model, determination of the control limits, to computation of the actual $\text{ARL}_0$ and $\text{ARL}_1$ values, is repeated for 100 times. The average of the 100 $\text{ARL}_0$ (or $\text{ARL}_1$) values is used as the final estimate of the true $\text{ARL}_0$ (or $\text{ARL}_1$) value in each case considered. The corresponding standard error of the $\text{ARL}_0$ (or $\text{ARL}_1$) estimate can also be computed.

## 3.1 | Performance of the estimated IC model

In this part, we evaluate the numerical performance of the proposed model estimation method discussed in Subsection 2.1. To this end, 100 simulated IC datasets are generated, as described above, for each combination of $(n,m)$ and $(\rho_t, \rho_s)$. The regression coefficients $(\beta_1, \beta_2)$ are then estimated from each of these IC datasets. The box-plots of the 100 sets of $(\beta_1, \beta_2)$ estimates are presented in Figure 2 in six cases when $(n,m) = (200,64)$ or (400,100) and $(\rho_t, \rho_s) = (0.2, 0.1)$, (0.4,0.2) or (0.6,0.3). From the figure, it can be seen that: (i) the median estimates of $(\beta_1, \beta_2)$ are close to their true values of $(0.3, 0.3)$ in all cases considered, implying that the proposed model estimation method is reliable, (ii) the estimates of $(\beta_1, \beta_2)$ are closer to their true values when $(\rho_t, \rho_s)$ are smaller (i.e., the temporal and spatial data correlation is weaker), which is intuitively reasonable, and (iii) the results get better when $(n,m)$ are larger, implying the statistical consistency of the estimates that has been verified theoretically in Theorem 1. Regarding the estimated mean and covariance/variance functions, it has been well studied in the literature that they would converge to the true functions when both $m$ and $n$ increase. See, for instance, Yang and Qiu[17,18] for a related discussion.

[Figure 2 about here.]

## 3.2 | Impact of the IC data size $(n,m)$ and the block size $l$ on the performance of MNEW

Performance of the proposed control charts NEW and MNEW depends on the IC data size $(n,m)$ and the block size $l$ of the block bootstrap procedure. In this part, we study such dependence for the chart MNEW. The results for the chart NEW are similar and thus omitted here. Let us consider cases when $\lambda = 0.2$, $\text{ARL}_{0,z} = 200$ and $\text{ARL}_0 = 200$. To study the impact of $(n,m)$ on the performance of MNEW, we fix $l$ at 10 and let $n$ change from 100 to 1200 and $m$ change among 16, 36, 64 and 100. Results of the actual $\text{ARL}_0$ values of MNEW after its control limit is determined by the block bootstrap procedure are presented in Figure 3, where the shaded area in each panel denotes those actual $\text{ARL}_0$ values that are within 5% of the nominal level 200. From the plots of the figure, it can be seen that: (i) the actual $\text{ARL}_0$ values become closer to the nominal level 200 when $m$ and $n$ increase, (ii) in cases when $m \geq 64$, the performance of MNEW is quite stable when $n \geq 200$, and (iii) the results are better when the data correlation is weaker.

[Figure 3 about here.]

Next, we study the impact of the block size $l$ on the performance of MNEW. To this end, $(n, m)$ are fixed at $(200, 64)$ or $(400, 100)$, and $l$ is allowed to change from 1 to 20. Other setups are kept the same as those in the example of Figure 3. The related results of the actual $\text{ARL}_0$ values of MNEW are shown in Figure 4. From the figure, it can be seen that: (i) when $l \in [10, 15]$, MNEW performs quite satisfactorily unless the data correlation is very strong (i.e., $(\rho_t, \rho_s) = (0.6, 0.3)$) and the IC data size is relatively small (i.e., $(n, m) = (200, 64)$), and (ii) the performance of MNEW becomes worse when the data correlation is stronger, as expected. Based on this example, it seems reasonable to choose $l$ in the interval $[10, 15]$.

[Figure 4 about here.]

## 3.3 | Impact of the parameters $\lambda$ and $\text{ARL}_{0,z}$ on the performance of MNEW

The proposed charts NEW and MNEW have two other parameters $\lambda$ and $\text{ARL}_{0,z}$ involved. In this part, we study their impact on the performance of NEW and MNEW. Because the results for NEW are similar to those for MNEW, they are not presented here. Let us consider cases when $\text{ARL}_0 = 200$, $(n, m) = (200, 64)$, $(\rho_t, \rho_s) = (0.4, 0.2)$, $l = 10$, $\lambda = 0.1, 0.2$ or $0.3$, and $\text{ARL}_{0,z} = 100, 200$ or $400$. The actual $\text{ARL}_1$ values and the corresponding standard errors of MNEW for detecting the four types of shifts described earlier are presented in Table 1. In each row of the table, the bold numbers denote the smallest ones for each $\text{ARL}_{0,z}$ value, and the bold italic number denotes the smallest number in the entire row. From the table, we can have the following conclusions. First, for detecting shifts of types (I) and (III) that are not due to covariates, MNEW performs the best when $\text{ARL}_{0,z} = 400$. This is reasonable because the control limit $\kappa$ of the EWMA chart for the covariates (cf., (8)) would be large due to a large value of $\text{ARL}_{0,z}$. Consequently, the weighting function $\text{W}(u; \lambda, \kappa)$ defined in (10) is closer to the regular weighting parameter $\lambda$ in this case, compared to cases when $\text{ARL}_{0,z} = 100$ or $200$. The resulting chart, which is more similar to the regular EWMA chart, would be more effective in such cases. From the table, for detecting shifts of types (II) and (IV) that are partially related to covariates, MNEW performs the best when $\text{ARL}_{0,z} = 100$. So, selection of $\text{ARL}_{0,z}$ depends on whether a future shift is related to a shift in covariates: if the answer is "yes", then $\text{ARL}_{0,z}$ should be chosen small, and it should be chosen large otherwise. In practice, we may not have such prior information. In such cases, we suggest choosing $\text{ARL}_{0,z} = \text{ARL}_0$. From Table 1, it can be seen that the corresponding results when $\text{ARL}_{0,z} = \text{ARL}_0 = 200$ may not be the best, but they are reasonably good. Second, when the value of $\text{ARL}_{0,z}$ is given, we can see that different values of $\lambda$ can result in different $ARL_1$ values, although the impact of $\lambda$ seems a little smaller than that of $\text{ARL}_{0,z}$. From the table, it seems that the general conclusion about the impact of $\lambda$ is still true that a smaller $\lambda$ is good for detecting a smaller shift and a larger $\lambda$ is good for detecting a larger shift (cf., Chapter 5, Qiu[24]).

[Table 1 about here.]

## 3.4 | Numerical comparison of different methods

In this part, we compare the numerical performance of the proposed charts NEW and MNEW with a number of alternative methods. The four alternative methods considered in the comparison are briefly described below.

(i) Zhao et al[25] suggested a disease outbreak detection method, denoted as DODZ, using the kernel smoothing method for estimating the IC pattern of the disease incidence rates. For sequential monitoring of the observed disease incidence rates, it first removes the estimated IC pattern from the observed data and calculates the related "raw" residuals. Then, at each observation location, an AR(2) model is used to decorrelate the "raw" residuals over time and obtain the "model-based" residuals. Finally, a signal of disease outbreak is triggered if a "model-based" residual exceeds a threshold value $c$, where $c$ is chosen by a parametric bootstrap procedure under the assumptions that the observations collected at different locations are independent and the "model-based" residuals follow a normal distribution.

(ii) Chen et al[26] suggested a distribution-free EWMA chart, denoted as DFEWMA, for monitoring high-dimensional processes under the assumptions that the IC process distribution is unchanged over time and observations at different time points are independent. In the DFEWMA chart, time-varying control limits are used and they are obtained from a permutation testing procedure.

(iii) In the proposed chart NEW, a regular weighting parameter $\lambda$ is used in (9), replacing the covariate-dependent weight $\mathrm{W}(E_{z,i}; \lambda, \kappa)$. The resulting chart does not use any covariate information. It is denoted as WOC, represeting "without covariate" information.

(iv) In the proposed chart MNEW, a regular weighting parameter $\lambda$ is used. The resulting chart is denoted as MWOC.

