A General Framework for Robust Monitoring of Multivariate Correlated Processes

Xiulin Xie and Peihua Qiu
Department of Biostatistics, University of Florida
2004 Mowry Road, Gainesville, FL 32610

Abstract

Statistical process control (SPC) charts provide an important analytic tool for online monitoring of sequential processes. Conventional SPC charts are designed for cases when in-control (IC) process observations are independent and identically distributed at different observation times and the IC process distribution belongs to a parametric (e.g., normal) family. In practice, however, these model assumptions are rarely valid. To address this issue, there have been some existing discussions in the SPC literature for handling cases when the IC process distribution cannot be described well by a parametric form, and some nonparametric SPC charts have been developed based on data ranking and/or data categorization. However, both data ranking and data categorization would lose information in the original process observations. Consequently, the effectiveness of the nonparametric SPC charts would be compromised. In this paper, we make another research effort to handle this problem by developing a general process monitoring framework that is robust to the IC process distribution and short-ranged serial correlation. The new method tries to preserve as much information in the original process observations as possible. Instead of using data ranking and/or data categorization, it is based on intensive data pre-processing, including data decorrelation, data transformation, and data integration. Because the distribution of the pre-processed data can be approximated well by a parametric distribution, the design and implementation of the new method is relatively simple. Numerical studies show that it is indeed robust to the IC process distribution and effective for online monitoring of multivariate processes with short-ranged serial correlation.

Key Words: Data decorrelation; Recursive computation; Robustness; Self-starting charts; Sequential learning; Transformation.

1 Introduction

Statistical process control (SPC) charts provide a powerful tool for online process monitoring in the manufacturing industry, disease surveillance, and many other applications (Hawkins and Olwell

1998; Montgomery 2012; Qiu 2014). Conventional SPC charts can be roughly classified into the following four categories: Shewhart charts (Shewhart 1931), cumulative sum (CUSUM) charts (Page 1954), exponentially weighted moving average (EWMA) charts (Robert 1959), and change-point detection (CPD) charts (Hawkins et al. 2003). These conventional charts are designed mainly for cases when in-control (IC) process observations at different time points are independent and identically distributed (i.i.d.) with a parametric (e.g., normal) distribution. In practice, however, these model assumptions are rarely valid. This paper focuses on multivariate process monitoring when the IC process observations are serially correlated and the IC process distribution cannot be described well by a parametric form.

In the SPC literature, it has been well demonstrated that traditional control charts are unreliable to use when their model assumptions are violated (Apley and Tsung 2002, Hackl and Ledolter 1991, Qiu and Xiang 2014). Recent SPC research has considered cases when one or more of these model assumptions are invalid. For instance, when the normality assumption is invalid, some nonparametric SPC charts have been developed (cf., Chakraborti and Graham 2019, Qiu 2018). One type of nonparametric control charts are constructed based on data ranking. See Capizzi and Masarotto (2013), Chakraborti et al. (2004), Krupskii et al. (2020), Qiu and Hawkins (2001), Zou and Tsung (2011), and more. For instance, Qiu and Hawkins (2001) suggested a nonparametric control chart using the ranking information across different components of a multivariate observation at a given time point. Zou and Tsung (2010) proposed a nonparametric EWMA chart based on a nonparametric goodness-of-fit test using the ranking information among process observations at different time points, and a multivariate extension of that method was discussed in Zhang et al. (2016). Some researchers also suggested constructing control charts based on spatial sign or spatial ranks (e.g., Li et al. 2017, Zou and Tsung 2011, Zou et al. 2012). Another type of nonparametric control charts takes an alternative strategy by first categorizing the original observations and then using methods of categorical data analysis for constructing control charts (Li 2021, Qiu 2008, Qiu and Xie 2022). However, information in the original process observations would be lost by using either data ranking or data categorization. Consequently, the effectiveness of these nonparametric charts would be compromised. As a side note, most nonparametric control charts mentioned above are designed for cases when process observations at different time points are assumed independent, while some others can handle cases when serial correlation is present in the observed data (e.g., Qiu and Xie 2022).

