The Annals of Applied Statistics
2023, Vol. 17, No. 1, 47–66
https://doi.org/10.1214/22-AOAS1615

CONTROL CHARTS FOR DYNAMIC PROCESS MONITORING WITH AN APPLICATION TO AIR POLLUTION SURVEILLANCE

BY XIULIN XIE[a] AND PEIHUA QIU[b]

Department of Biostatistics, University of Florida, [a]*xiulin.xie@ufl.edu,* [b]*pqiu@ufl.edu*

Air pollution is a major global public health risk factor. Among all air pollutants, $PM_{2.5}$ is especially harmful. It has been well demonstrated that chronic exposure to $PM_{2.5}$ can cause many health problems, including asthma, lung cancer and cardiovascular diseases. To tackle problems caused by air pollution, governments have put a huge amount of resources to improve air quality and reduce the impact of air pollution on public health. In this effort it is extremely important to develop an air pollution surveillance system to constantly monitor the air quality over time and to give a signal promptly once the air quality is found to deteriorate so that a timely government intervention can be implemented. To monitor a sequential process, a major statistical tool is the statistical process control (SPC) chart. However, traditional SPC charts are based on the assumptions that process observations at different time points are independent and identically distributed. These assumptions are rarely valid in environmental data because seasonality and serial correlation are common in such data. To overcome this difficulty, we suggest a new control chart in this paper, which can properly accommodate dynamic temporal pattern and serial correlation in a sequential process. Thus, it can be used for effective air pollution surveillance. This method is demonstrated by an application to monitor the daily average $PM_{2.5}$ levels in Beijing and shown to be effective and reliable in detecting the increase of $PM_{2.5}$ levels.

1. Introduction. Air pollution is a major global public health risk factor, especially in countries such as China, India and other low- and middle-income countries (Health Effects Institute (2019)). Among all air pollutants, fine particle masses with aerodynamic diameters $\leq 2.5\mu m$ (i.e., $PM_{2.5}$) are especially harmful, since they are small enough to penetrate deep into our respiratory tract and lungs and, consequently, damage lung function and the human respiratory system (Boogaard, Walker and Cohen (2019), Cohen et al. (2017), Xing et al. (2016)). It has been confirmed that chronic exposure to $PM_{2.5}$ can cause many health problems, including asthma, lung cancer, accelerated atherosclerosis and cardiovascular diseases (e.g., Pope et al. (2004), Wu, Jin and Carlsten (2018)). Therefore, $PM_{2.5}$ pollution has attracted attention of many governments, and much financial and human resource has been spent for improving air quality. In this effort, early detection of severe air pollution is especially important, because it can help governments to figure out the pollution source and to take proper and timely measures to prevent and/or control the pollution. To this end, some pollution surveillance systems have been developed to monitor pollution levels in different regions. For instance, the municipal government of Beijing in China developed a four-level, color-coded (i.e., blue, yellow, orange and red) pollution alert system in 2013 for collecting and monitoring the $PM_{2.5}$ data of the city, where the four colors denoted the four different levels of the $PM_{2.5}$ pollution with "blue" denoting the lowest level and "red" denoting the highest level. The city government issued the first "orange" alert on 11/30/2015 and the first

Received July 2021; revised December 2021.
Key words and phrases. Data correlation, dynamic process monitoring, environmental protection, pollution surveillance, sequential learning, statistical process control.

FIG. 1. *Pictures of a region around the Beijing National Stadium on* 11/23/2015 (*left panel*) *and* 11/30/2015 (*right panel*).

"red" alert on 12/05/2015 due to extreme air pollution in 2015. To reduce air pollution and minimize its impact on public health, the city government implemented several preventive measures, including closing some factories and schools, halting certain construction sites and urging city residents to minimize outdoor activities. To demonstrate the air pollution in 2015, Figure 1 presents two pictures of a region around the Beijing National Stadium taken on 11/23/2015 and 11/30/2015, that is, before and after the first "orange" alert on 11/30/2015. From the pictures, it can be seen that heavy smog lingered over the Beijing National Stadium and the surrounding area on 11/30/2015, and the air quality was much better one week before.

Environmental data often have complicated structure, including complex data distributions, serial correlation, seasonality, other dynamic patterns and more. Therefore, it is challenging to analyze them properly. In the atmospheric environment literature there are some existing methods on quantitative assessment of air quality (Liang et al. (2015, 2016), Seaman (2000)). For instance, Liang et al. (2015) used the local kernel smoothing procedure to assess $PM_{2.5}$ pollution in Beijing during the years 2010–2014. These existing methods, however, are retrospective and cannot effectively monitor the air quality sequentially over time. For proactive decision making, some researchers used the conventional statistical process control (SPC) charts for air pollution surveillance (e.g., Al-Rashed, Al-Mutairi and Attar (2019), Barratt et al. (2007), Chelani (2011)). These conventional SPC charts, such as the Shewhart, cumulative sum (CUSUM), exponentially weighted moving average (EWMA) and change-point detection (CPD) charts (cf. Qiu (2014)), were originally developed for monitoring production lines in the manufacturing industry and require the assumptions that process observations at different observation times are independent and identically distributed with a specific parametric distribution. In practice, however, these assumptions are rarely valid. For instance, observations of $PM_{2.5}$ concentration usually have seasonal variation (Jacob and Winner (2009), Zhao et al. (2009)), and serial correlation almost always exists in such sequential time series data. In the SPC literature it has been well demonstrated that the conventional control charts would not be reliable to use in cases when one or more of their assumptions are violated (e.g., Lee and Apley (2011), Xue and Qiu (2021)). Therefore, they would not be appropriate to use for air pollution surveillance.

The underlying process of an air quality index (e.g., $PM_{2.5}$ concentration) over time in a specific region can be regarded as a dynamic process in the sense that its distribution would change over time even when the index readings are at normal levels. Thus, air pollution surveillance is for sequential monitoring of a dynamic process whose observations could be serially correlated. For sequential process monitoring, SPC provides a major statistical tool. Besides the conventional SPC charts, many newer control charts have been developed for various applications, where the conventional model assumptions are invalid, which include nonparametric charts for monitoring processes whose distributions do not belong to any parametric distribution families (e.g., Chakraborti and Graham (2019), Qiu (2018)) and charts for monitoring processes with serially correlated observations (e.g., Apley and Tsung (2002),

Capizzi and Masarotto (2008), Qiu and Xie (2021)). Some control charts have been developed in the framework of dynamic screening system (DySS) for applications such as disease screening for individual patients (e.g., Qiu and Xiang (2014), You and Qiu (2020)). However, the DySS method was developed for monitoring many individual dynamic processes based on an in-control (IC) dataset that contains observed data of some IC processes. For instance, if sequential observations of a set of disease risk factors (e.g., cholesterol level, systolic blood pressure, diastolic blood pressure) of a given patient are regarded as observations of a dynamic process, then the DySS method is for monitoring many such processes of different patients. In the air pollution surveillance problem considered in the current paper, however, only one sequential process is involved, which is the sequential observations of certain air pollutants at a single place like Beijing. Also, the DySS method requires a quite large IC data set to be available in advance for estimating the IC longitudinal pattern of the individual dynamic processes, which is unavailable in the current problem. Therefore, for air pollution surveillance and other applications of dynamic process monitoring, new SPC methods are needed.

In this paper, we develop a new SPC chart for dynamic process monitoring, which should be effective for applications such as air pollution surveillance, where process distributions could be nonparametric and time-varying, and process observations could be serially correlated. This new method is described in detail in Section 2. Some numerical justifications of the new method are given in Section 3. It is applied to an air quality dataset collected in Beijing in Section 4. Some remarks conclude the article in Section 5.

2. New control chart for dynamic process monitoring. Our proposed new control chart can be described briefly as follows. First, it needs an IC data set to obtain an initial estimate of the IC longitudinal pattern of the dynamic process under monitoring. Second, at the current time point during online process monitoring, the observed data are first standardized using the estimated IC longitudinal pattern and then decorrelated with all historical data. Third, a control chart is then applied to the decorrelated data and makes a decision whether the process has a distributional shift at the current time point. If the decision is "yes," then stop process monitoring and communicate with the related personnel for a subsequent reaction to the signal. Otherwise, update the estimate of the IC longitudinal pattern using the IC data at the previous time point and the observed data at the current time point. These major steps are described in detail below.

2.1. *Initial estimation of the IC longitudinal pattern.* Let $\mathbf{y}(t) = (y_1(t), y_2(t), \ldots, y_p(t))'$ be a vector of p variables to monitor about a dynamic process at time t. Before online process monitoring, assume that an IC dataset $\mathcal{Y}_{IC} = \{\mathbf{y}(t_{-m_0+1}), \mathbf{y}(t_{-m_0+2}), \ldots, \mathbf{y}(t_0)\}$ of size m_0 is available. This IC dataset follows the multivariate nonparametric longitudinal model:

$$\mathbf{y}(t_j) = \boldsymbol{\mu}(t_j) + \boldsymbol{\epsilon}(t_j) \quad \text{for } j = -m_0 + 1, -m_0 + 2, \ldots, 0, \tag{1}$$

where $t_j \in [0, T]$ is the jth time point, $\boldsymbol{\mu}(t_j) = (\mu_1(t_j), \mu_2(t_j), \ldots, \mu_p(t_j))'$ is the mean of $\mathbf{y}(t_j)$ and $\boldsymbol{\epsilon}(t_j)$ is the p-dimensional zero-mean error term. In Model (1) the covariance structure is described by $V(t', t) = \text{Cov}(\boldsymbol{\epsilon}(t'), \boldsymbol{\epsilon}(t))$, for any $t', t \in [0, T]$. Besides the regularity condition that both $\boldsymbol{\mu}(t)$ and $V(t', t)$ are continuous functions, we do not impose any other assumptions on Model (1). Thus, it is flexible.