We first study the IC performance of all six methods discussed above. In the charts DFEWMA, WOC, MWOC, NEW and MNEW, the weighting parameter $\lambda$ is chosen to be $0.1, 0.2$ or $0.3$. In the method DODZ, the threshold value $c$ is determined by a parametric bootstrap procedure, as discussed in Zhao et al.[25] In the chart DFEWMA, there is a window size parameter $w$ to choose. In this paper, we adopt the suggestion by Chen et al[26] that $w$ is chosen to be the smallest integer satisfying $(1-\lambda)^w \leq 0.05$. As mentioned above, the control limit of DFEWMA is chosen by a permutation testing procedure suggested by Chen et al.[26] The control limits of WOC, MWOC, NEW and MNEW are determined by the block bootstrap procedure discussed in Subsection 2.2 with $B = 10,000$ and $l = 10$. In both NEW and MNEW, $\mathrm{ARL}_{0,z}$ is chosen to be equal to $\mathrm{ARL}_0$. Then, in cases when the nominal $\mathrm{ARL}_0$ is 200, $(n, m) = (200, 64)$ or $(400, 100)$, and $(\rho_t, \rho_s) = (0.2, 0.1)$, $(0.4, 0.2)$ or $(0.6, 0.3)$, the calculated actual $\mathrm{ARL}_0$ values and their standard errors are presented in Table 2. From the table, it can be seen that (i) the actual $\mathrm{ARL}_0$ values of DODZ and DFEWMA are quite far away from the nominal $\mathrm{ARL}_0$ value of 200, due to the fact that some of their model assumptions are violated in this example, (ii) the actual $\mathrm{ARL}_0$ values of WOC, MWOC, NEW and MNEW are all close to the nominal $\mathrm{ARL}_0$ value, and thus these four charts all have a reliable IC performance, and (iii) for each method, its actual $\mathrm{ARL}_0$ values are closer to the nominal $\mathrm{ARL}_0$ value when $(n, m)$ are larger and/or $(\rho_t, \rho_s)$ are smaller, which is intuitively reasonable.

[Table 2 about here.]

Next, we compare the OC performance of all six methods in cases when the nominal $\mathrm{ARL}_0$ value is fixed at 200, $(n, m) = (200, 64)$ or $(400, 100)$, and $(\rho_t, \rho_s) = (0.2, 0.1)$, $(0.4, 0.2)$ or $(0.6, 0.3)$. To make the comparison fair, the control limits of the six charts, especially the charts DODZ and DFEWMA, are all adjusted such that their actual $\mathrm{ARL}_0$ values equal to 200. In addition, a weighting parameter $\lambda$ is involved in the charts DFEWMA, WOC, MWOC, NEW and MNEW, and the OC performance of the related charts may not be comparable if they use a same value of $\lambda$ (cf., Qiu[27]). To overcome this difficulty, their optimal OC performance is considered here, which is obtained by changing the value of $\lambda$ for each method such that its $\mathrm{ARL}_1$ value is minimized for detecting a given shift. In the charts NEW and MNEW, $\mathrm{ARL}_{0,z}$ is chosen to be equal to $\mathrm{ARL}_0$, and other setups of this example are the same as those in Tables 1 and 2. Then, the optimal $\mathrm{ARL}_1$ values of the six methods are shown in Figures 5 and 6. From the figures, we can have the following conclusions. (i) The four charts WOC, MWOC, NEW and MNEW perform uniformly better than the remaining two charts DODZ and DFEWMA in all cases considered. (ii) WOC performs slightly better than MWOC, NEW and MNEW for detecting shifts of type (I). (iii) For detecting shifts of type (II), NEW and MNEW are better than WOC and MWOC. (iv) MNEW and MWOC are better than NEW and WOC for detecting shifts of type (III). (v) To detect shifts of type (IV), MNEW performs the best. The conclusion (i) is reasonable because some model assumptions of DODZ and DFEWMA are invalid in this example. All shifts of type (I) are positive and not due to covariates. So, the conclusion (ii) is reasonable because the chart WOC did not use any covariate information in its construction, it did not consider any modifications for handling negative shifts, and thus the variability of its charting statistic would be smaller than that of the charts MWOC, NEW and MNEW. Since the shifts of type (II) are partially due to covariates and they are all positive, NEW and MNEW would be better than WOC and MWOC, and NEW would be better than MNEW. This explains why the conclusion (iii) is also reasonable. The remaining two conclusions can be explained in a similar way. From this example, it can be seen that the chart MNEW performs the best or close to the best in all cases considered. Because it is often difficult to know in practice whether a future shift would be related to the covariates and whether it could be negative at some spatial locations, the chart MNEW is recommended in this paper.

[Figure 5 about here.]

[Figure 6 about here.]

## 4 | A REAL-DATA APPLICATION

In this section, we demonstrate our proposed methodology using a real-data example. As discussed in Section 1, because infectious diseases could have a great damage to public health, many disease reporting systems have been developed in US at the federal or state level. In the state of Florida, the disease reporting system called ESSENCE has been developed by the Florida Department of Health. ESSENCE collects daily numbers of incidences of the influenza-like illness (ILI) from 264 participating emergency departments and urgent care centers across the entire state. Researchers can have an access to the database after a proper application, but the provided data have been organized in county level by the reporting system. In this section, its data collected during 2012-2014 are used. The observed ILI incidence rates on 06/15 (a summer time) and 12/15 (a winter time) during the three years are presented in Figure 7. From the figure, it can be seen that the disease incidence rates in the winters are much higher than those in the summers, which is mainly due to the seasonality of ILI. However, the disease incidence rates in the winter of 2014 seem much higher than those in the winters of 2012 and 2013. Thus, it is interesting to see whether there was a disease outbreak occurred in that year. In the literature, it has been well discussed that diseases like ILI could be highly associated with certain climate conditions, such as air temperature and relative humidity (cf., Noort et al[28]). For this reason, we downloaded the observed data of the time-dependent relative humidity and the space/time-dependent air temperature in Florida during 2012-2014 from the official website of the National Oceanic and Atmospheric Administration of the United States, and the two variables Relative Humidity (in the unit of %) and Air Temperature (in the unit of °F) will be used as covariates in our proposed method.

[Figure 7 about here.]

As discussed above, the observed data of the disease incidence rates during 2012 and 2013 look stable. Thus, they are used as the IC data in our proposed method. This IC data are then splitted into two parts: those in 2013 are used for estimating the IC model (1), and those in 2012 are used for determining the control limits of the related control charts. From the estimated IC model, the estimated regression coefficients of Relative Humidity and Air Temperature are $-1.17 \times 10^{-6}$ and $-1.06 \times 10^{-6}$, respectively. Therefore, both covariates are negatively associated with the ILI incidence rates, which is consistent with our intuition and with the conclusions found in the literature (cf., Pica and Bouvier,[29] Schulman and Kilbourne[30]). In the proposed charts NEW and MNEW, we choose $\lambda = 0.1$ and $\text{ARL}_0 = 200$, which are commonly used in an EWMA chart. Also, we choose $\text{ARL}_{0,z}$ to be equal to $\text{ARL}_0$, as suggested in Section 3. The control limits $\kappa$ and $L$ are determined by the block bootstrap procedure discussed in Subsection 2.2 with the bootstrap sample size $B = 10,000$ and the block length $l = 10$. For comparison purposes, the methods DODZ, DFEWMA, WOC and MWOC are also considered here. For the method DODZ, its control limit is determined by the parametric bootstrap approach suggested by Zhao et al.[25] The time-varying control limits of DFEWMA are chosen by the permutation testing procedure suggested by Chen et al.[26] In the charts DFEWMA, WOC and MWOC, $\lambda$, $\text{ARL}_0$, $B$ and $l$ are chosen to be the same as those in NEW and MNEW. The charting statistics of the six methods for monitoring the observed disease incidence rates in 2014 are shown in Figure 8. From the figure, it can be seen that (i) the charting statistic of DODZ is very noisy and it gives the first signal of disease outbreak in middle January of 2014, (ii) DFEWMA gives signals almost every day and its first signal is on 01/01/2014, (iii) the charts WOC and MWOC are quite similar and their first signals are on 10/07 and 10/10, respectively, and (iv) the charts NEW and MNEW are also similar and their first signals are both on 09/27. This example shows that the signals from the charts DODZ and DFEWMA are too frequent to be really helpful. As a comparison, the performance of the charts WOC and MWOC seems more reasonable because they have taken into account the dynamic nature (e.g., seasonality) of the IC process distribution and the possible spatio-temporal data correlation. Finally, the charts NEW and MNEW seem more effective than WOC and MWOC after accommodating the helpful information in covariates for disease surveillance.

[Figure 8 about here.]