Because some information in the observed data would be lost if a nonparametric control chart is used as discussed above, some researchers argue that the conventional control charts should be used in practice even when the normality assumption is violated (cf., Borror et al. 1999, Stoumbos and Sullivan 2002, Testik et al. 2003). The main argument is that the conventional control charts are constructed based on the original process observations (thus, no information in the observed data is lost), and they would be robust to the normality assumption if their procedure parameters are chosen properly (cf., Borror et al. 1999, Stoumbos and Sullivan 2002, Testik et al. 2003). For instance, the conventional EWMA chart is a weighted average of process observations and the weights are controlled by its weighting parameter. If the weighting parameter is chosen small (e.g., 0.05), then the resulting EWMA charting statistic is a weighted average of a relatively large number of process observations. In such cases, its distribution would be approximately normal by the Central Limit Theorem (cf., Qiu 2014, Section 8.1). However, how small the weighting parameter should be chosen depends on the difference between the actual IC process distribution and a normal distribution, which is rarely known in practice. In addition, when the weighting parameter is chosen small, the resulting EWMA chart would be ineffective for detecting relatively large process distributional shifts. Therefore, much research is needed to develop control charts that are both reliable and effective in cases when the normality assumption is invalid.

In this paper, we make another research effort for online monitoring of multivariate processes with arbitrary IC distributions. A general process monitoring framework is developed, which is robust to the IC process distribution. The new method tries to preserve as much information in the original process observations as possible. Instead of using data ranking and/or data categorization, it is based on intensive data pre-processing, including data decorrelation, data transformation, and data integration. After the data pre-processing, the original multivariate process observations are transformed to univariate data whose IC distribution can be approximated well by a parametric distribution. Then, a univariate control chart can be applied to the transformed data for online process monitoring. Because the approximated IC distribution of the transformed data is parametric, design of the new process monitoring method is relatively simple, and the control limits of the related control charts can be determined in advance by Monte Carlo simulations. The proposed new process monitoring framework is self-starting in the sense that the IC data get expanded every time when the process under monitoring is claimed to be IC and the estimates of the related IC quantities get updated accordingly. Numerical studies show that it is indeed robust to the IC

process distribution and effective for online monitoring of multivariate processes with short-ranged serial correlation.

The remainder of the paper is organized as follows. In Section 2, our proposed new method is described in detail. Its numerical performance is evaluated in Section 3 by some simulation studies. Then, it is demonstrated using a real dataset about a semiconductor manufacturing process in Section 4. Several remarks conclude the paper in Section 5. The proof of a theoretical result is given in the supplementary material.

2 Proposed Method

Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ be p quality variables to monitor for a sequential process, $\mathcal{X}_{IC}^{(0)} = \{\mathbf{X}_{-m_0+1}, \mathbf{X}_{-m_0+2}, \ldots, \mathbf{X}_0\}$ be an initial IC dataset collected before online process monitoring, and $\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{np})'$ be the observation collected at the current time point n for online process monitoring. For the observed data $\{\mathbf{X}_n, n \geq 1\}$, it is assumed that their IC process distribution, including the IC serial correlation, does not change over time, and the IC serial correlation is short-ranged as well. Namely, it is assumed that $\boldsymbol{\gamma}(s) = \text{Cov}(\mathbf{X}_i, \mathbf{X}_{i+s})$ depends on s only when i changes, for any i and s, and $\boldsymbol{\gamma}(s) \approx 0$ when $s > b_{max}$, where b_{max} denotes the time range of serial data correlation. In practice, the serial correlation between $\mathbf{X}_i$ and $\mathbf{X}_{i+s}$ usually decays when s increases. Therefore, the above assumptions are routinely made in the SPC literature on monitoring serially correlated data (e.g., Apley and Tsung 2002, Capizzi and Masarotto 2008). To make a decision about the status of the process under monitoring, its observed data $\{\mathbf{X}_n, n \geq 1\}$ should be pre-processed so that the distribution of the pre-processed data can be approximated well by a parametric distribution. Then, a conventional control chart can be applied to the pre-processed data for process monitoring. If the process is claimed to be IC at the current time point n, then $\mathbf{X}_n$ is combined with the existing IC data for process monitoring at the next time point $n+1$. Otherwise, a signal is given. The entire proposed method is described below in several parts.