To obtain an initial estimate of $\boldsymbol{\mu}(t)$, we can compute the local linear kernel (LLK) estimates of all components of $\boldsymbol{\mu}(t)$ (cf. Fan and Gijbels (1996)). In matrix notation, let $\mathbf{Y} = (y_1(t_{-m_0+1}), \ldots, y_1(t_0), \ldots, y_p(t_{-m_0+1}), \ldots, y_p(t_0))'$, and $\mathbf{K} = \text{diag}\{K(\frac{t_j - t}{h_l}), j = -m_0 + 1, \ldots, 0, l = 1, \ldots, p\}$, where $K(\cdot)$ is a kernel function and $\{h_l, l = 1, \ldots, p\}$ are

bandwidths. Then, the initial estimator of $\boldsymbol{\mu}(t)$ can be obtained by the following LLK smoothing procedure:

$$\min_{\boldsymbol{\beta}\in R^{2p}} \big[\mathbf{Y} - (I_{p\times p}\otimes \mathbf{X})\boldsymbol{\beta}\big]'\mathbf{K}\big[\mathbf{Y} - (I_{p\times p}\otimes \mathbf{X})\boldsymbol{\beta}\big], \tag{2}$$

where $\otimes$ denotes the Kronecker product, $I_{p\times p}$ is the $p\times p$ identity matrix, $\boldsymbol{\beta} = (\beta_{01}, \beta_{11}, \ldots, \beta_{0p}, \beta_{1p})'$ are coefficients and $\mathbf{X} = ((1, t_{-m_0+1} - t)', (1, t_{-m_0+2} - t)', \ldots, (1, t_0 - t)')'$. The solution of (2) has the expression $\widehat{\boldsymbol{\beta}}^{(0)} = [\mathbf{Q}^{(0)}]^{-1}\mathbf{J}^{(0)}$, where $\mathbf{Q}^{(0)} = (I_{p\times p}\otimes \mathbf{X})'\mathbf{K}(I_{p\times p}\otimes \mathbf{X})$, and $\mathbf{J}^{(0)} = (I_{p\times p}\otimes \mathbf{X})'\mathbf{K}\mathbf{Y}$. The initial LLK estimate of $\boldsymbol{\mu}(t)$ is defined to be

$$\widehat{\boldsymbol{\mu}}^{(0)}(t) = \big[\widehat{\boldsymbol{\beta}}^{(0)}\big]'(I_{p\times p}\otimes \mathbf{e}_1), \tag{3}$$

where $\mathbf{e}_1 = (1, 0)'$. In the above LLK procedure the kernel function $K(\cdot)$ is usually chosen to be the Epanechnikov kernel function, that is, $K(u) = \frac{3}{4}(1-u^2)I(|u|\le 1)$, because of its good properties (Epsnečnikov (1969)). For the bandwidths $\{h_l, l = 1, 2, \ldots, p\}$, we suggest choosing them using the following modified cross-validation (MCV) procedure that was originally suggested by De Brabanter et al. (2011) for handling bandwidth selection in a univariate regression setup with correlated data. By this approach the bandwidths $\{h_l, l = 1, \ldots, p\}$ can be chosen by minimizing the following MCV score:

$$\mathrm{MCV}(h_1, h_2, \ldots, h_p) = \frac{1}{m_0}\sum_{i=-m_0+1}^{0} \big[\mathbf{y}(t_i) - \widehat{\boldsymbol{\mu}}_{-i}(t_i)\big]'\big[\mathbf{y}(t_i) - \widehat{\boldsymbol{\mu}}_{-i}(t_i)\big],$$

where $\widehat{\boldsymbol{\mu}}_{-i}(t_i)$ is the leave-one-out estimate of $\boldsymbol{\mu}(t_i)$ by (2) when the observation $\mathbf{y}(t_i)$ is excluded in the computation and when the kernel function $K(\cdot)$ is modified to be

$$K_\varepsilon(u) = \frac{4}{4 - 3\varepsilon - \varepsilon^3}\begin{cases} \dfrac{3}{4}\big(1-u^2\big)I\big(|u|\le 1\big) & \text{when } |u|\ge \varepsilon, \\ \dfrac{3(1-\varepsilon^2)}{4\varepsilon}|u| & \text{when } |u| < \varepsilon, \end{cases}$$

where $\varepsilon \in (0, 1)$ is a constant.

After $\widehat{\boldsymbol{\mu}}^{(0)}(t)$ is obtained, the initial estimate of the covariance function $V(t', t)$ can be defined to be the following weighted moment estimate: for any $t', t \in [0, T]$,

$$\widehat{V}(t', t) = \frac{\sum_{j=-m_0+1}^{0}\sum_{k=-m_0+1}^{0}(\mathbf{y}(t_j) - \widehat{\boldsymbol{\mu}}^{(0)}(t_j))(\mathbf{y}(t_k) - \widehat{\boldsymbol{\mu}}^{(0)}(t_k))'K(\frac{t_j - t'}{q})K(\frac{t_k - t}{q})}{\sum_{j=-m_0+1}^{0}\sum_{k=-m_0+1}^{0}K(\frac{t_j - t'}{q})K(\frac{t_k - t}{q})}, \tag{4}$$

where the kernel function $K(\cdot)$ is still chosen to be the Epanechnikov kernel function and the bandwidth q is chosen by minimizing the following cross-validated prediction error (PE):

$$\mathrm{PE}(q) = \frac{1}{m_0}\sum_{i=-m_0+1}^{0} \big(\mathbf{y}(t_i) - \widehat{\mathbf{y}}_{-i}(t_i)\big)'\big(\mathbf{y}(t_i) - \widehat{\mathbf{y}}_{-i}(t_i)\big),$$

where $\widehat{\mathbf{y}}_{-i}(t_i)$ is the predicted value of $\mathbf{y}(t_i)$ obtained by the kriging method (cf. Cressie and Wikle (2011)) described below. For $-m_0 + 1 \le i \le 0$, let $\mathbf{Y}_{-i}$ be the matrix with $\{\mathbf{y}(t_k), |t_k - t_i| \le q \text{ and } k \ne i\}$ as its columns and $\widehat{\mathbf{e}}_{-i}$ be the corresponding residual matrix. Then, the predicted value of $\mathbf{y}(t_i)$ is defined to be $\widehat{\mathbf{y}}_{-i}(t_i) = \widehat{\boldsymbol{\mu}}^{(0)}(t_i) + \widehat{V}_{i,-i}'\widehat{V}_{-i}^{-1}\widehat{\mathbf{e}}_{-i}$, where $\widehat{V}_{i,-i}$ is the estimated covariance matrix between $\mathbf{y}(t_i)$ and $\mathbf{Y}_{-i}$, $\widehat{V}_{-i}$ is the estimated covariance matrix of $\mathbf{Y}_{-i}$, and both of them can be computed from $\widehat{V}(t', t)$ defined in (4). It should be pointed out that the estimate $\widehat{V}(t', t)$ may not be a positive semidefinite matrix for each (t, t'). Thus, it may not be a legitimate covariance matrix. In this paper, we suggest using the matrix modification method discussed in Higham (1988) to modify it properly to be a valid covariance matrix.

2.2. *Online monitoring of dynamic processes.* Next, we discuss online monitoring of dynamic processes. Assume that observations of a process to monitor follow the model

$$\mathbf{y}(t_n) = \boldsymbol{\mu}(t_n) + \boldsymbol{\epsilon}(t_n) \quad \text{for } n \geq 1,$$

where $t_n > T$ are observation times. To account for possible seasonality of the process, it is assumed that the time interval $[0, T]$ of the IC data contains a whole season, and the mean function $\boldsymbol{\mu}(t)$ is periodic in time with the period T when the process is IC, so that $\boldsymbol{\mu}(t) = \boldsymbol{\mu}(t^*)$, where $t = t^* + lT$, $t^* \in [0, T]$ and $l \geq 1$ is an integer. In this setup the regular longitudinal pattern of the process in $[0, T]$ can be used as a baseline pattern, and the proposed control chart described below is for detecting a shift in the longitudinal pattern of the process from this baseline pattern.

The process observations $\{\mathbf{y}(t_n), n \geq 1\}$ could be serially correlated in applications. Because the conventional control charts are designed for cases with uncorrelated process observations only, we try to decorrelate them properly before a control chart is used for process monitoring. To this end, it is often reasonable, in practice, to assume that the correlation between two process observations is weaker when their observation times are farther away. Thus, we assume that $\text{Cov}(\mathbf{y}(t_i), \mathbf{y}(t_j)) = 0$, when $|t_i - t_j| > b_{\max}$ where $b_{\max} > 0$ denotes the time range of serial correlation. Based on this assumption, at the current time point t_n, $\mathbf{y}(t_n)$ should be decorrelated with its previous $b_{\max}$ observations. Because data decorrelation needs to be implemented at each observation time during process monitoring, reduction of computing time is important. To accomplish that, the concept of *spring length*, suggested by Chatterjee and Qiu (2009), will be used in the proposed method. This concept is based on the restarting mechanism of a CUSUM chart (Qiu (2014), Chapter 4). At a given time point, if the CUSUM chart finds that the likelihood to have a process distributional shift is small, then its charting statistic would be reset to 0 and all process observations collected at the current and previous observation times would be ignored in subsequent process monitoring. At the current time point t_n, the spring length b_n is then defined to be the number of observation times from the previous reset of the charting statistic to the current time t_n. So, based on the concept of spring length, $\mathbf{y}(t_n)$ only needs to be decorrelated with the previous b_{n-1} observations since process observations collected before the time $t_{n-b_{n-1}}$ would not be used in process monitoring at t_n, where b_{n-1} is used here because b_n is not defined yet before the chart makes a decision about the process status at t_n. Let $\phi_n = \min\{b_{\max}, b_{n-1}\}$. Then, at the current time point t_n, we need to decorrelate $\mathbf{y}(t_n)$ with the previous ϕ_n observations. Since b_{n-1} is often a single-digit integer (cf. You and Qiu (2019)), much computation could be saved by using ϕ_n in the sequential data decorrelation.

In cases when monitoring a conventional process with time-independent IC process distribution, Li and Qiu (2017) proposed a sequential data decorrelation and standardization procedure. This procedure has been generalized for monitoring dynamic processes in this paper and the generalized procedure is described below:

- When $n = 1$, the decorrelated and standardized observation of $\mathbf{y}(t_1)$ is defined to be $\widehat{\mathbf{y}}^*(t_1) = [\widehat{V}(t_1, t_1)]^{-1/2}(\mathbf{y}(t_1) - \widehat{\boldsymbol{\mu}}^{(0)}(t_1))$.
- When $n > 1$, define $\widehat{\mathbf{e}}_{n-1} = ((\mathbf{y}(t_{n-\phi_n}) - \widehat{\boldsymbol{\mu}}^{(n-1)}(t_{n-\phi_n}))', \ldots, (\mathbf{y}(t_{n-1}) - \widehat{\boldsymbol{\mu}}^{(n-1)}(t_{n-1}))')'$ and $\mathbf{W}_n = [\mathbf{y}(t_{n-\phi_n}), \mathbf{y}(t_{n-\phi_n+1}), \ldots, \mathbf{y}(t_n)]'$, where the estimate $\widehat{\boldsymbol{\mu}}^{(n-1)}(t)$ is defined in Expression (7) below. The estimates of $\text{Cov}(\mathbf{W}_n, \mathbf{W}_n)$ and $\text{Cov}(\mathbf{W}_{n-1}, \mathbf{y}(t_n))$ are denoted as $\widehat{\boldsymbol{\Sigma}}_{n,n}$ and $\widehat{\boldsymbol{\Sigma}}_{n-1,n}$, respectively, and we have

$$\widehat{\boldsymbol{\Sigma}}_{n,n} = \begin{pmatrix} \widehat{V}(t_{n-\phi_n}, t_{n-\phi_n}) & \cdots & \widehat{V}(t_{n-\phi_n}, t_n) \\ \vdots & \ddots & \vdots \\ [\widehat{V}(t_{n-\phi_n}, t_n)]' & \cdots & \widehat{V}(t_n, t_n) \end{pmatrix} =: \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{n-1,n-1} & \widehat{\boldsymbol{\Sigma}}_{n-1,n} \\ \widehat{\boldsymbol{\Sigma}}'_{n-1,n} & \widehat{V}(t_n, t_n) \end{pmatrix}.$$