To verify the signals of NEW and MNEW, the ILI incidence rates of the entire Florida state during 09/01-12/31 in all three years of 2012-2014 are presented in the left panel of Figure 9, where the dark dashed line denotes the estimated mean function of the IC model and the vertical thin like denotes the first signal time of NEW and MNEW. To better perceive the difference between the observed ILI incidence rates and the estimated IC mean function, the right panel of Figure 9 presents the residuals of the 2014 ILI incidence rates (i.e, differences between the observations and the estimated mean values). From the plots in the figure, it can be seen that (i) the estimated IC mean function describes the longitudinal pattern of the IC data well, (ii) the observed ILI incidence rates start to deviate from the estimated IC mean function in early or middle September of 2014, and (iii) the charts NEW and MNEW can react to such a systemic deviation promptly.

[Figure 9 about here.]

## 5 | CONCLUDING REMARKS

In the previous sections, we present a new sequential monitoring method for disease surveillance. The new method can accommodate spatio-temporal data correlation, time-varying IC process distribution (e.g., seasonality), and nonparametric data distribution. One novelty of the new method is that it takes advantage of the covariate information in the way that a signal from the proposed chart can only be triggered by unusual patterns in the observed disease incidence data but the helpful covariate information can improve its effectiveness. Extensive simulation studies and a real data application have shown that the new method performs well in different cases. However, there are still some issues about the proposed method that need to be addressed in the future research. For instance, in practice there could be a large number of covariates relevant to the incidence rates of a disease. Intuitively, only those highly related to the disease incidence rates should be included in the IC model, to reduce the variability of the estimated IC model and improve the efficiency of the subsequent process monitoring. Therefore, a reliable and effective variable selection procedure should be helpful when estimating the IC model. Also, although the semiparametric IC model (1) is already very flexible, it assumes that the relationship between the covariates and the response is linear, which may not be appropriate in some applications. This model is possible to be further generalized to a semiparametric model with space/time-varying coefficients or even a completely nonparametric model. However, such possible generalizations are not straightforward and they require much future research effort. In addition, the proposed method assumes that the time period $T$ is known, which is reasonable in some applications (e.g., the seasonality of some infectious diseases like influenza is usually in years). For some other applications, especially those involve diseases that we are unfamiliar with, $T$ might be unknown and it needs to be properly estimated in advance, which requires further research.

**Acknowledgments:** The authors thank the editor, the associate editor and two referees for their constructive comments and suggestions, which improved the quality of the paper greatly. This research was supported in part by an NSF grant with the number DMS-1914639.

## References

1. Chen H, Zeng D, Yan P. *Infectious Disease Informatics*, New York: Springer; 2010.

2. Shmueli G, Burkom H. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*. 2010;52:39–51.

3. Fiore AE, Uyeki TM, Broder K, Finelli L, Euler GL, Singleton JA, Iskander JK, Wortley PM, Shay DK, Bresee JS, Cox NJ. Prevention and control of influenza with vaccines: recommendations of the advisory committee on immunization practices. *MMWR Recomm Rep*. 2010;59: 1–62.

4. Jacquez GM. A $k$ nearest neighbor test for spacetime interaction. *Stat Med*. 1996;15:1935–1949.

5. Knox E, Bartlett M. The detection of space-time interactions. *Journal Royal Stat Soc Ser C*. 1964;13:25–30.

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

7. Kulldorff M. A spatial scan statistic. *Commun Stat–Theory Methods*. 1997;26:1481–1496.

8. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *Int J Health Geogr*. 2008;7:article 14.

9. Woodall WH, Marshall JB, Joner MD, Fraker SE, Abdel-Salam, ASG. On the use and evaluation of prospective scan methods for health-related surveillance. *Journal Royal Stat Soc Ser A*. 2008;171:223–237.

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

11. Zhang J, Qiu P, Chen X. Statistical monitoring-based alarming systems in modeling the AIDS epidemic in the US, 1985-2011. *Curr HIV Res*. 2016;14:130–137.

12. Dassanayake S, French, JP. An improved cumulative sum-based procedure for prospective disease surveillance for count data in multiple regions. *Stat Med*. 2016;35:2593–2608.

13. Dong Y, Hedayat AS, Sinha BK. Surveillance strategies for detecting changepoint in incidence rate based on exponentially weighted moving average methods. *J Am Stat Assoc*. 2008;103:843–853.

14. Yang K, Qiu P. Online sequential monitoring of spatio-temporal disease incidence rates. *IISE Trans*. 2020;52:1218–1233.

15. Yang K, Qiu, P. Adaptive process monitoring using covariate information. *Technometrics*. 2020. https://doi.org/10.1080/00401706.2020.1772115.

16. Speckman P. Kernel smoothing in partial linear models. *Journal Royal Stat Soc Ser B*. 1988;50:413–436.

17. Yang K, Qiu P. Spatio-temporal incidence rate data analysis by nonparametric regression. *Stat Med*. 2018;37:2094–2107.

18. Yang K, Qiu P. Nonparametric estimation of the spatio-temporal covariance structure. *Stat Med*. 2019;38:4555–4565.

19. Epanechnikov VA. Non-parametric estimation of a multivariate probability density. *Theory Probab Its Appl*. 1969;14:153–158.

20. Altman NS. Kernel smoothing of data with correlated errors. *J Am Stat Assoc*. 1990;85:749–758.

21. Opsomer J, Wang Y, Yang Y. Nonparametric regressin with correlated errors. *Stat Sci*. 2001;16:134–153.

22. Brabanter KD, Brabanter JD, Suykens JAK, Moor BD. Kernel regression in the presence of correlated errors. *J Mach Learn Res*. 2011;12:1955–1976.

23. Roberts SW. Control chart tests based on geometric moving averages. *Technometrics*. 1959;1:239–250.

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

25. Zhao Y, Zeng D, Herring AH, Ising A, Waller A, Richardson D, Kosorok MR. Detecting disease outbreaks using local spatiotemporal methods. *Biometrics*. 2011;67:1508–1517.

26. Chen N, Zi X, Zou C. A distribution-free multivariate control chart. *Technometrics*. 2016;58:448–459.

27. Qiu P. Distribution-free multivariate process control based on log-linear modeling. *IIE Trans*. 2008;40:664–677.

28. Noort SP, Aguas R, Ballesteros S, Gomes MG. The role of weather on the relation between influenza and influenza-like illness. *J Theor Biol*. 2012;298:131-137.

29. Pica N, Bouvier NM. Environmental factors affecting the transmission of respiratory viruses. *Curr Opin Vriol*. 2012;2: 90–95.

30. Schulman JL, Kilbourne ED. Airborne transmission of influenza virus infection in mice. *Nature*. 1962;195:1129–1130.

# APPENDIX

In the appendix, we present some statistical properties of the estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\mu}(t,\boldsymbol{s})$, $\widehat{\mu}_z(t,\boldsymbol{s})$, $\widehat{\mu}_y(t,\boldsymbol{s}) = \widehat{\mu}(t,\boldsymbol{s})+\widehat{\mu}_z(t,\boldsymbol{s})$, $\widehat{V}_y(t,t';\boldsymbol{s},\boldsymbol{s}')$ and $\widehat{V}_z(t,t';\boldsymbol{s},\boldsymbol{s}')$ defined in Section 2.1. To this end, the strong mixing coefficient of order $k$ in the time domain for the random errors $\{\varepsilon(t_i,\boldsymbol{s}_{ij})\}$ in model (1) is defined to be

$$\alpha_\varepsilon(k) = \sup_{n\geq k+1, 1\leq i\leq n-k} \sup_{A,B} \left\{ |P(AB) - P(A)P(B)| \,:\, A \in \mathcal{F}_1^i, B \in \mathcal{F}_{i+k}^n \right\},$$

where $\mathcal{F}_{i_1}^{i_2}$ denotes the $\sigma$-algebra generated by $\{\varepsilon(t_i,\boldsymbol{s}_{ij}), j = 1,\ldots,m_i, i_1 \leq i \leq i_2\}$. For the random errors $\{\varepsilon_z(t_i,\boldsymbol{s}_{ij}) = z(t_i,\boldsymbol{s}_{ij}) - E(z(t_i,\boldsymbol{s}_{ij}))\}$, the strong mixing coefficient can be defined in the same way, which is denoted as $\alpha_z(k)$.