2.1 Initial estimation of the IC distribution

From the initial IC data $\mathcal{X}_{IC}^{(0)}$, we first calculate initial estimates of the IC mean $\boldsymbol{\mu}$ and the IC covariance matrices $\{\boldsymbol{\gamma}(s), 0 \leq s \leq b_{max}\}$. Because no parametric form is imposed on the IC

process distribution, the maximum likelihood estimation cannot be used. Instead, the following moment estimates are considered:

$$\begin{aligned} \widehat{\boldsymbol{\mu}}^{(0)} &= \frac{1}{m_0}\sum_{i=-m_0+1}^{0}\mathbf{X}_i, \\ \widehat{\boldsymbol{\gamma}}^{(0)}(s) &= \frac{1}{m_0-s}\sum_{i=-m_0+1}^{-s}\left(\mathbf{X}_{i+s}-\widehat{\boldsymbol{\mu}}^{(0)}\right)\left(\mathbf{X}_i-\widehat{\boldsymbol{\mu}}^{(0)}\right)', \quad \text{for } 0\le s\le b_{max}. \end{aligned} \tag{1}$$

In (1), it has been assumed that $m_0 > b_{max}$, which is a mild assumption since b_{max} is usually chosen in the range $[10, 20]$ (cf., Li and Qiu 2020). In cases when $p > m_0$ or $b_{max} > m_0$, the estimated covariance matrices defined in (1) could be singular. In such cases, alternative estimation approaches for analyzing high dimensional data should be considered (e.g., Bickel and Levina 2008, Pourahmadi 2013). In addition, there may be outliers in the initial IC data. When the number of outliers is small, their impact on the IC parameter estimates and the performance of the resulting control chart would be limited. However, when the number of outliers is relatively large, the outliers should be removed from the IC data in advance, or some outlier robust estimation methods should be considered here. In the statistical literature, there have been some discussions on parameter estimation from data with outliers. See, for instance, Peña and Prieto (2001) and Yu et al. (2012). Since high-dimensional covariance estimation and outlier robust estimation are not the focus of the current paper, they will not be discussed here in details.

After $\widehat{\boldsymbol{\mu}}^{(0)}$ and $\{\widehat{\boldsymbol{\gamma}}^{(0)}(s), 0\le s\le b_{max}\}$ are computed, the initial IC data $\mathcal{X}_{IC}^{(0)}$ can be standardized and decorrelated by a data decorrelation algorithm based on the Cholesky decomposition of sample covariance matrices. More specifically, let $\mathbf{W}_i = (\mathbf{X}_{i-b}', \mathbf{X}_{i-b+1}', \ldots, \mathbf{X}_i')'$ be a long vector consisting of the observation $\mathbf{X}_i$ and all its previous observations that need to be decorrelated with, where $b = \min(i+m_0-1, b_{max})$ and $-m_0+1 \le i \le 0$. Then, the estimate of $\mathrm{Cov}(\mathbf{W}_i, \mathbf{W}_i)$ can be defined to be

$$\widehat{\boldsymbol{\Sigma}}_{i,i} = \begin{pmatrix} \widehat{\boldsymbol{\gamma}}^{(0)}(0) & \cdots & \widehat{\boldsymbol{\gamma}}^{(0)}(b) \\ \vdots & \ddots & \vdots \\ [\widehat{\boldsymbol{\gamma}}^{(0)}(b)]' & \cdots & \widehat{\boldsymbol{\gamma}}^{(0)}(0) \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{i-1,i-1} & \widehat{\boldsymbol{\Sigma}}_{i-1,i} \\ \widehat{\boldsymbol{\Sigma}}_{i-1,i}' & \widehat{\boldsymbol{\gamma}}^{(0)}(0) \end{pmatrix}.$$