Then, the decorrelated and standardized observation at time t_n is defined to be

$$\widehat{\mathbf{y}}^*(t_n) = \widehat{D}_n^{-1/2}\big[-\widehat{\mathbf{\Sigma}}'_{n-1,n}\widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1}\widehat{\mathbf{e}}_{n-1} + \big(\mathbf{y}(t_n) - \widehat{\boldsymbol{\mu}}^{(n-1)}(t_n)\big)\big],$$

where $\widehat{D}_n = \widehat{V}(t_n, t_n) - \widehat{\mathbf{\Sigma}}'_{n-1,n}\widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1}\widehat{\mathbf{\Sigma}}_{n-1,n}$, and $\widehat{\mathbf{\Sigma}}^{-1}_{n,n}(n)$ can be computed recursively by the following formula: for $n \geq 2$,

$$\widehat{\mathbf{\Sigma}}^{-1}_{n,n} = \begin{pmatrix} \widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1} + \widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1}\widehat{\mathbf{\Sigma}}_{n-1,n}\widehat{D}^{-1}_{n-1}\widehat{\mathbf{\Sigma}}'_{n-1,n}\widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1}, & -\widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1}\widehat{\mathbf{\Sigma}}_{n-1,n}\widehat{D}^{-1}_{n-1} \\ -\widehat{D}^{-1}_{n-1}\widehat{\mathbf{\Sigma}}'_{n-1,n}\widehat{\mathbf{\Sigma}}^{-1}_{n-1,n-1}, & \widehat{D}^{-1}_{n-1} \end{pmatrix}.$$

Then, the decorrelated data $\{\widehat{\mathbf{y}}^*(t_1), \ldots, \widehat{\mathbf{y}}^*(t_n)\}$ would be asymptotically uncorrelated with each other, and each decorrelated observation vector would have the asymptotic mean of $\mathbf{0}$ and the asymptotic covariance matrix of $I_{p\times p}$.

After the data decorrelation and standardization, we are ready to use a control chart for online process monitoring. To this end, a nonparametric control chart will be considered, since the process distribution is not restricted to any parametric distribution family in this paper. In the SPC literature there are many nonparametric control charts available (cf. Chakraborti and Graham (2019), Qiu (2018)). Theoretically speaking, most of them can be used here. In this paper, we use the antirank-based nonparametric chart that was suggested by Qiu and Hawkins (2003), since this chart was shown to have a reasonably good performance in various different cases (cf. Qiu (2018)). Then, our proposed process monitoring procedure based on this chart is described below.

Let $\mathbf{Z}(n) = (\hat{y}^*_1(t_n), \hat{y}^*_2(t_n), \ldots, \hat{y}^*_p(t_n), 0)'$ be a combination of the decorrelated observation $\widehat{\mathbf{y}}^*(t_n)$ and the IC mean 0 of each of its component. The antirank vector of $\mathbf{Z}(n)$ is denoted as $\mathbf{A}(n) = (A_1(n), A_2(n), \ldots, A_p(n), A_{p+1}(n))'$. Then, by the definition of antiranks, the $A_1(n)$th component of $\mathbf{Z}(n)$ is the smallest among all $p+1$ components, the $A_2(n)$th component of $\mathbf{Z}(n)$ is the second smallest component and so forth. Thus, $\mathbf{A}(n)$ is the vector of indices of the order statistics of the $p+1$ components of $\mathbf{Z}(n)$. Qiu and Hawkins (2003) demonstrated that the first antirank $A_1(n)$ was sensitive to "downward" shifts in the mean of $\mathbf{Z}(n)$, the last antirank $A_{p+1}(n)$ was sensitive to "upward" mean shifts and the pair $(A_1(n), A_{p+1}(n))$ was sensitive to arbitrary mean shifts. For this reason the proposed control chart will be constructed, based on the pair $(A_1(n), A_{p+1}(n))$, which can take $p(p+1)$ possible values in the set $\mathcal{S} = \{(i, j), 1 \leq i \neq j \leq p+1\}$. Let $\mathbf{g}(n)$ be a $p(p+1)$-dimensional vector whose lth element equal to 1 when $(A_1(n), A_{p+1}(n))$ equals the lth element of $\mathcal{S}$ and whose remaining elements are all 0, and $\mathbf{f}$ be the IC mean of $\mathbf{g}(n)$. At the previous time point t_{n-1}, $\mathbf{f}$ can be estimated by the relative frequencies of the observed values of $\{(A_1(i), A_{p+1}(i)), -m_0 + 1 \leq i \leq n-1\}$, and the estimate is denoted as $\widehat{\mathbf{f}}^{(n-1)}$. Then, the following CUSUM chart can be constructed based on the comparison of the observed and expected values of $\mathbf{g}(n)$:

$$C_n = (\mathbf{S}^{\text{obs}}_n - \mathbf{S}^{\text{exp}}_n)'\big[\text{diag}(\mathbf{S}^{\text{exp}}_n)\big]^{-1}(\mathbf{S}^{\text{obs}}_n - \mathbf{S}^{\text{exp}}_n), \tag{5}$$

where

$$\begin{cases} \mathbf{S}^{\text{obs}}_n = \mathbf{S}^{\text{exp}}_n = \mathbf{0} & \text{if } U_n \leq \rho \\ \mathbf{S}^{\text{obs}}_n = \big[\mathbf{S}^{\text{obs}}_{n-1} + \mathbf{g}(n)\big](U_n - \rho)/U_n & \text{if } U_n > \rho \\ \mathbf{S}^{\text{exp}}_n = \big[\mathbf{S}^{\text{exp}}_{n-1} + \widehat{\mathbf{f}}^{(n-1)}\big](U_n - \rho)/U_n & \text{if } U_n > \rho, \end{cases}$$

$$\begin{aligned} U_n = &\big[(\mathbf{S}^{\text{obs}}_{n-1} - \mathbf{S}^{\text{exp}}_{n-1}) + (\mathbf{g}(n) - \widehat{\mathbf{f}}^{(n-1)})\big]'\big[\text{diag}(\mathbf{S}^{\text{exp}}_{n-1} + \widehat{\mathbf{f}}^{(n-1)})\big]^{-1} \\ &\times \big[(\mathbf{S}^{\text{obs}}_{n-1} - \mathbf{S}^{\text{exp}}_{n-1}) + (\mathbf{g}(n) - \widehat{\mathbf{f}}^{(n-1)})\big], \end{aligned}$$

diag($\boldsymbol{a}$) denotes a diagonal matrix with the diagonal elements being the related elements of the vector $\boldsymbol{a}$ and $\rho > 0$ is a prespecified small constant. Then, the chart gives a signal when

$$C_n > \gamma, \tag{6}$$

where $\gamma > 0$ is a control limit. In (5), when ρ is chosen to be 0, $\mathbf{S}_n^{\text{obs}}$ is just the vector of cumulative counts of the observed elements in $\mathcal{S}$ by the current time point t_n, and $\mathbf{S}_n^{\text{exp}}$ is the vector of cumulative expected counts. The use of the small constant ρ is for setting up the restarting mechanism that the charting statistic C_n in (5) would be reset to 0 each time when $U_n \le \rho$, since U_n measures the difference between the cumulative counts of the observed elements in $\mathcal{S}$ and the related expected counts and a mean shift in the original process is unlikely when $U_n \le \rho$. The restarting mechanism mentioned here has been used by the concept of spring length discussed earlier.

In the nonparametric dynamic process monitoring (NDPM) chart (5)–(6), the constant ρ is usually prespecified. It has been shown in the literature that large values of ρ are good for detecting large shifts and small values are good for detecting small shifts (Qiu and Hawkins (2003)). Once ρ is prespecified, the control limit γ can be determined by a Monte Carlo simulation to achieve a prespecified value of ARL_0 in cases when observation times are equally spaced. More specifically, $\widehat{\mathbf{f}}^{(0)}$ can be obtained from the decorrelated and standardized IC data using the related relative frequencies. Then, for each simulation run the observations $\{\mathbf{g}(n), n \ge 1\}$ can be generated from a multinomial distribution specified by the IC distribution $\widehat{\mathbf{f}}^{(0)}$. For a given value of γ, the NDPM chart (5)–(6) is then applied to the observations $\{\mathbf{g}(n), n \ge 1\}$, and the run length (RL) value, defined to be the number of observation times from the beginning of process monitoring to the signal time, is recorded. This simulation is repeated for M times, and the average of the M RL values provides an estimate of the actual ARL_0 value of the chart, denoted as $ARL_0(\gamma)$. Then, γ can be searched by a numerical algorithm (e.g., the bisection algorithm) so that the assumed ARL_0 value is reached by $ARL_0(\gamma)$ with a given accuracy; see Qiu and Hawkins (2003) for a more detailed discussion. In cases when observation times are unequally spaced, ARL_0 would not be appropriate for measuring the IC performance of a control chart, because ARL_0 focuses only on the number of observation times between the start of online process monitoring and the signal time of the chart. In such cases we should use the IC average time to signal, denoted as ATS_0, instead (cf. Qiu and Xiang (2014)).

2.3. *Update of the IC parameter estimates.* At the current observation time t_n, if the condition (6) is satisfied, then a signal of shift is given by the chart. Otherwise, the observation $\mathbf{y}(t_n)$ should be combined with the existing IC dataset, and the estimates of the IC parameters should be updated using the combined IC dataset. In this latter case the formulas for updating the IC parameter estimates are given below.

First, for the estimate of the IC mean function $\boldsymbol{\mu}(t)$, it can be calculated by the following updating formulas:

$$\widehat{\boldsymbol{\mu}}^{(n)}(t) = \big[\widehat{\boldsymbol{\beta}}^{(n)}\big]'(I_{p\times p} \otimes \mathbf{e}_1), \tag{7}$$

where $\widehat{\boldsymbol{\beta}}^{(n)} = [\mathbf{Q}^{(n)}]^{-1}\mathbf{J}^{(n)}$, and $\mathbf{Q}^{(n)}$ and $\mathbf{J}^{(n)}$ can be updated recursively by

$$\mathbf{Q}^{(n)} = \mathbf{Q}^{(n-1)} + [\mathbf{k}_n \otimes \mathbf{X}_n](I_{p\times p} \otimes \mathbf{X}_n)',$$
$$\mathbf{J}^{(n)} = \mathbf{J}^{(n-1)} + \big[\mathbf{k}_n \mathbf{y}(t_n)\big] \otimes \mathbf{X}_n,$$

$\mathbf{X}_n = (1, t_n^* - t)'$, $\mathbf{k}_n = \text{diag}(k_{h_1}(t_n^* - t), k_{h_2}(t_n^* - t), \ldots, k_{h_p}(t_n^* - t))$, and $k_{h_l}(t_n^* - t) = K((t_n^* - t)/h_l)/h_l$, for $l = 1, 2, \ldots, p$.