**Theorem 1.** For model (1) and its estimation, assume that: i) $\Omega$ is a bounded and closed set in $\mathbb{R}^2$, $\{m_i, i=1,\dots,n\}$ are all in the same order as $m$, $\{\boldsymbol{s}_{ij}, j=1,\dots,m_i\}$ follow a distribution with density $f(\boldsymbol{s})$ for each $i$, they are independent of each other, and independent of $\{\varepsilon(t_i,\boldsymbol{s}_{ij})\}$ and $\{\mathbf{X}_1(t_i),\mathbf{X}_2(t_i,\boldsymbol{s}_{ij})\}$; ii) $f(\boldsymbol{s})$ is twice continuously differentiable and has a non-zero lower bound in $\Omega$; iii) the functions $\mu(t,\boldsymbol{s})$, $\mu_z(t,\boldsymbol{s})$, $V_y(t,t';\boldsymbol{s},\boldsymbol{s}')$ and $V_z(t,t';\boldsymbol{s},\boldsymbol{s}')$ are all twice continuously differentiable in $[0,T]\times\Omega$; iv) there exist two constants $C_0, C_1>0$ such that, $\alpha(k)\le C_0\exp(-C_1k)$, for any $k\ge 0$, where $\alpha(k)=\max\left(\alpha_\varepsilon(k),\alpha_z(k)\right)$; v) there are constants $C_2, C_3, \omega>0$ such that $P(|\varepsilon(t,\boldsymbol{s})|\ge k)\le C_2k^\omega\exp(-C_3k)$ and $P(|\varepsilon_z(t,\boldsymbol{s})|\ge k)\le C_2k^\omega\exp(-C_3k)$, for any $(t,\boldsymbol{s})\in[0,T]\times\Omega$ and $k\ge 0$; vi) the kernel functions $K_1(u)$ and $K_2(u)$ are both symmetric about 0 and Lipschitz-1 continuous density functions with finite supports; vii) $h_{s,\ell}=o(1)$, $h_{t,\ell}/h_{s,\ell}=O(1)$, $\log(n)/(mh^2_{s,\ell})=o(1)$, and $\log(n)^2/(nh^2_{t,\ell})=o(1)$, for $1\le\ell\le 4$. Then, we have

1) $||\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}||_\infty = O_p\left(h^2_{t,1}+h^2_{s,1}+[1/(nh_{t,1})]^{1/2}\right)$,

2) $\sup\limits_{(t,\boldsymbol{s})\in[0,T]\times\Omega}\left|\widehat{\mu}(t,\boldsymbol{s})-\mu(t,\boldsymbol{s})\right| = O_p\left(h^2_{t,1}+h^2_{s,1}+[\log(n)^2/(nh^2_{t,1})]^{1/2}\right)$,

3) $\sup\limits_{(t,\boldsymbol{s})\in[0,T]\times\Omega}\left|\widehat{\mu}_z(t,\boldsymbol{s})-\mu_z(t,\boldsymbol{s})\right| = O_p\left(h^2_{t,3}+h^2_{s,3}+[\log(n)^2/(nh^2_{t,3})]^{1/2}\right)$,

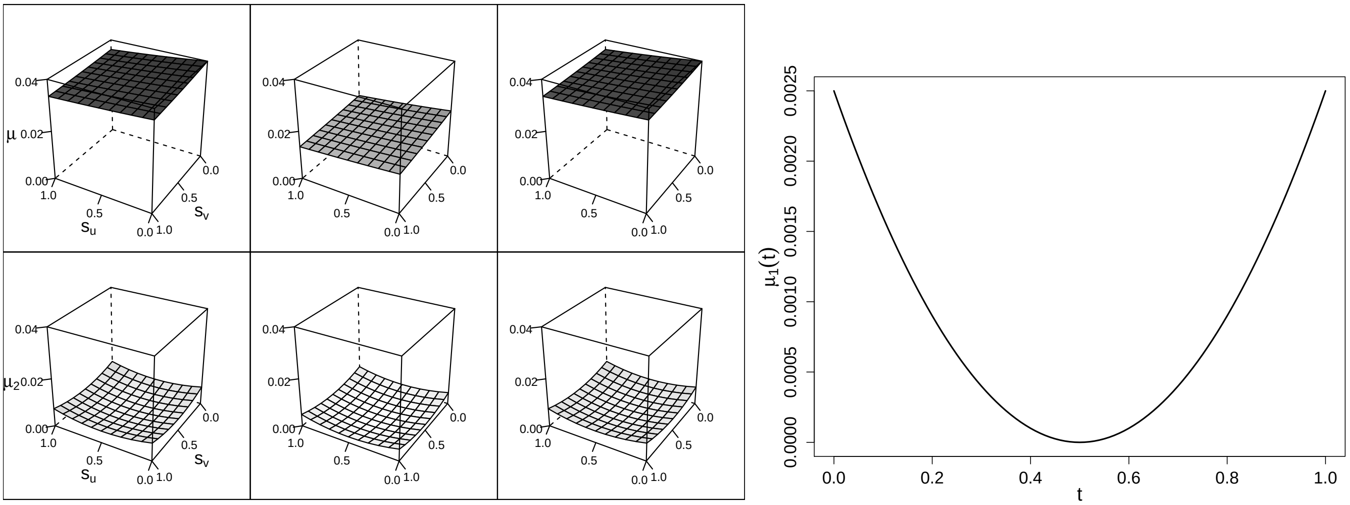
4) $\sup\limits_{(t,\boldsymbol{s})\in[0,T]\times\Omega}\left|\widehat{\mu}_y(t,\boldsymbol{s})-\mu_y(t,\boldsymbol{s})\right| = O_p\left(h^2_{t,1}+h^2_{s,1}+h^2_{t,3}+h^2_{s,3}+[\log(n)^2/(nh^2_{t,1})]^{1/2}+[\log(n)^2/(nh^2_{t,3})]^{1/2}\right)$,

5) $\sup\limits_{(t,\boldsymbol{s}),(t',\boldsymbol{s}')\in[0,T]\times\Omega}\left|\widehat{V}_y(t,t';\boldsymbol{s},\boldsymbol{s}')-V_y(t,t';\boldsymbol{s},\boldsymbol{s}')\right| = O_p\left(h^2_{t,max}+h^2_{s,max}+[\log(n)^2/(nh^2_{t,min})]^{1/2}\right)$,

6) $\sup\limits_{(t,\boldsymbol{s}),(t',\boldsymbol{s}')\in[0,T]\times\Omega}\left|\widehat{V}_z(t,t';\boldsymbol{s},\boldsymbol{s}')-V_z(t,t';\boldsymbol{s},\boldsymbol{s}')\right| = O_p\left(\widetilde{h}^2_{t,max}+\widetilde{h}^2_{s,max}+[\log(n)^2/(n\widetilde{h}^2_{t,min})]^{1/2}\right)$,

where $\|\cdot\|_\infty$ is the maximum norm, $h_{t,max}=\max\{h_{t,1},h_{t,2},h_{t,3}\}$, $h_{t,min}=\min\{h_{t,1},h_{t,2},h_{t,3}\}$, $\widetilde{h}_{t,max}=\max\{h_{t,1},h_{t,3},h_{t,4}\}$, and $\widetilde{h}_{t,min}=\min\{h_{t,1},h_{t,3},h_{t,4}\}$.

The proof of Theorem 1 is given in the supplementary file.



(a)

(b)

**FIGURE 1** (a) IC mean functions $\mu(t,\boldsymbol{s})$ (1st row) and $\mu_2(t,\boldsymbol{s})$ (2nd row) when $t=0$ (1st column), $t=0.5$ (2nd column) and $t=1$ (3rd column). (b) IC mean function $\mu_1(t)$.