The decorrelated and standardized IC observation at time i is then defined to be

$$\mathbf{X}_i^* = \begin{cases} [\widehat{\boldsymbol{\gamma}}^{(0)}(0)]^{-1/2}\left(\mathbf{X}_i - \widehat{\boldsymbol{\mu}}^{(0)}\right), & \text{when } i = -m_0+1, \\ \widehat{\mathbf{D}}_i^{-1/2}\left[\mathbf{X}_i - \widehat{\boldsymbol{\mu}}^{(0)} - \widehat{\boldsymbol{\Sigma}}_{i-1,i}'\widehat{\boldsymbol{\Sigma}}_{i-1,i-1}^{-1}\widehat{\mathbf{e}}_{i-1}\right], & \text{when } i > -m_0+1, \end{cases}$$

where $\widehat{\mathbf{e}}_{i-1} = ((\mathbf{X}_{i-b}-\widehat{\boldsymbol{\mu}}^{(0)})', (\mathbf{X}_{i-b+1}-\widehat{\boldsymbol{\mu}}^{(0)})', \ldots, (\mathbf{X}_{i-1}-\widehat{\boldsymbol{\mu}}^{(0)})')'$, and $\widehat{\mathbf{D}}_i = \widehat{\boldsymbol{\gamma}}^{(0)}(0) - \widehat{\boldsymbol{\Sigma}}_{i-1,i}'\widehat{\boldsymbol{\Sigma}}_{i-1,i-1}^{-1}\widehat{\boldsymbol{\Sigma}}_{i-1,i}$.

It should be pointed out that the inverse matrices $\widehat{\mathbf{\Sigma}}_{i-1,i-1}^{-1}$, $[\widehat{\gamma}^{(0)}(0)]^{-1/2}$ and $\widehat{\mathbf{D}}_i^{-1/2}$ discussed above may not always exist in practice, especially in cases when the IC sample size m_0 is relatively small. In such cases, we suggest using the matrix modification method discussed in Higham (1988) to modify the related matrices to positive semidefinite matrices, which can be implemented using the function **nearPD()** in the R-package **Matrix**. Theoretically speaking, if $\widehat{\boldsymbol{\mu}}^{(0)}$ and $\{\widehat{\gamma}^{(0)}(s), 0 \le s \le b_{max}\}$ are the true IC process mean and the IC covariance matrices, then it can be checked that the decorrelated and standardized IC observations $\{\mathbf{X}_i^*, i = -m_0+1, -m_0+2, \ldots, 0\}$ would be uncorrelated with each other and each of them would have mean $\mathbf{0}$ and the identity covariance matrix. Let $F_j(x)$ be the IC cumulative distribution function (cdf) of the jth component of the decorrelated data, for $j = 1, 2, \ldots, p$. Then, $F_j(x)$ can be estimated by the following empirical cdf:

$$\widehat{F}_j^{(0)}(x) = \frac{1}{m_0} \sum_{i=-m_0+1}^{0} I\left(X_{ij}^* \le x\right), \tag{2}$$

where X_{ij}^* denotes the jth component of $\mathbf{X}_i^*$, and $I(u)$ is the indicator function that equals 1 when u is "true" and 0 otherwise.

2.2 Data transformation and online process monitoring

To monitor the multivariate process observations $\{\mathbf{X}_n, n \ge 1\}$, the observed data need to be pre-processed so that a conventional control chart is appropriate to use. To this end, at the current observation time n, the observation $\mathbf{X}_n$ should be decorrelated with all its previous observations using a procedure similar to the one used for decorrelating the initial IC data discussed in the previous subsection. The resulting decorrelated and standardized observation at time n is denoted as $\mathbf{X}_n^* = (X_{n1}^*, X_{n2}^*, \ldots, X_{np}^*)$. Next, we want to find a transformation so that the IC distribution of the transformed data $\mathbf{X}_n^*$ can be approximated well by a parametric distribution. To this end, let $\{\widehat{F}_j^{(n-1)}(x), j = 1, 2, \ldots, p\}$ be the estimates of the IC cdf's of the p components of $\mathbf{X}_n^*$ obtained at time $n-1$ by the recursive formulas given in Subsection 2.3 below. Then, $\widehat{F}_1^{(n-1)}(X_{n1}^*), \widehat{F}_2^{(n-1)}(X_{n2}^*), \ldots, \widehat{F}_p^{(n-1)}(X_{np}^*)$ would be asymptotically independent of each other, and each of them has the asymptotic distribution of $U[0,1]$. So, their joint distribution can be approximated well by the product of p $U[0,1]$ distributions. Let $G(\cdot)$ be the cdf of the product of p independent $U[0,1]$ random variables. Then, by the Glivenko-Cantelli theorem (cf., Tucker 1959), each empirical cdf $\widehat{F}_j^{(n-1)}(x)$ would converge uniformly to the true cdf $F_j(x)$, for $j = 1, 2, \ldots, p$,