Since the serial correlation is allowed to be nonstationary in this paper and at the next observation time t_{n+1} we only need to estimate the covariance between $\mathbf{y}(t_{n+1})$ and its previous $b_{\max}$ observations, these covariance estimates can be defined by

$$\widehat{V}(t_{n+1-j}, t_{n+1}) = \Bigg(\sum_{l=n+1-j-w}^{n+1-j} \sum_{k=n+1-w}^{n+1} \big(\mathbf{y}(t_l) - \widehat{\boldsymbol{\mu}}^{(n)}(t_{n+1-j})\big)\big(\mathbf{y}(t_k) - \widehat{\boldsymbol{\mu}}^{(n)}(t_{n+1})\big)' \times K\Big(\frac{t_l - t_{n+1-j}}{q}\Big) K\Big(\frac{t_k - t_{n+1}}{q}\Big)\Bigg) \Bigg/ \Bigg(\sum_{l=n+1-j-w}^{n+1-j} \sum_{k=n+1-w}^{n+1} K\Big(\frac{t_l - t_{n+1-j}}{q}\Big) K\Big(\frac{t_k - t_{n+1}}{q}\Big)\Bigg), \tag{8}$$

where $0 \le j \le b_{\max}$, w is a prespecified window size, and $K(\cdot)$ and q are the same as those in (4).

Finally, the estimate of the IC distribution of $(A_1(n), A_{p+1}(n))$ can also be updated by the following formula: for $n \ge 1$,

$$\widehat{\mathbf{f}}^{(n)} = \frac{m_0 + n - 1}{m_0 + n} \widehat{\mathbf{f}}^{(n-1)} + \frac{1}{m_0 + n} \mathbf{g}(n).$$

2.4. *Practical guidelines on parameter selection.* There are a number of parameters in the NDPM chart (5)–(6) that need to be selected properly in advance. To this end, some practical guidelines based on extensive numerical studies are given below.

On selection of $b_{\max}$: In the NDPM chart, two process observations are assumed to be uncorrelated when their observation times are, at least, $b_{\max}$ apart. In practice, $b_{\max}$ is often unknown and needs to be prespecified. Of course, it is better to choose a larger value for $b_{\max}$, but the related computation in data decorrelation would be more extensive. Based on our extensive numerical experience, the performance of the NDPM chart would be reasonably good when we choose $b_{\max} \in [10, 20]$.

On selection of w: The window size w is used when computing the estimate $\widehat{V}(t_{n+1-j}, t_{n+1})$ in (8). Based on our numerical experience, it can be chosen to be $w = \alpha \times b_{\max}$ with $\alpha \in [4, 6]$.

3. Numerical justifications of the proposed method. In this section we provide some numerical justifications of the proposed method using Monte Carlo simulations. For simplicity, the IC observation times are assumed to be $\{t_j = (m_0 + j)\tau, j = -m_0 + 1, -m_0 + 2, \ldots, 0\}$, which are equally spaced in the baseline time interval $[0, T] = [0, 1]$, where $\tau = 1/m_0$ is the basic time unit. For the IC model (1) it is assumed that $p = 3$ (i.e., there are three variables to monitor), and the following six different cases are considered. In Cases I–III the process mean functions are time-independent and assumed to be $\boldsymbol{\mu}(t) = (0, 0, 0)'$, for any $t \in [0, 1]$. In Case I, the random errors $\{\boldsymbol{\epsilon}(t_j), j \ge 1\}$ are assumed to be independent and identically distributed (i.i.d.) at different observation times, and each random error vector has the distribution $N_3(\mathbf{0}, I_{3\times 3})$. In Case II, the random errors $\{\boldsymbol{\epsilon}(t_j), j \ge 1\}$ are assumed to follow the vector AR(1) model $\boldsymbol{\epsilon}(t_j) = 0.2\boldsymbol{\epsilon}(t_{j-1}) + \boldsymbol{\eta}(t_j)$, for $j \ge 1$, where $\boldsymbol{\epsilon}(t_0) = \mathbf{0}$ and $\{\boldsymbol{\eta}(t_j), j \ge 1\}$ are i.i.d. at different time points, each component of $\boldsymbol{\eta}(t_j)$ has the standardized chi-square distribution $(\chi_3^2 - 3)/\sqrt{6}$ and the covariance matrix of $\boldsymbol{\eta}(t_j)$ is

$$\begin{pmatrix} 1 & 0.2 & 0.2^2 \\ 0.2 & 1 & 0.2 \\ 0.2^2 & 0.2 & 1 \end{pmatrix}.$$

In Case III, the random error vector is $\boldsymbol{\epsilon}(t_j) = \text{diag}(1, \exp(t), \frac{1}{1+t})\boldsymbol{\epsilon}^*(t_j)$, for each j, where $\boldsymbol{\epsilon}^*(t_j)$ follows the vector time series model $\boldsymbol{\epsilon}^*(t_j) = 0.2t_j\boldsymbol{\epsilon}^*(t_{j-1}) + \boldsymbol{\eta}(t_j)$ and $\boldsymbol{\eta}(t_j)$ has the same distribution as that in Case II. In Cases IV–VI, the process observations are generated in the same way as those in Cases I–III, respectively, except that the IC mean functions are assumed to be $\boldsymbol{\mu}(t) = (0, t, \sin(2\pi t))'$, which are time-dependent. Obviously, Case I is the conventional case considered in the SPC literature with i.i.d. process observations and the normal IC distribution. Cases II and III consider cases with stationary and nonstationary serial correlation, respectively, and with a nonnormal error distribution. While the IC process distribution considered in these first three cases is time-independent, the IC process distribution considered in Cases IV–VI is time-varying, and these three cases are considered for studying the impact of the time-varying IC distribution on the performance of the related process monitoring methods.

Besides the proposed method NDPM, we also consider six alternative methods for comparison purposes, including three simplified versions of NDPM, the method by Qiu and Hawkins (2003) and two machine learning methods by Sukchotrat, Kim and Tsung (2010) and He, Jiang and Deng (2018). The first simplified version of NDPM, denoted as NDPM-D-S, is the same as NDPM, except that the covariance structure is assumed to be stationary. In the label NDPM-D-S, "D" implies that the dynamic nature of the process under monitoring is considered in the chart, and "S" denotes stationary serial correlation. The second simplified version, denoted as NDPM-ND-NS, assumes that the IC process distribution is time-independent (i.e., nondynamic), and the other setups are the same as those for NDPM. The third simplified version, denoted as NDPM-ND-S, assumes that the process under monitoring is nondynamic and the serial correlation is stationary. The method by Qiu and Hawkins (2003) is based on the first and last antiranks of the p variables. This method, which is denoted as AR, assumes that process observations at different observation times are i.i.d. when the process is IC. The machine learning method by Sukchotrat, Kim and Tsung (2010) is based on the K-nearest-neighbor (KNN) data description procedure. It uses the average distance between a given observation and its k nearest observations in the IC dataset as the charting statistic. Its control limit is chosen by a bootstrap procedure from the IC dataset. This method is denoted as KNN. The method by He, Jiang and Deng (2018), denoted as DSVM, uses the support vector machine (SVM) framework. Its control limit is also determined by a bootstrap procedure from the IC data.

In all simulation examples we assume that the nominal ARL_0 value is 200 for all control charts. By the suggestion in Sukchotrat, Kim and Tsung (2010), the number of nearest observations in KNN is chosen to be $k = 30$. In DSVM, the moving window size is chosen to be 10, as suggested by He, Jiang and Deng (2018). The constant ρ (cf. the expression after (5)) in the CUSUM charts of NDPM, NDPM-D-S, NDPM-ND-NS, NDPM-ND-S and AR is chosen to be 0.5, if there is no further specifications. In NDPM, the parameter $b_{\max}$ is chosen to be 15, and the moving window size w is chosen to be $5b_{\max}$, as suggested in Section 2.4.

3.1. *Evaluation of the IC performance.* We first evaluate the IC performance of the related methods. To compute the actual ARL_0 value of a chart, an IC dataset of size m_0 is first generated. Then, a control chart is applied to a sequence of 2000 IC process observations for online process monitoring, and its RL value is recorded. The online process monitoring is then repeated for 1000 times, and the average of the 1000 RL values is used as an estimate of the actual conditional ARL_0 value, conditional on the IC data. Finally, to obtain an estimate of the actual unconditional ARL_0 value, all steps described above, starting from the generation of the IC data to computation of the estimate of the actual conditional ARL_0 value, are repeated for 100 times. The actual (unconditional) ARL_0 value of the chart is then estimated by the average of the 100 estimates of the actual conditional ARL_0 value. The IC

TABLE 1
Actual ARL_0 values and their standard errors (in parentheses) of seven control charts when their nominal ARL_0 values are fixed at 200 and the IC sample size m_0 is 500

Case	NDPM	NDPM-D-S	NDPM-ND-NS	NDPM-ND-S	AR	KNN	DSVM
I	188 (3.55)	193 (3.43)	191 (3.30)	197 (3.62)	195 (2.89)	166 (4.33)	178 (5.23)
II	184 (3.58)	192 (3.48)	190 (3.27)	195 (3.56)	136 (2.40)	147 (4.93)	152 (5.03)
III	182 (3.11)	153 (3.33)	189 (3.77)	162 (3.37)	120 (3.10)	119 (3.55)	104 (3.32)
IV	188 (3.25)	194 (3.40)	105 (3.28)	147 (3.00)	63 (1.30)	125 (3.19)	149 (4.02)
V	185 (3.37)	193 (3.38)	118 (2.40)	160 (3.04)	61 (1.28)	114 (3.45)	134 (3.11)
VI	183 (3.97)	152 (3.29)	140 (3.76)	138 (3.24)	64 (1.34)	79 (2.92)	94 (2.55)

sample size m_0 is first fixed at 500. The results of the estimated actual ARL_0 values of the seven charts in various cases are presented in Table 1, along with their standard errors. From the table we can have the following conclusions: (i) The charts AR, KNN and DSVM have a reasonable performance in Case I, when the process observations are i.i.d. with a normal distribution, but they are unreliable to use in all other cases when some or all of these assumptions are violated because their estimated actual ARL_0 values are substantially different from the nominal ARL_0 level of 200 in these cases. (ii) The chart NDPM-ND-NS performs well in Cases I–III when its model assumption of nondynamic IC process distribution is valid, but its performance is quite poor in Cases IV–VI when this assumption is violated. (iii) The chart NDPM-D-S performs well in Cases I, II, IV and V, when its assumption of stationary serial correlation is valid, and quite poorly in Cases III and VI when the serial correlation is actually nonstationary. (iv) The chart NDPM-ND-S performs well in Cases I and II, when its assumptions of nondynamic IC process distribution and stationary serial correlation are valid, and quite poorly in all other cases when either or both of these assumptions are violated. (v) As a comparison, the chart NDPM has a reasonably good performance in all cases considered, since its estimated actual ARL_0 values are always within 10% of the nominal ARL_0 level. It should be noticed that observation times are equally spaced in this example. In cases when observation times are unequally spaced, the sampling rate to collect process observations is steady over time, and ATS_0 values are considered; our numerical studies confirm that similar conclusions to those mentioned above can be made about the IC performance of the related control charts.