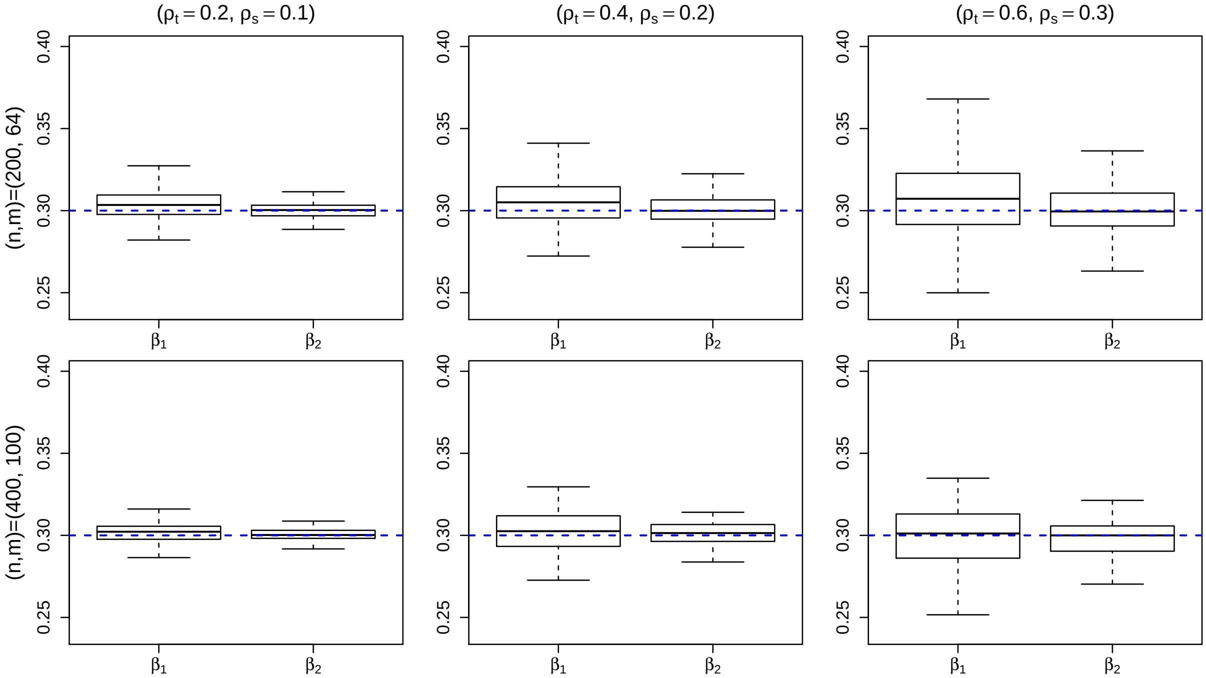


**FIGURE 2** Boxplots of the estimates $(\widehat{\beta}_1, \widehat{\beta}_2)$ based on 100 replicated simulations. In each panel, the y-axis denotes the values of $\widehat{\beta}_1$ or $\widehat{\beta}_2$, and the bold dashed horizontal line denotes the true values of $\beta_1$ and $\beta_2$.

**TABLE 1** $\mathrm{ARL}_1$ values and their standard errors (in parentheses) of the proposed chart MNEW when $\mathrm{ARL}_0 = 200$, $(n, m) = (200, 64)$, $(\rho_t, \rho_s) = (0.4, 0.2)$ and $l = 10$. In each row, the bold numbers denote the smallest ones for each $\mathrm{ARL}_{0,z}$ value, and the bold italic number denotes the smallest number in the entire row.

| Shift | | $\mathrm{ARL}_{0,z} = 100$ | | | $\mathrm{ARL}_{0,z} = 200$ | | | $\mathrm{ARL}_{0,z} = 400$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Type | $\nu$ | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.3$ | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.3$ | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.3$ |
| (I) | 1 | 88.23(2.62) | 79.04(2.58) | **77.48**(2.12) | 58.71(2.30) | **56.65**(1.63) | 59.63(1.67) | ***50.19***(1.34) | 52.10(1.31) | 56.70(1.28) |
| | 2 | 47.91(2.17) | 34.16(2.07) | **32.24**(1.73) | 18.90(0.93) | **18.54**(0.90) | 21.76(1.04) | ***17.24***(0.70) | 18.18(0.77) | 20.84(0.80) |
| | 3 | 24.14(1.64) | 16.24(0.92) | **14.89**(0.71) | 11.04(0.40) | **10.35**(0.43) | 10.85(0.49) | 10.49(0.37) | ***10.19***(0.41) | 10.46(0.43) |
| | 4 | 13.10(1.01) | 8.01(0.31) | **7.28**(0.27) | 6.92(0.21) | 6.09(0.23) | **6.02**(0.25) | 6.13(0.20) | 5.63(0.21) | ***5.62***(0.22) |
| (II) | 1 | ***38.72***(1.28) | 39.11(1.26) | 50.53(1.27) | **39.41**(1.26) | 46.60(1.29) | 52.45(1.28) | **40.63**(1.24) | 47.70(1.27) | 53.18(1.28) |
| | 2 | ***14.95***(0.63) | 16.49(0.70) | 19.56(0.76) | **15.79**(0.68) | 17.45(0.72) | 19.64(0.78) | **15.99**(0.68) | 17.92(0.72) | 20.07(0.76) |
| | 3 | ***7.59***(0.29) | 7.69(0.34) | 8.15(0.38) | **8.18**(0.34) | 8.24(0.37) | 8.72(0.40) | **8.46**(0.35) | 8.52(0.38) | 8.94(0.41) |
| | 4 | 4.81(0.15) | ***4.66***(0.17) | 4.76(0.18) | 5.19(0.18) | **4.94**(0.19) | 4.96(0.20) | 5.44(0.19) | **5.15**(0.20) | 5.16(0.22) |
| (III) | 1 | **162.28**(4.27) | 163.56(4.30) | 164.30(4.32) | **155.82**(4.10) | 156.50(4.11) | 156.73(4.14) | ***146.20***(3.89) | 151.24(3.91) | 154.00(3.98) |
| | 2 | **132.26**(3.12) | 133.43(3.12) | 134.93(3.17) | **115.90**(3.04) | 116.56(3.06) | 120.33(3.11) | ***112.10***(2.87) | 113.01(3.01) | 118.40(3.05) |
| | 3 | 111.62(2.84) | **109.52**(2.65) | 111.41(2.71) | **80.81**(2.54) | 84.28(2.57) | 91.24(2.64) | ***73.64***(2.17) | 80.75(2.25) | 86.95(2.49) |
| | 4 | 93.18(2.61) | **87.53**(2.48) | 90.29(2.56) | **55.89**(1.56) | 60.39(1.73) | 70.78(1.91) | ***48.75***(1.35) | 56.79(1.49) | 65.03(1.69) |
| (IV) | 1 | ***133.95***(3.62) | 141.52(3.78) | 144.73(3.72) | **136.08**(3.78) | 144.13(3.88) | 146.11(3.93) | **136.49**(3.80) | 145.77(3.78) | 148.47(3.99) |
| | 2 | ***90.40***(2.67) | 105.63(2.83) | 115.19(2.91) | **92.99**(2.68) | 106.65(2.97) | 115.47(3.04) | **96.59**(2.64) | 109.85(3.01) | 116.87(3.07) |
| | 3 | ***56.93***(1.64) | 72.45(2.11) | 80.98(2.46) | **57.75**(1.71) | 74.39(2.24) | 81.43(2.51) | **60.31**(1.84) | 77.01(2.19) | 82.92(2.61) |
| | 4 | ***30.61***(1.31) | 45.67(1.51) | 56.96(1.61) | **31.51**(1.41) | 46.01(1.53) | 58.51(1.74) | **33.95**(1.48) | 48.89(1.53) | 60.52(1.76) |