and consequently

$$z_n = \Phi^{-1}\left[G\left(\Pi_{j=1}^{p}\widehat{F}_j^{(n-1)}(X_{nj}^*)\right)\right] \tag{3}$$

would have the asymptotic distribution of $N(0,1)$, where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal cdf. Regarding the cdf $G(\cdot)$, its analytic formula is given in Proposition 1 below with the derivation given in Appendix.

Proposition 1. *Let $U_1, U_2, \ldots, U_p$ be p independent uniform random variables on the interval $[0,1]$. Then, the cdf of $\Pi_{j=1}^{p}U_j$ is given by the following formula:*

$$G(z) = 1 - \eta\left(-ln(z); p, 1\right),$$

where $\eta(x;\alpha,\beta) = 1 - \sum_{l=0}^{\alpha-1}(\beta x)^l e^{-\beta x}/l!$ is the cdf of the Gamma distribution with parameters α and β.

The derivation of the result in Proposition 1 is given in Appendix.

After obtaining the transformed data $\{z_n, n \geq 1\}$ by (3), it is natural to consider the following univariate EWMA charting statistic (cf., Robert 1959):

$$E_{n,z} = \lambda_1 z_n + (1-\lambda_1)E_{n-1,z}, \tag{4}$$

where $E_{0,z} = 0$, and $\lambda_1 \in (0,1]$ is a weighting parameter. Then, the chart gives a signal of process mean shift at time n if

$$\sqrt{\frac{2-\lambda_1}{\lambda_1}}|E_{n,z}| > h_1, \tag{5}$$

where $h_1 > 0$ is a control limit chosen to achieve a given value of the IC average run length, denoted as ARL_0. The chart (4)-(5) is called EWMA-P chart hereafter, where the last letter "P" reflects the fact that the transformed observation z_n is based on the *product* of p transformed components $\{\widehat{F}_j^{(n-1)}(X_{nj}^*), j = 1, 2, \ldots, p\}$ of the original process observation $\mathbf{X}_n$.

To transform the original p-dimensional process observations $\{\mathbf{X}_n, n \geq 1\}$ into the univariate data $\{z_n, n \geq 1\}$ by (3), there are several benefits. First, the dimension of the process monitoring problem has been reduced from p to 1, which simplifies the process monitoring problem greatly. Second, although we do not impose any parametric form on the IC distribution of the original process observations, the transformed data would have the asymptotic $N(0,1)$ distribution. Consequently, proper design of the resulting process monitoring procedure (4)-(5) becomes convenient,

since the control limit h_1 can be determined by Monte Carlo simulations. See a more detailed discussion in Subsection 2.4. However, to reduce the original p-dimensional data to the transformed univariate data, some shifts in the original process observations $\{\mathbf{X}_n, n \geq 1\}$ may not be reflective in the transformed data $\{z_n, n \geq 1\}$, which is addressed in Proposition 2 below whose proof is straightforward and thus omitted.

Proposition 2. *Assume that the p-dimensional process observations $\{\mathbf{X}_n, n \geq 1\}$ have an IC distribution with mean $\boldsymbol{\mu}$, and are stationarily serially correlated with the invertible covariance matrices $\{\boldsymbol{\gamma}(s), 0 \leq s \leq b_{max}\}$. After the original process observations have a mean shift from $\boldsymbol{\mu}$ to $\boldsymbol{\mu}_1$ with the shift size $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu} \neq \mathbf{0}$, the distribution of the transformed data $\{z_n, n \geq 1\}$ defined in (3) would not change if and only if*