The performance of the seven charts discussed above could be affected by the IC data size m_0. To study the impact of m_0 on their performance, we consider the following example in which m_0 changes among 200, 300, 400, 500, 800, 1000 and all other setups remain the same as those in the example of Table 1. The estimated actual ARL_0 values of the seven charts in such cases are presented in Figure 2. From the figure we can have the following conclusions. First, the IC performance of the charts AR, KNN and DSVM improves in Case I when m_0 increases, the chart AR has a reasonably reliable IC performance when $m_0 \geq 300$ and the charts KNN and DSVM have a quite reliable IC performance when $m_0 > 500$. But, in all other cases, the improvement of their IC performance is minimal when m_0 increases. Second, the IC performance of the charts NDPM-D-S, NDPM-ND-NS and NDPM-ND-S improves when m_0 increases in cases when their model assumptions are all valid (e.g., in Cases I, II, IV and V for the chart NDPM-D-S). In cases when some of their model assumptions are invalid, the improvement is minimal. Third, the IC performance of NDPM improves in all cases considered when m_0 increases, and its IC performance is quite reliable when $m_0 \geq 400$ since its actual ARL_0 values are within 10% of the nominal ARL_0 value of 200 in such cases.

The results in Figure 2 and Table 1 are for cases when the number of variables, p, is three. Of course, the necessary IC sample size m_0 would depend on p. To investigate this, let us

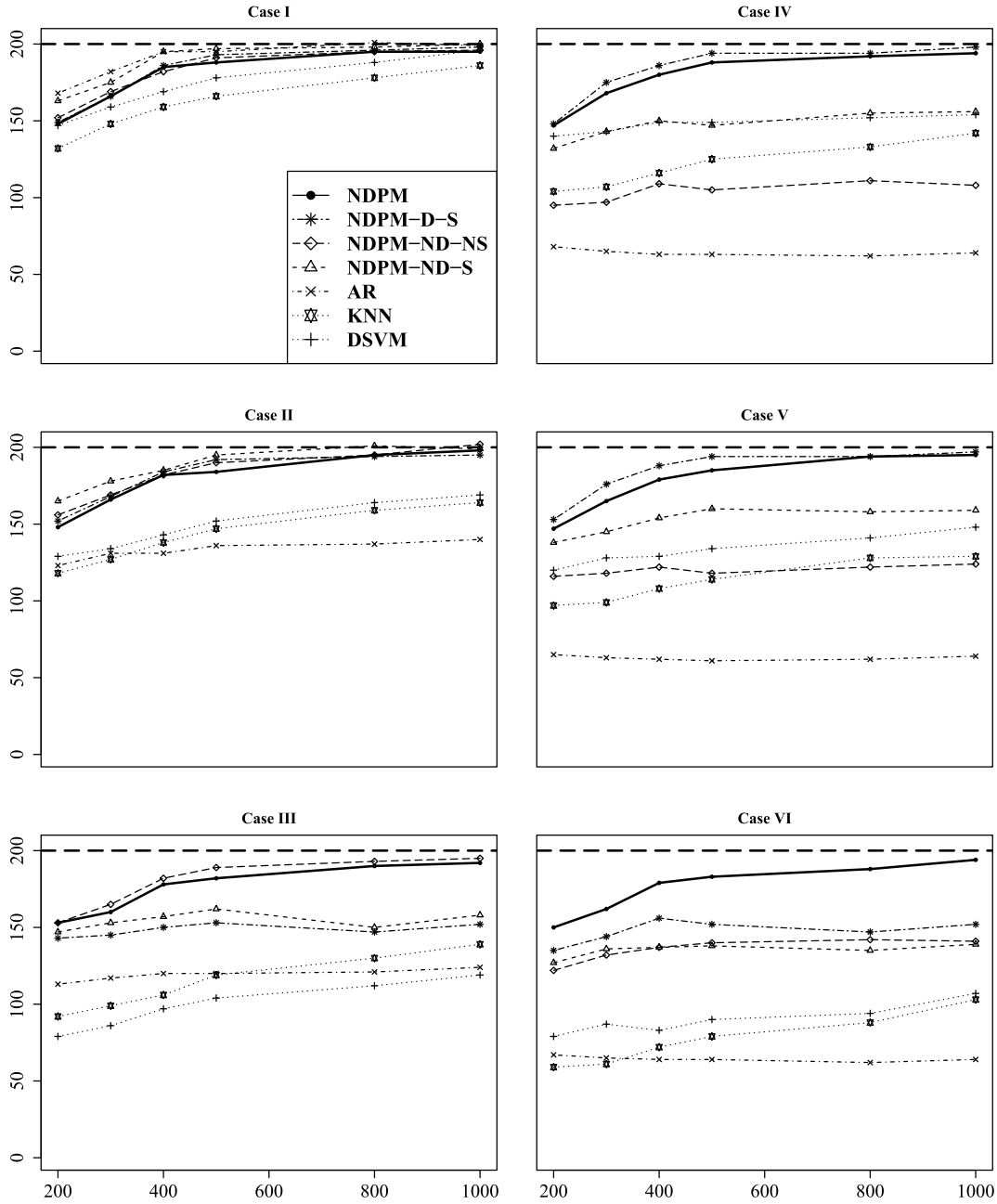

FIG. 2. *Estimated actual ARL_0 values of seven control charts when their nominal ARL_0 values are fixed at* 200, *the IC sample size m_0 changes among* 200, 300, 400, 500, 800 *and* 1000 *and other setups remain unchanged from those in the example of Table* 1.

consider a case extended from Case V to a p-dimensional case as follows. In the new case, the p-dimensional random errors $\{\boldsymbol{\epsilon}(t_j), j \geq 1\}$ are generated as described in Case V. For the mean vector $\boldsymbol{\mu}(t)$, we choose $\mu_{3\ell+1}(t) = \mu_1(t)$, $\mu_{3\ell+2}(t) = \mu_2(t)$ and $\mu_{3\ell+3}(t) = \mu_3(t)$ for $\ell \geq 1$, where $\mu_1(t)$, $\mu_2(t)$ and $\mu_3(t)$ are the same as those in Case V. Because the IC process is dynamic and the serial correlation is stationary in this case, we choose to study the IC performance of the charts NDPM-D-S and NDPM only. When p changes among $\{3, 5, 7\}$ and other setups are the same as those in Figure 2, their estimated actual ARL_0 values are

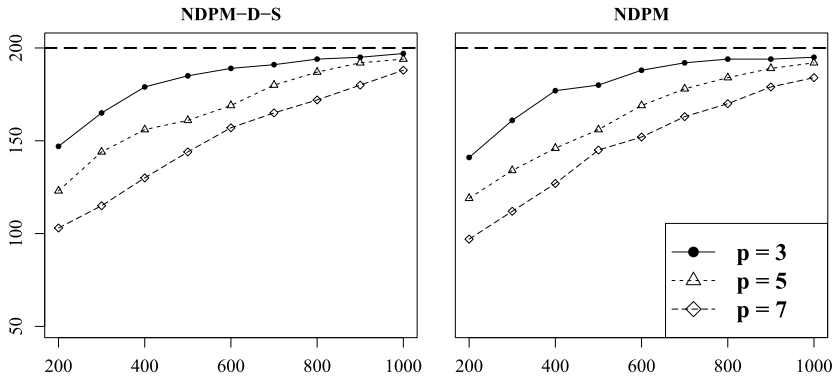

FIG. 3. *Estimated actual ARL_0 values of the charts NDPM-D-S and NDPM in a p-dimensional case extended from Case V when the nominal ARL_0 value of each chart is fixed at* 200 *and p changes among* $\{3, 5, 7\}$.

presented in Figure 3. From the figure it can be seen that the necessary IC sample size m_0 should indeed be larger to have a reliable IC performance of these two charts when p is larger.

3.2. *Evaluation of the OC performance.* Next, we evaluate the OC performance of the related charts in cases when m_0 is fixed at 500. To this end, a mean shift starting at the beginning of process monitoring is considered, and the shifted mean becomes $\boldsymbol{\mu}_1(t) = \boldsymbol{\mu}(t) + \delta(\sigma_1(t), \sigma_2(t), \sigma_3(t))'$, where $\sigma_j(t)$ is the standard deviation of $y_j(t)$, for $j = 1, 2, 3$ and δ is a constant that changes among 0.2, 0.4, 0.6, 0.8 and 1. To make the comparison among different charts fair, the control limits of the related charts have been adjusted properly so that their actual ARL_0 values all equal to the nominal ARL_0 value of 200. First, we let the procedure parameters of the charts be the same as those in Table 1. In such cases the computed ARL_1 values of the seven charts are presented in Figure 4. From the figure we can have the following conclusions: (i) The chart NDPM has the best performance in Case VI when the process under monitoring is dynamic with nonstationary serial correlation. (ii) The charts NDPM and NDPM-ND-NS perform better than the other control charts in Case III when the process under monitoring is nondynamic and the serial correlation is nonstationary. (iii) The charts NDPM and NDPM-D-S perform better than the other methods in Case V when the process is dynamic and the serial correlation is stationary. (iv) The charts NDPM-ND-S NDPM-D-S and NDPM-ND-NS perform better than the remaining charts in Case II when the process is nondynamic and the serial correlation is stationary. (v) The charts AR, KNN and DSVM have a reasonable performance in Case I when the process observations are i.i.d. and normally distributed, but their performance in other cases are not satisfactory. Of course, the performance of the related charts may depend on the selection of their parameters (e.g., the constant ρ in the NDPM chart (5)–(6)). To avoid this impact on the method comparison, next, we adjust the procedure parameters for each method so that its ARL_1 value reaches the minimum for detecting a given shift while its ARL_0 value is kept at the nominal level of 200. Namely, the optimal performance of the related charts is considered here. As a result, the optimal ARL_1 values of the seven charts are shown in Figure 5. From the figure it can seen that similar conclusions to those from the example of Figure 4 can be made here regarding the optimal OC performance of the charts. These two examples show that the charts NDPM-D-S, NDPM-ND-NS, NDPM-ND-S, AR, KNN and DSVM have reasonable performance only when their model assumptions are valid and that the chart NDPM has a reasonable performance in most cases considered.