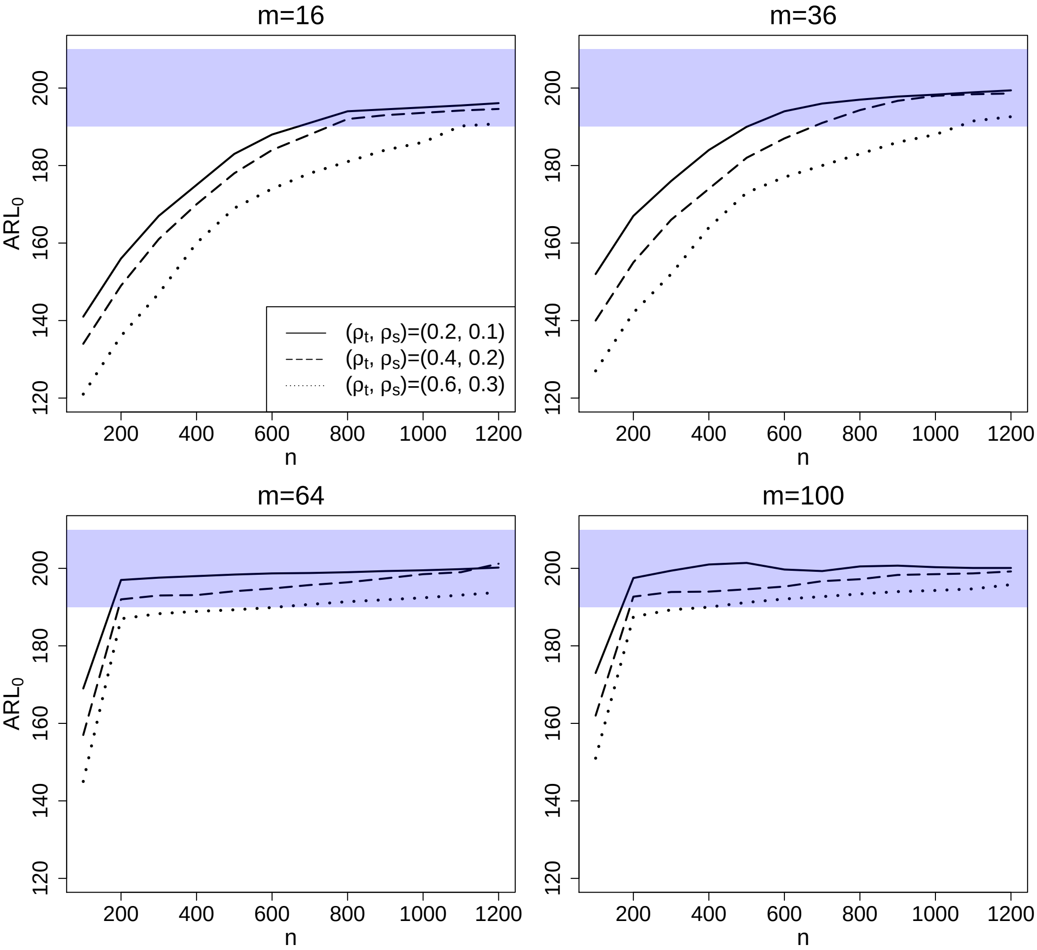


**FIGURE 3** Actual ARL$_0$ values of MNEW when the nominal ARL$_0$ value is 200, $n$ changes from 100 to 1200 and m changes among 16, 36, 64 and 100. In each panel, the shaded area denotes those ARL$_0$ values that are within 5% of the nominal ARL$_0$ level.

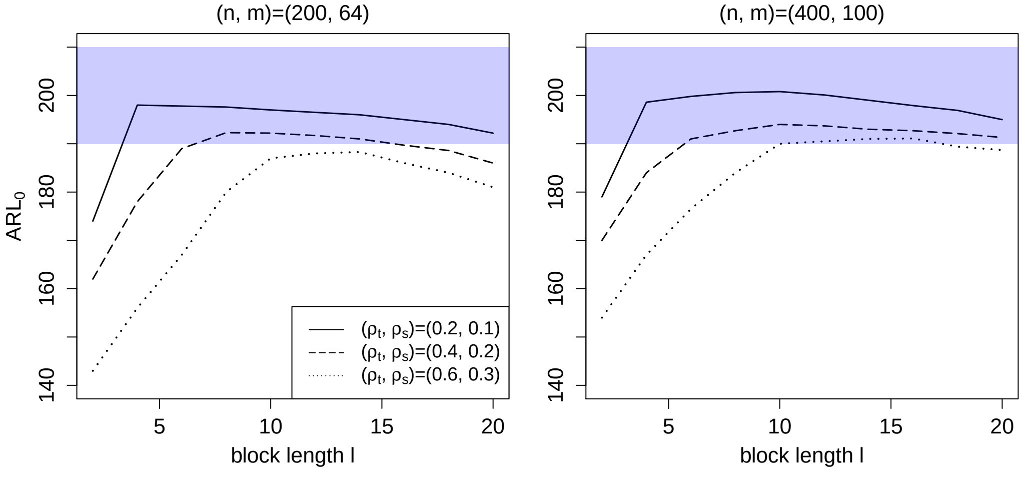


**FIGURE 4** Actual ARL$_0$ values of MNEW when the nominal ARL$_0$ value is 200, $(n, m)$ are fixed at $(200, 64)$ or $(400, 100)$, and $l$ changes from 1 to 20. In each panel, the shaded area denotes those ARL$_0$ values that are within 5% of the nominal ARL$_0$ level.

**TABLE 2** Actual ARL$_0$ values and their standard errors (in parentheses) of six control charts in different cases when the nominal ARL$_0$ value is 200.

| | | DODZ | DFEWMA | | WOC | | MWOC | | NEW | | MNEW | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(n,m)$ | $(\rho_t,\rho_s)$ | ARL$_0$ | $\lambda$ | ARL$_0$ | $\lambda$ | ARL$_0$ | $\lambda$ | ARL$_0$ | $\lambda$ | ARL$_0$ | $\lambda$ | ARL$_0$ |
| (200,64) | (0.2,0.1) | 174(9.1) | 0.1 | 112(5.8) | 0.1 | 193(11.1) | 0.1 | 195(11.3) | 0.1 | 192(11.3) | 0.1 | 194(11.3) |
| | | | 0.2 | 127(6.9) | 0.2 | 197(11.2) | 0.2 | 197(11.4) | 0.2 | 196(11.4) | 0.2 | 197(11.5) |
| | | | 0.3 | 140(7.3) | 0.3 | 198(11.3) | 0.3 | 203(11.6) | 0.3 | 198(11.5) | 0.3 | 200(11.7) |
| | (0.4,0.2) | 161(8.4) | 0.1 | 97(5.1) | 0.1 | 187(10.9) | 0.1 | 189(11.0) | 0.1 | 185(10.8) | 0.1 | 186(11.1) |
| | | | 0.2 | 108(5.7) | 0.2 | 192(11.0) | 0.2 | 192(11.2) | 0.2 | 191(11.1) | 0.2 | 192(11.2) |
| | | | 0.3 | 114(6.2) | 0.3 | 195(11.1) | 0.3 | 197(11.3) | 0.3 | 192(11.2) | 0.3 | 195(11.4) |
| | (0.6,0.3) | 144(7.6) | 0.1 | 86(4.6) | 0.1 | 182(10.6) | 0.1 | 185(10.8) | 0.1 | 181(10.7) | 0.1 | 183(11.2) |
| | | | 0.2 | 94(4.9) | 0.2 | 190(11.0) | 0.2 | 190(11.0) | 0.2 | 189(11.2) | 0.2 | 187(11.3) |
| | | | 0.3 | 104(5.7) | 0.3 | 193(11.1) | 0.3 | 192(11.1) | 0.3 | 192(11.3) | 0.3 | 193(11.3) |
| (400,100) | (0.2,0.1) | 189(9.7) | 0.1 | 121(4.8) | 0.1 | 198(8.5) | 0.1 | 197(8.5) | 0.1 | 204(8.8) | 0.1 | 203(8.9) |
| | | | 0.2 | 129(5.3) | 0.2 | 199(8.5) | 0.2 | 202(8.7) | 0.2 | 198(8.6) | 0.2 | 201(8.8) |
| | | | 0.3 | 147(6.1) | 0.3 | 200(8.6) | 0.3 | 201(8.7) | 0.3 | 200(8.7) | 0.3 | 200(8.8) |
| | (0.4,0.2) | 174(8.9) | 0.1 | 103(4.6) | 0.1 | 191(8.2) | 0.1 | 191(8.3) | 0.1 | 190(8.3) | 0.1 | 190(8.4) |
| | | | 0.2 | 111(4.8) | 0.2 | 195(8.5) | 0.2 | 192(8.4) | 0.2 | 193(8.5) | 0.2 | 194(8.7) |
| | | | 0.3 | 121(5.2) | 0.3 | 199(8.6) | 0.3 | 202(8.7) | 0.3 | 198(8.8) | 0.3 | 199(8.9) |
| | (0.6,0.3) | 158(7.9) | 0.1 | 92(3.7) | 0.1 | 187(8.0) | 0.1 | 189(8.1) | 0.1 | 188(8.2) | 0.1 | 186(8.2) |
| | | | 0.2 | 102(4.1) | 0.2 | 192(8.3) | 0.2 | 191(8.3) | 0.2 | 191(8.5) | 0.2 | 190(8.4) |
| | | | 0.3 | 111(4.7) | 0.3 | 195(8.4) | 0.3 | 196(8.5) | 0.3 | 196(8.7) | 0.3 | 195(8.8) |