$$\Pi_{j=1}^{p} F_j(0) = \Pi_{j=1}^{p} F_j(\delta_j^*), \tag{6}$$

where $(\delta_1^*, \delta_2^*, \ldots, \delta_p^*)' = [\boldsymbol{\gamma}(0) - \boldsymbol{\Sigma}'_{n-1,n} \boldsymbol{\Sigma}^{-1}_{n-1,n-1} \boldsymbol{\Sigma}_{n-1,n}]^{-1/2} \boldsymbol{\delta}$, $\boldsymbol{\Sigma}_{n,n} = Cov(\mathbf{W}_n, \mathbf{W}_n)$, $\boldsymbol{\Sigma}_{n-1,n} = Cov((\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_{n-1})', \mathbf{X}_n)$, $\mathbf{W}_n = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_n)'$, and $F_j(x)$ is the IC cdf of the jth component of the decorrelated data, for $j = 1, 2, \ldots, p$, as discussed in Equation (2).

From Proposition 2, any shift in the original process observations with the shift size $\boldsymbol{\delta}$ satisfying Equation (6) cannot be reflected in the transformed data $\{z_n, n \geq 1\}$. Thus, such a shift cannot be detected by the EWMA-P chart (4)-(5). While the original shift size $\boldsymbol{\delta}$ is p-dimensional, the ones satisfying (6) are in one-dimensional space. Thus, the EWMA-P chart can actually detect most shifts in the original process observations when $p \geq 2$. To have an intuitive perspective on the shifts satisfying Equation (6), let us consider the following two examples in which p equals 2 and 3, respectively.

Example 1. *Assume that $\{\mathbf{X}_n, n \geq 1\}$ are i.i.d. two-dimensional random variables, and each component of $\mathbf{X}_n$ has the standardized χ^2_3 distribution. Then, a mean shift of size $\boldsymbol{\delta} = (\delta_1, \delta_2)$ satisfies Equation (6) if and only if*

$$\delta_1 = \sqrt{6} F_1^{-1} \left[\frac{F_1(0)^2}{F_1(\delta_2/\sqrt{6})} \right],$$

where $F_1(\cdot)$ is the cdf of the standardized χ^2_3 distribution. The above equation defines a curve in the two-dimensional space of (δ_1, δ_2) shown in the left panel of Figure 1.

Example 2. *Assume that $\{\mathbf{X}_n, n \geq 1\}$ are i.i.d. three-dimensional random variables, and the three components of $\mathbf{X}_n$ have the $N(0,1)$, the standardized t_3 and the standardized χ^2_3 distributions,*

respectively. Then, a mean shift of size $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3)$ *satisfies Equation (6) if and only if*

$$\delta_1 = F_1^{-1}\left[\frac{F_1(0)F_2(0)F_3(0)}{F_2(\delta_2/\sqrt{3})/F_3(\delta_3/\sqrt{6})}\right],$$

where $F_1(\cdot), F_2(\cdot)$ *and* $F_3(\cdot)$ *are the cdf's of the N(0,1), the standardized* χ^2_3*, and the standardized* t_3 *distributions, respectively. The above equation defines a surface in the three-dimensional space of* $(\delta_1, \delta_2, \delta_3)$ *that is shown in the right panel of Figure 1.*

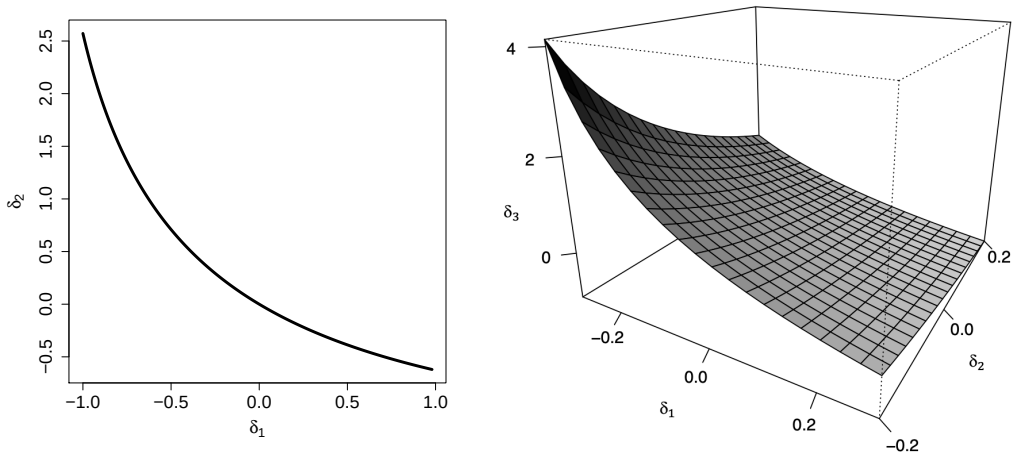

Figure 1: The line and surface in the left and right panels denote the shift sizes $\boldsymbol{\delta}$ that satisfy Equation (6) in Examples 1 and 2, respectively.