4. Application to air pollution surveillance in Beijing. With the rapid industrial development over the last several decades in China, the environmental pollution in that

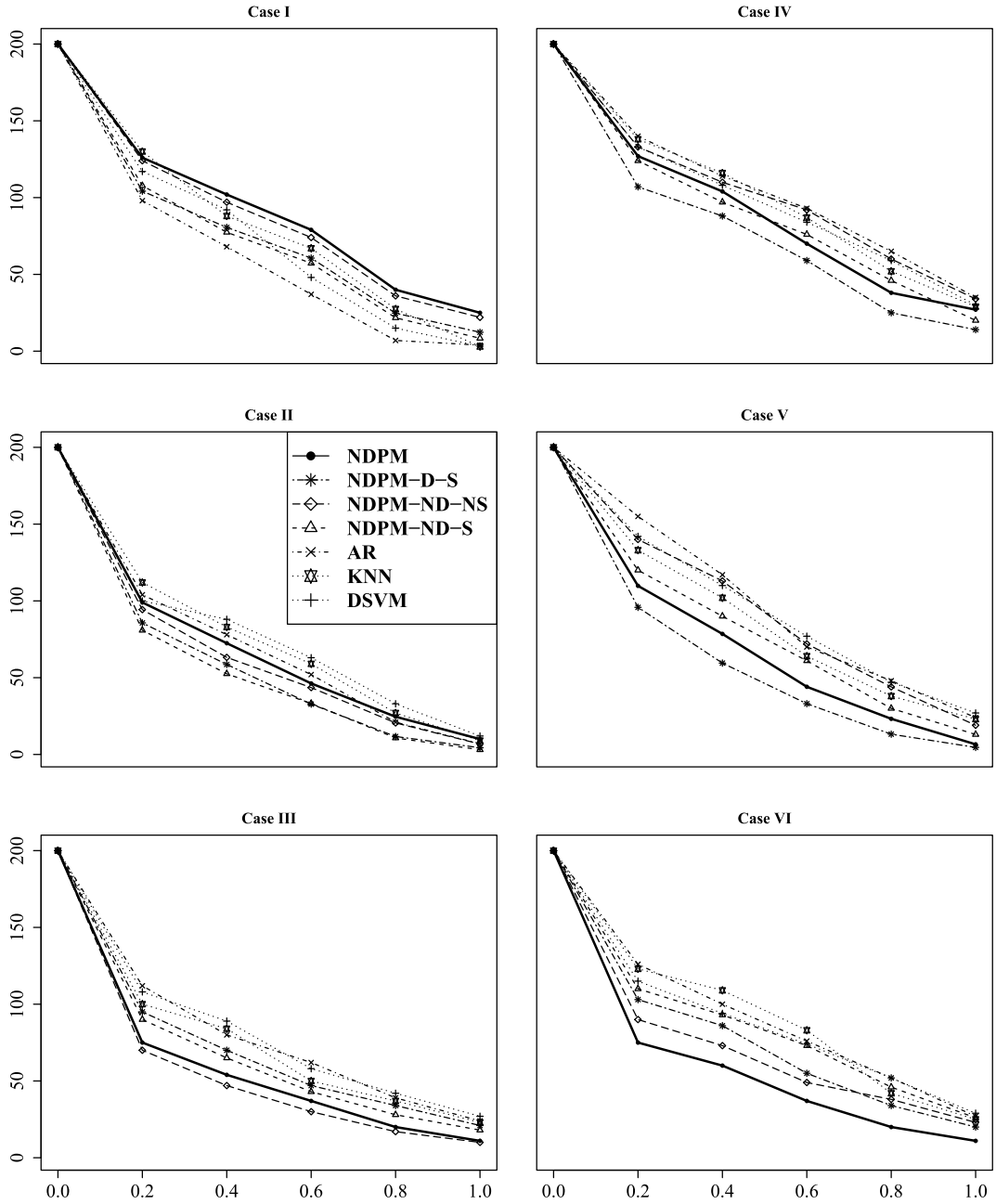

FIG. 4. *ARL_1 values of the seven charts when their nominal ARL_0 values are all fixed at* 200, $p = 3$, $m_0 = 500$, *their procedure parameters are chosen as in the example of Fig.* 2, *and the shift size parameter* δ *changes among* 0.2, 0.4, 0.6, 0.8 *and* 1.0.

country has become a problem that seriously damages public health. Beijing, the capital and one of the most populous cities of China, has suffered severe environmental pollution (cf. Figure 1 in Section 1). Possible causes of Beijing's air pollution include large-scale coal combustion, increasing number of motor vehicles, meteorological conditions and more (Liang et al. (2015)). This problem has got the government's attention in the past 10–15 years, and they are implementing several preventional measures to control the coal combustion and vehicle emissions. Beijing government also developed a pollution alert system in 2013 to collect air pollution data. Part of the data is saved in the UC

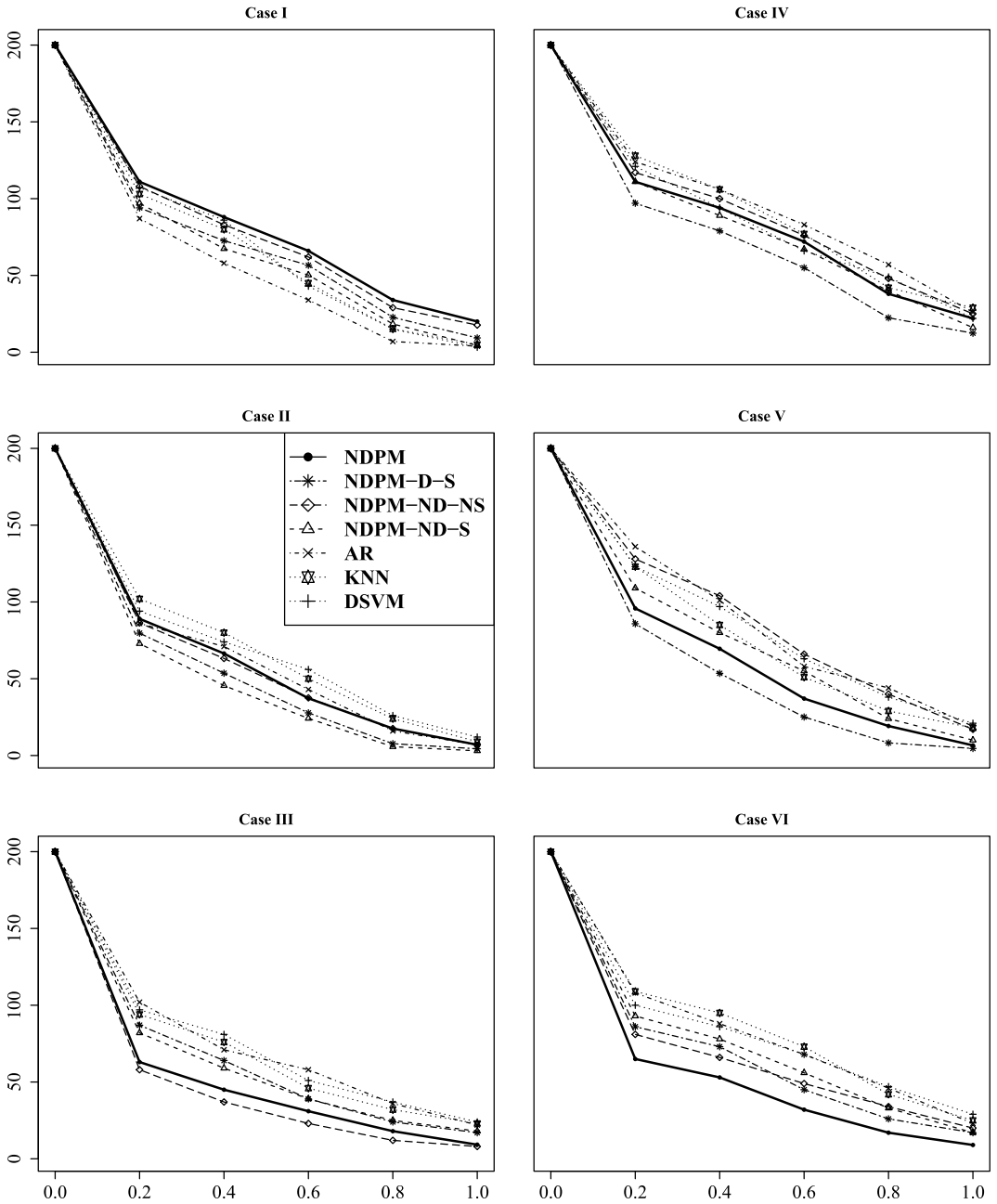

FIG. 5. *Optimal ARL_1 values of the seven charts when their nominal ARL_0 values are all fixed at* 200, $p = 3$, $m_0 = 500$, *and the shift size parameter* δ *changes among* 0.2, 0.4, 0.6, 0.8 *and* 1.0.

Irvine Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data). This dataset contains several air quality variables and meteorological variables. It has been used by environmental researchers for accessing the air quality in Beijing (e.g., Zhang et al. (2017)). In this section we apply the proposed dynamic process monitoring method NDPM to this dataset for air pollution surveillance. In this analysis the two most important air quality variables, that is, the density levels of $PM_{2.5}$ and CO, as well as the meteorological variable "dew point temperature (DEW)" are considered. While not a pollutant, per se, the meteorological variable DEW is considered here because it has been

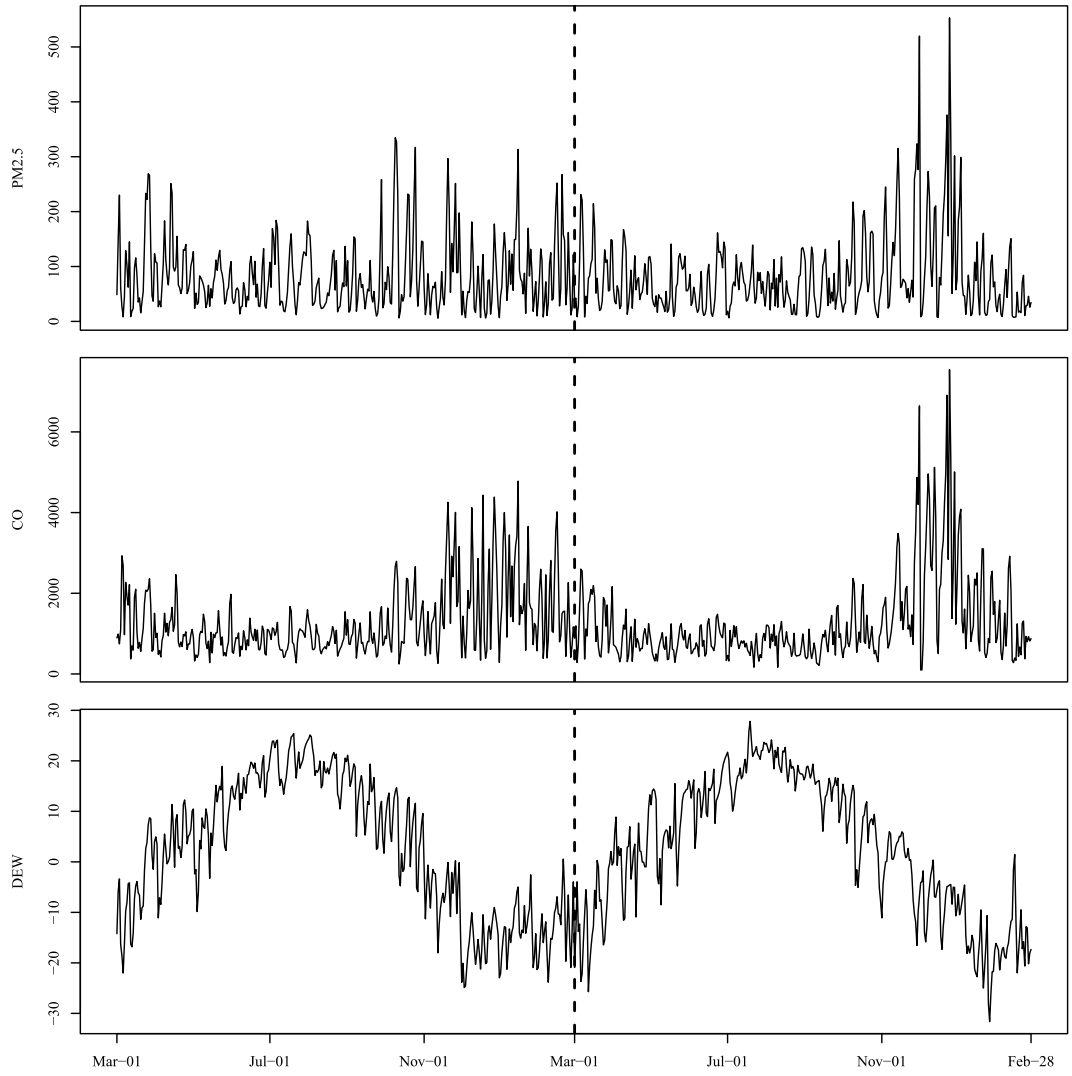

FIG. 6. *Original observations of $PM_{2.5}$, CO and DEW in Beijing during* 3/1/2014–2/28/2016. *The vertical dashed line in each plot separates the IC data from the data for online process monitoring.*

confirmed in the meteorological and environmental research that it is one of the most important meteorological variables that can substantially influence the air quality, especially the $PM_{2.5}$ level (Chaloulakou et al. (2003), Liu, Zhou and Lu (2020), Zhang et al. (2017)). That is because a higher DEW often implies a higher humidity and a high temperature, and the fog and/or haze formed in that meteorological condition are ideal for many air pollutants, including $PM_{2.5}$. The dataset used here contains observations of the three variables in two whole years from March 1, 2014 to February 28, 2016. The original data of the three variables are shown in Figure 6.