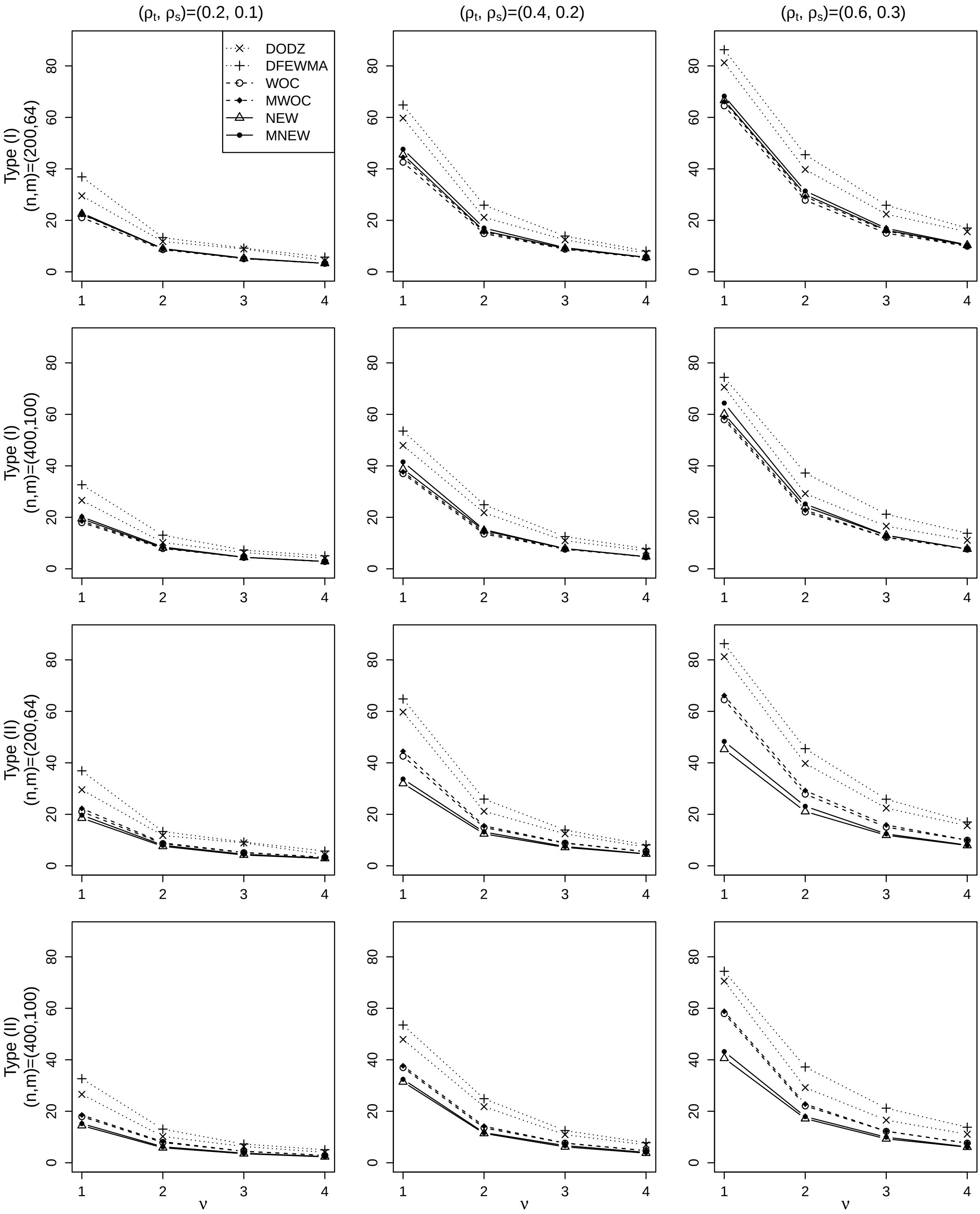


**FIGURE 5** Optimal ARL$_1$ values of the six charts DODZ, DFEWMA, WOC, MWOC, NEW and MNEW for detecting shifts of types (I) and (II).

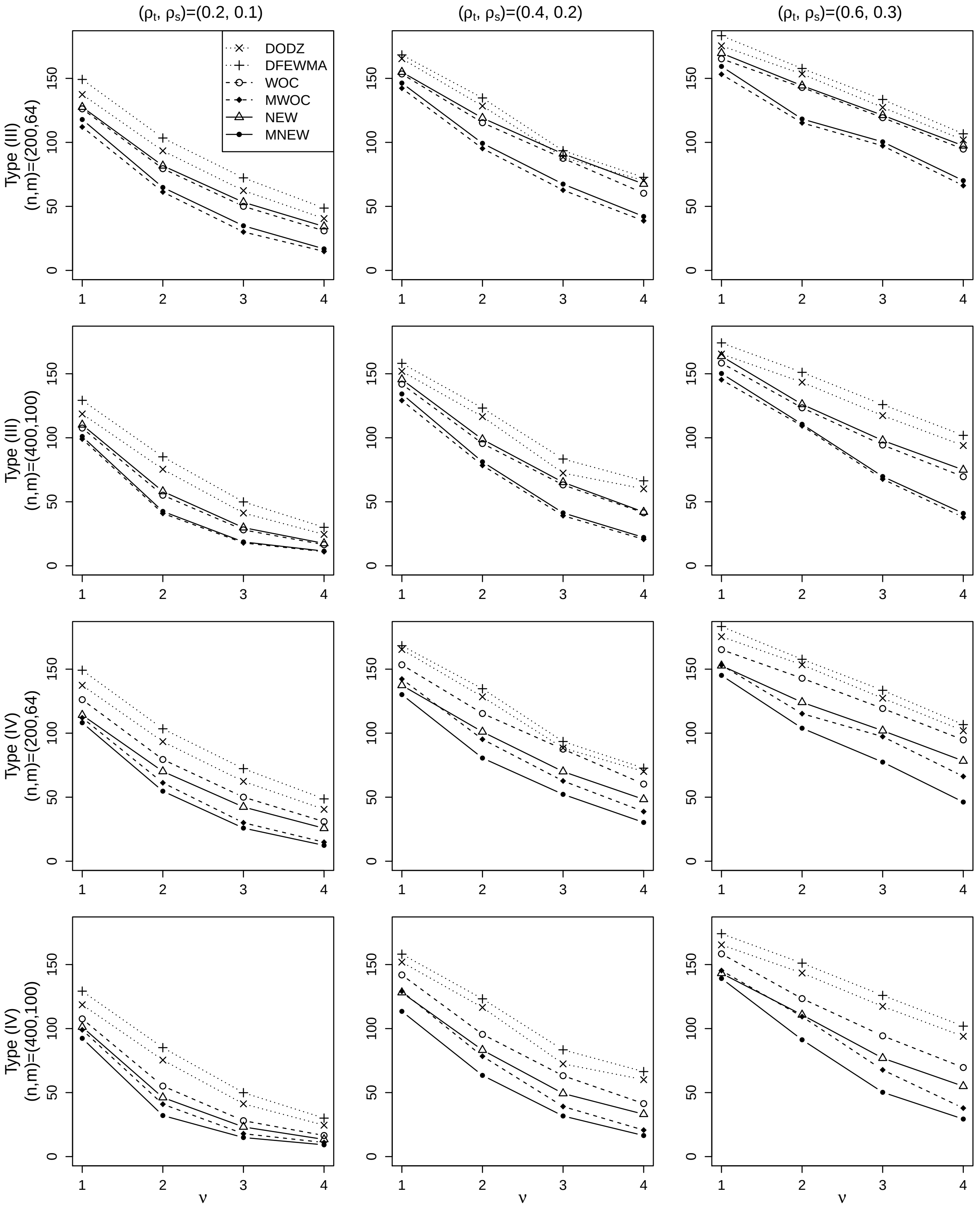


**FIGURE 6** Optimal $\mathrm{ARL}_1$ values of the six charts DODZ, DFEWMA, WOC, MWOC, NEW and MNEW for detecting shifts of types (III) and (IV).