To overcome the limitation of the EWMA-P chart (4)-(5) that it cannot detect shifts satisfying Equation (6) as discussed above, one natural idea is to first transform each component of the decorrelated data $\mathbf{X}_n^*$ by

$$Z_{nj} = \Phi^{-1}\left[\widehat{F}_j^{(n-1)}(X_{nj}^*)\right], \text{ for } j = 1, 2, \ldots, p,$$

and then consider using the squared sum $T_n^2 = \sum_{j=1}^p Z_{nj}^2$. It is obvious that the IC distribution of Z_{nj} is asymptotically $N(0,1)$, for each j, and thus the IC distribution of T_n^2 would be close to χ_p^2. However, it would not be a good idea to use T_n^2 directly for process monitoring because its variability could be large, especially when the initial IC sample size m_0 is small. This is because the empirical cdf's $\{\widehat{F}_j^{(n-1)}(x), j = 1, 2, \ldots, p\}$ used in the above data transformation would have quite large variability, which has been confirmed by our numerical studies. Consequently, online process monitoring using T_n^2 as the charting statistic would not be effective. To address this issue, we propose an alternative charting statistic based on the EWMA idea. First, for each component

of the transformed data, let us consider the following EWMA statistic:

$$E_{nj} = \lambda_2 Z_{nj} + (1-\lambda_2)E_{n-1,j}, \quad \text{for } j = 1, 2, \ldots, p, \ n \geq 1, \tag{7}$$

where $E_{0j} = 0$, for each j, and $\lambda_2 \in (0, 1]$ is a weighting parameter. Because the sequence $\{Z_{nj}, n \geq 1\}$, for each j, is asymptotically uncorrelated and each variable has the asymptotic $N(0,1)$ distribution, E_{nj} would have the asymptotic distribution of $N(0, \lambda_2/(2-\lambda_2))$. Since the above EWMA operation is a weighted averaging procedure, variability in the transformed data $\{Z_{nj}, j = 1, 2, \ldots, p, n \geq 1\}$ would be reduced by this operation. As a matter of fact, because the weighting parameter λ_2 is often chosen to be a small number (e.g., 0.05 and 0.1), the asymptotic variance $\lambda_2/(2-\lambda_2)$ of E_{nj} would be much smaller than the asymptotic variance 1 of Z_{nj}, for each j. Then, $\frac{2-\lambda_2}{\lambda_2}\sum_{j=1}^{p} E_{nj}^2$ would be asymptotically χ_p^2 distributed. Define

$$B_n = \Phi^{-1}\left\{ Q_p\left[\frac{2-\lambda_2}{\lambda_2} \sum_{j=1}^{p} E_{nj}^2 \right] \right\}, \tag{8}$$

where $Q_p(\cdot)$ is the cdf of the χ_p^2 distribution. Then, B_n is our proposed charting statistic and has the asymptotic distribution of $N(0,1)$. The chart gives a signal of process mean shift at time n if

$$B_n > h_2, \tag{9}$$

where $h_2 > 0$ is a control limit. The chart (8)-(9) is called EWMA-Q chart hereafter, to reflect the fact that it is based on the *quadratic* form of $\{E_{nj}, j = 1, 2, \ldots, p\}$.