From Figure 6 there is a quite obvious seasonality in the observed data; the density levels of $PM_{2.5}$ and CO seem higher during winter times, and DEW seems higher in summer times. Also, the data in the first year, from March 1, 2014 to February 28, 2015, seem quite stable, although the CO levels during the winter times in that year are relatively high due mainly to a large amount of coal consumption in northern China during winter times (cf. Li et al. (2020)). Therefore, the data in the first year are used as the IC data for estimating the regular longitudinal pattern of the three variables, and the data in the second year are used for online

process monitoring. For the IC data we first compute the initial LLK estimate $\widehat{\boldsymbol{\mu}}^{(0)}(t)$, using (3), and then obtain the residuals $\mathbf{y}(t_j) - \widehat{\boldsymbol{\mu}}^{(0)}(t_j)$, for each j. To check the normality of the IC data, the Shapiro–Wilk test is then applied to the residuals, and the test gives a p-value of 2.2×10^{-26}, implying that the IC data are significantly nonnormal. To check the serial correlation, the Durbin–Watson test is applied to the residuals for each variable. The p-value of this test for each variable is $< 10^{-10}$, implying a significant autocorrelation in the IC data. To check the stationarity of autocorrelation, the augmented Dickey–Fuller (ADF) test is used for each variable. The p-value of this test for each variable is larger than 0.1. Thus, we fail to reject the null hypothesis that "autocorrelation is nonstationary" for each variable and conclude that the autocorrelation in the IC data is nonstationary. To check the correlation among the three variables in the IC data, the following sample correlation coefficient matrix is obtained:

$$\begin{bmatrix} 1.000 & 0.749 & 0.730 \\ 0.749 & 1.000 & 0.601 \\ 0.730 & 0.601 & 1.000 \end{bmatrix}.$$

Pearson's correlation test for checking whether the pairwise correlation is significant gives the p-values of 2.2×10^{-26} for all three pairs ($PM_{2.5}$, CO), ($PM_{2.5}$, DEW) and (CO, DEW). Therefore, there is a significant pairwise correlation among the three variables in the IC data.

Next, we apply the proposed dynamic process monitoring method NDPM to this dataset. As discussed in Section 2.1, we first estimate the IC mean functions and the IC covariance function from the IC data by (3) and (4). The IC data and the estimated IC mean functions are shown in the first row of Figure 7 from which it can be seen that the estimated IC mean functions describe the IC longitudinal pattern of the three variables well. Then, the proposed method NDPM is used for sequentially monitoring of the observed data in the second year starting from March 1, 2015, as discussed in Sections 2.2 and 2.3. The setups of the chart (5)–(6), for example, ARL_0 and computation of its control limit, are the same as those in the example of Table 1. The chart gives the first signal on November 28, 2015, which is two days earlier than the orange alert issued by the Beijing government on November, 30, 2015. To investigate whether this signal is real, observations of the three variables in the second year, during March 1, 2015 and February 28, 2016, are shown in the second row of Figure 7 along with the estimated IC mean functions shown by the solid curves in the plots. It can be seen that the longitudinal pattern of the observations is quite different from the estimated IC mean functions around the signal time shown by the vertical dashed lines in the plots. To further show the difference, the observed data of the three variables during October 1st and December 31st in years 2014 and 2015 are shown in the same plots of the third row of Figure 7. It can be seen that the two sets of data are similar at the beginning and then start to deviate around November 1. Such deviations are detected by NDPM on November 28, 2015.

As a comparison, the other six control charts, discussed in Section 3, are also applied to this dataset with the same setups as those in the example of Table 1. These control charts are shown in Figure 8 along with the chart NDPM. From the figure it can be seen that the chart AR gives signals almost everyday, and the charts NDPM-D-S, NDPM-ND-NS, NDPM-ND-S, KNN and DSVM give their first signals on December 7th, 22th, 8th, 8th and 15th of 2015, respectively. As mentioned earlier, the proposed chart NDPM gives its first signal on November 28, 2015. By the visualization of the observed data shown in Figure 7 and the hypothesis test results discussed earlier, which confirm that the observed IC data are nonnormal and have nonstationary autocorrelation, we can conclude that the frequent signals from AR may not be reliable in this example because its assumptions that the IC process observations are i.i.d. would be violated here and that the proposed chart NDPM should be more reliable and effective than its peers in this example for early detection of the air quality deterioration started around November 1, 2015.

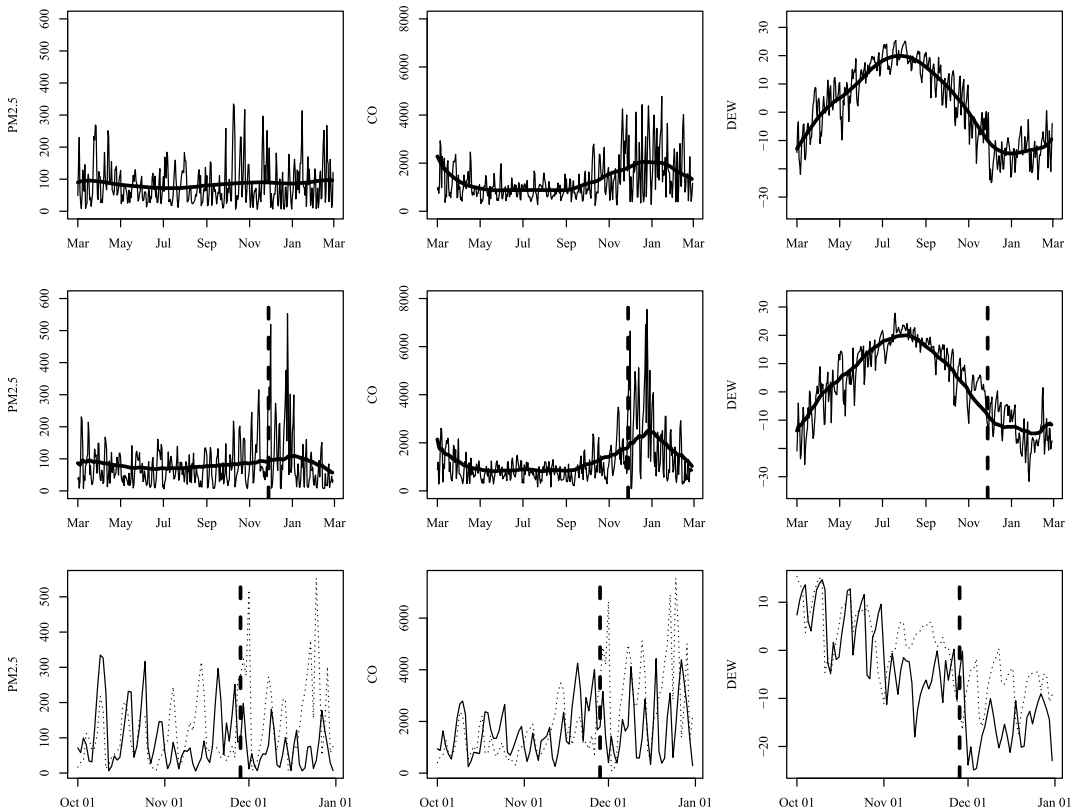

FIG. 7. *The first and second rows show observations of the three variables $PM_{2.5}$, CO and DEW in Beijing during* 3/1/2014–2/28/2015 *and* 3/1/2015–2/28/2016, *respectively. The solid curve in each plot denotes the estimated IC mean function of the related variable. The third row shows observations of the three variables during October* 1*st and December* 31*st in years* 2014 *and* 2015 *by the solid and dotted lines*, *respectively. The vertical dashed line in each plot denotes the signal time of the proposed method NDPM.*

5. Concluding remarks. Environmental pollution has become a major global problem, causing serious consequences on public health. Governments in the world are taking proper measures to reduce and control pollution emissions in order to improve the quality of our environment. In this effort, effective online monitoring of the air pollutant concentrations is especially important for governments to take proper interventions in a timely manner and protect public health. In this paper, we have developed a new method for air pollution surveillance. The new method can properly accommodate the dynamic longitudinal pattern of the process under monitoring and serial correlation in the observed data. It also is not limited to parametric distributional families. Both simulation studies and the application to monitor the air quality in Beijing show that it performs well in various cases. Although air pollution surveillance is focused in this paper, we would like to point out that our proposed method is actually quite general and can be applied to other dynamic process monitoring problems, including sequential monitoring of incidence rates of one or more infectious diseases (e.g., flu, COVID-19) in a region, online monitoring of sea-level pressures in oceanography and seismic monitoring in physical geography.