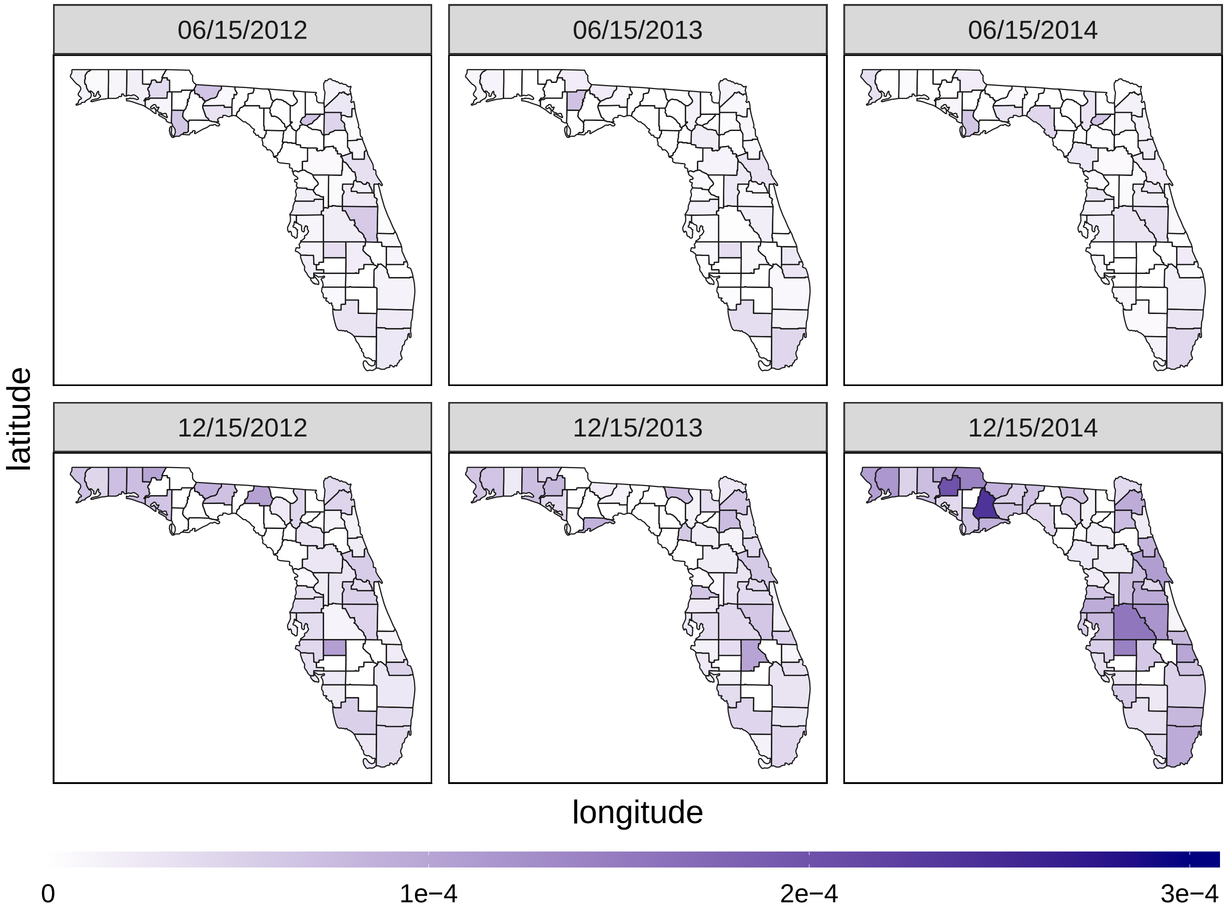


**FIGURE 7** Observed ILI incidence rates of all 67 Florida counties on 06/15 and 12/15 in years 2012-2014.

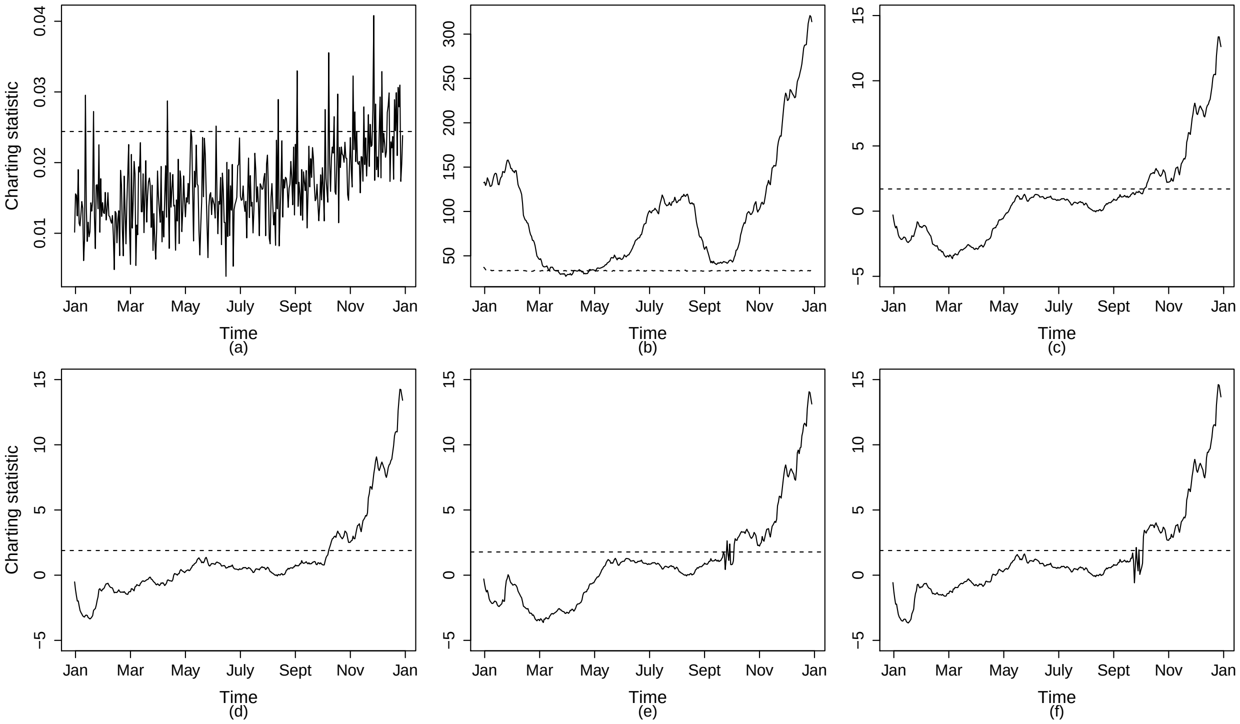


**FIGURE 8** Six control charts for monitoring the observed spatial incidence rates of ILI in 2014: DODZ (plot (a)), DFEWMA (plot (b)), WOC (plot (c)), MWOC (plot (d)), NEW (plot (e)) and MNEW (plot (f)). The horizontal line in each plot denotes a control limit of the related chart.

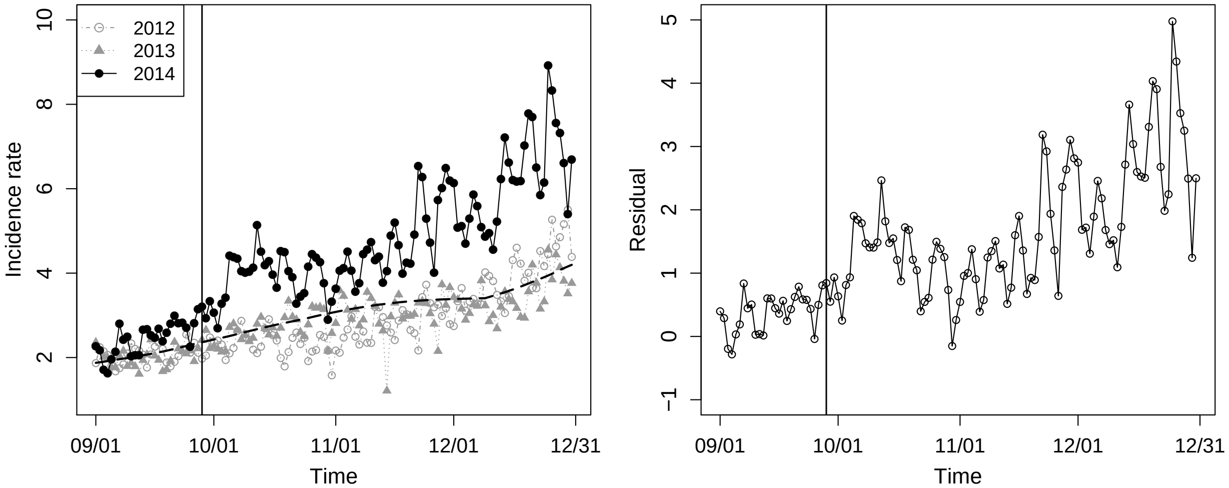


**FIGURE 9** Left panel: observed daily ILI incidence rates during 09/01-12/31 in years 2012-2014 and the estimated IC mean function (dark dashed line). Right panel: residuals of the observed ILI incidence rates during 09/01/2014-12/31/2014. In each panel, the vertical line denotes the first signal time of NEW and MNEW, and the y-axis is in the scale of $10^{-5}$.