The proposed method still has much room for improvement. For instance, when monitoring the air quality in Beijing, besides the major variables $PM_{2.5}$, CO and DEW that need to be monitored online, there could be some covariates that provide useful information about the air quality, including air temperature, air pressure, wind speed and other weather conditions. Proper use of such covariates could potentially improve the performance of the proposed

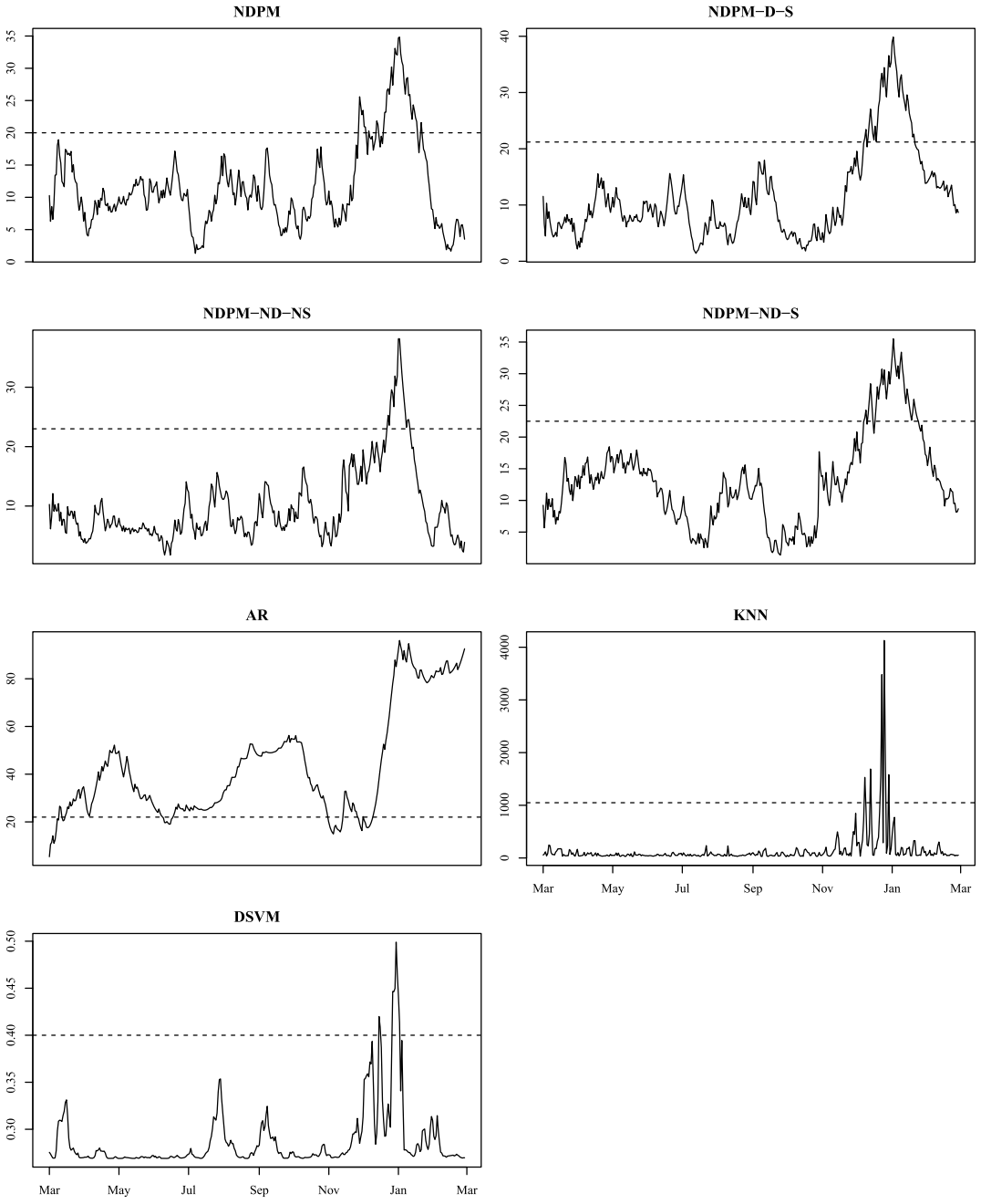

FIG. 8. *Seven control charts for online monitoring of the air pollution data in Beijing during March* 1, 2015 *and February* 28, 2016. *The horizontal dashed line in each plot denotes the control limit of the related control chart.*

method. Also, the current method is for monitoring the air quality at a single location. In practice, the air quality of multiple locations (e.g., Beijing and its surrounding cities) could be spatially correlated, and it could improve the effectiveness of the control chart to monitor the air quality at multiple locations simultaneously. This is related to the spatiotemporal process monitoring problem (e.g., Yang and Qiu (2020)). However, the current methods for spatiotemporal process monitoring consider a single quality variable only, while there could be multiple quality variables in practice. Also, the nonparametric spatiotemporal process monitoring methods require many spatial locations to be in the observed data since they are based

on nonparametric spatial smoothing that cannot work well in cases with only a few spatial locations. All these topics will be studied carefully in our future research.

Acknowledgments. The authors thank the Editor, the Associate Editor and two referees for many constructive comments and suggestions, which improved the quality of the paper greatly.

Funding. This research is supported in part by an NSF grant.

REFERENCES

AL-RASHED, A., AL-MUTAIRI, N. and ATTAR, M. A. (2019). Air pollution analysis in Kuwait using a statistical technique (CUSUM). *International Journal of Geosciences* **10** 254–294.

APLEY, D. W. and TSUNG, F. (2002). The autoregressive T^2 chart for monitoring univariate autocorrelated processes. *J. Qual. Technol.* **34** 80–96.

BARRATT, B., ATKINSON, R., ANDERSON, H. R., BEEVERS, S., KELLY, F., MUDWAY, I. and WILKINSON, P. (2007). Investigation into the use of the CUSUM technique in identifying changes in mean air pollution levels following introduction of a traffic management scheme. *Atmos. Environ.* **41** 1784–1791.

BOOGAARD, H., WALKER, K. and COHEN, A. J. (2019). Air pollution: The emergence of a major global health risk factor. *International Health* **11** 417–421.

CAPIZZI, G. and MASAROTTO, G. (2008). Practical design of generalized likelihood ratio control charts for autocorrelated data. *Technometrics* **50** 357–370. MR2528658 https://doi.org/10.1198/004017008000000280

CHAKRABORTI, S. and GRAHAM, M. A. (2019). Nonparametric (distribution-free) control charts: An updated overview and some results. *Quality Engineering* **31** 523–544.

CHALOULAKOU, A., KASSOMENOS, P., SPYRELLIS, N., DEMOKRITOU, P. and KOUTRAKIS, P. (2003). Measurements of PM10 and PM2.5 particle concentrations in Athens, Greece. *Atmos. Environ.* **37** 649–660.

CHATTERJEE, S. and QIU, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits. *Ann. Appl. Stat.* **3** 349–369. MR2668711 https://doi.org/10.1214/08-AOAS197

CHELANI, A. B. (2011). Change detection using CUSUM and modified CUSUM method in air pollutant concentrations at traffic site in Delhi. *Stoch. Environ. Res. Risk Assess.* **25** 827–834.

COHEN, A. J., BRAUER, M., BURNETT, R., ANDERSON, H. R., FROSTAD, J., ESTEP, K., BALAKRISHNAN, K., BRUNEKREEF, B., DANDONA, L. et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389** 1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

DE BRABANTER, K., DE BRABANTER, J., SUYKENS, J. A. K. and DE MOOR, B. (2011). Kernel regression in the presence of correlated errors. *J. Mach. Learn. Res.* **12** 1955–1976. MR2819023

EPSNEČNIKOV, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.* **14** 153–158.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

HE, S., JIANG, W. and DENG, H. (2018). A distance-based control chart for monitoring multivariate processes using support vector machines. *Ann. Oper. Res.* **263** 191–207. MR3775193 https://doi.org/10.1007/s10479-016-2186-4

HEALTH EFFECTS INSTITUTE (2019). *State of Global Air 2019*. Health Effects Institute, Boston, MA.

HIGHAM, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.* **103** 103–118. MR0943997 https://doi.org/10.1016/0024-3795(88)90223-6

JACOB, D. J. and WINNER, D. A. (2009). Effect of climate change on air quality, atmospheric environment. *Atmos. Environ.* **43** 51–63.

LEE, H. C. and APLEY, D. W. (2011). Improved design of robust exponentially weighted moving average control charts for autocorrelated processes. *Qual. Reliab. Eng. Int.* **27** 337–352.

LI, J. and QIU, P. (2017). Construction of an efficient multivariate dynamic screening system. *Qual. Reliab. Eng. Int.* **30** 1969–1981.

LI, W., SHAO, L., WANG, W., LI, H., WANG, X., LI, Y., LI, W., JONES, T. and ZHANG, D. (2020). Air quality improvement in response to intensified control strategies in Beijing during 2013–2019. *Sci. Total Environ.* **744** 140776.

LIANG, X., ZOU, T., GUO, B., LI, S., ZHANG, H., ZHANG, S., HUANG, H. and CHEN, S. X. (2015). Assessing Beijing's PM2.5 pollution: Severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society A* **471** 20150257.

LIANG, X., LI, S., ZHANG, S., HUANG, H. and CHEN, S. X. (2016). PM2.5 data reliability, consistency, and air quality assessment in five Chinese cities. *J. Geophys. Res., Atmos.* **121** 220–236.

LIU, Y., ZHOU, Y. and LU, J. (2020). Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Sci. Rep.* **10** 14518. https://doi.org/10.1038/s41598-020-71338-7

POPE, C. A., BURNETT, R. T., THURSTON, G. D., THUN, M. J., CALLE, E. E., KREWSKI, D. and GODLESKI, J. J. (2004). Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. *Circulation* **109** 71–77.

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

QIU, P. (2018). Some perspectives on nonparametric statistical process control. *J. Qual. Technol.* **50** 49–65.

QIU, P. and HAWKINS, D. (2003). A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions. *Statistician* **52** 151–164. MR1977257 https://doi.org/10.1111/1467-9884.00348

QIU, P. and XIANG, D. (2014). Univariate dynamic screening system: An approach for identifying individuals with irregular longitudinal behavior. *Technometrics* **56** 248–260. MR3207851 https://doi.org/10.1080/00401706.2013.822423

QIU, P. and XIE, X. (2021). Transparent sequential learning for statistical process control of serially correlated data. *Technometrics*. https://doi.org/10.1080/00401706.2021.1929493

SEAMAN, N. L. (2000). Meteorological modeling for air-quality assessments. *Atmos. Environ.* **34** 2231–2259.

SUKCHOTRAT, T., KIM, S. B. and TSUNG, F. (2010). One-class classification-based control charts for multivariate process monitoring. *IIE Trans.* **42** 107–120.

WU, W., JIN, Y. and CARLSTEN, C. (2018). Inflammatory health effects of indoor and outdoor particulate matter. *Journal of Allergy and Clinical Immunology* **141** 833–844.

XING, Y. F., XU, Y. H., SHI, M. H. and LIAN, Y. X. (2016). The impact of PM2.5 on the human respiratory system. *Journal of Thoracic Disease* **8** 69–74.

XUE, L. and QIU, P. (2021). A nonparametric CUSUM chart for monitoring multivariate serially correlated processes. *J. Qual. Technol.* **53** 396–409.

YANG, K. and QIU, P. (2020). Online sequential monitoring of spatio-temporal disease incidence rates. *IISE Trans.* **52** 1218–1233.

YOU, L. and QIU, P. (2019). Fast computing for dynamic screening systems when analyzing correlated data. *J. Stat. Comput. Simul.* **89** 379–394. MR3893031 https://doi.org/10.1080/00949655.2018.1552273

YOU, L. and QIU, P. (2020). An effective method for online disease risk monitoring. *Technometrics* **62** 249–264. MR4095749 https://doi.org/10.1080/00401706.2019.1625813

ZHANG, S., GUO, B., DONG, A., HE, J., XU, Z. and CHEN, S. X. (2017). Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A* **473** 20170457.

ZHAO, X., ZHANG, X., XU, X., XU, J., MENG, W. and PU, W. (2009). Seasonal and diurnal variations of ambient PM2.5 concentration in urban and rural environments in Beijing. *Atmos. Environ.* **43** 2893–2900.