2.3 Recursive update of the IC distribution estimates

The initial estimates of certain IC quantities discussed in Subsection 2.1 may not be able to approximate their true values well, since it is usually hard to collect a large initial IC dataset in many SPC applications. To overcome that difficulty, Hawkins (1987) suggested the novel self-starting idea for online monitoring of a sequential process with only a handful of IC process observations collected in advance. Its basic idea is that if the process is declared to be IC at the current observation time n, then the observation $\mathbf{X}_n$ can be combined with the IC data $\mathcal{X}_{IC}^{(n-1)} = \{\mathbf{X}_j, -m_0+1 \leq j \leq n-1\}$ at the previous time $n-1$. The estimates of the IC quantities can then be updated using the combined IC dataset $\mathcal{X}_{IC}^{(n)}$ for process monitoring at the next time point $n+1$. In the proposed method, the following recursive formulas can be used for updating the estimates of the related IC

quantities: for $1 \le j \le p$ and $0 \le s \le b_{max}$,

$$\begin{aligned}
\widehat{F}_j^{(n)}(x) &= \frac{m_0+n-1}{m_0+n}\widehat{F}_j^{(n-1)}(x) + \frac{1}{m_0+n} I\left(X_{nj}^* \le x\right), \qquad (10)\\
\widehat{\boldsymbol{\mu}}^{(n)} &= \frac{1}{m_0+n}\mathbf{X}_n + \frac{m_0+n-1}{m_0+n}\widehat{\boldsymbol{\mu}}^{(n-1)},\\
\widehat{\boldsymbol{\gamma}}^{(n)}(s) &= \frac{1}{m_0+n-s}\left(\mathbf{X}_n - \widehat{\boldsymbol{\mu}}^{(n)}\right)\left(\mathbf{X}_{n-s} - \widehat{\boldsymbol{\mu}}^{(n)}\right)' + \frac{m_0+n-s-1}{m_0+n-s}\widehat{\boldsymbol{\gamma}}^{(n-1)}(s).
\end{aligned}$$

2.4 Design of the proposed control charts

From the description in the previous subsections, it can be seen that both the EWMA-P chart (4)-(5) and the EWMA-Q chart (8)-(9) are based on robust process monitoring by recursive data transformation and sequential data decorrelation. Because the data structure of the original process observations has been simplified by the data-preprocessing procedures mentioned above, design of both charts becomes relatively simple since their control limits can be determined by Monte Carlo simulations. From Expressions (3)-(5), it can be seen that design of the EWMA-P chart (4)-(5) can proceed in the same way as that for the conventional EWMA chart (cf., Qiu 2014, Chapter 5). For the EWMA-Q chart (8)-(9), its control limit h_2 can be determined by a Monte Carlo simulation with the following several steps.

- First, a sequence of p-dimensional random vectors can be generated from the $N_p(\mathbf{0}, I_{p\times p})$ distribution as the transformed process observations $\{\mathbf{Z}_n = (Z_{n1}, Z_{n2}, \ldots, Z_{np})', n \ge 1\}$.
- Then, the charting statistic values $\{B_n, n \ge 1\}$ are computed by (7) and (8).
- For a given value of h_2, 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.
- The previous three steps are then repeated for M times, and the average of the M RL values provides an estimate of the actual ARL_0 value of the chart for the given value of h_2.
- Finally, the control limit h_2 can be determined by a numerical search algorithm (e.g., the bisection search) from the above simulation so that the pre-specified ARL_0 value is reached.

Between the two charts, it has been pointed out in Subsection 2.2 that the EWMA-P chart is effective for detecting most shifts except those satisfying Expression (6). As a comparison, it is obvious that the EWMA-Q chart is effective for detecting any shifts in the original process

observations since any such shifts will be reflected in the distribution of its charting statistic B_n. See the related numerical results in Section 3. It should be pointed out that although the two proposed control charts are described for detecting process mean shifts in cases when the serial correlation in process observations is stationary, their basic idea is actually quite general and can be used for solving other SPC problems. For instance, in cases when the serial correlation is non-stationary, the covariance function of the sequential process under monitoring can be estimated using a kernel smoothing approach (cf., Qiu and Xie 2022). Then, the observed data can be pre-processed using the estimated time-varying covariance matrices, and the EWMA-P chart (4)-(5) and the EWMA-Q chart (7)-(9) can be applied to the pre-processed data for process monitoring.

For readers' convenience, the EWMA-P chart (4)-(5) and the EWMA-Q chart (7)-(9) are summarized in the pseudo code below.

3 Simulation

In this section, we present some simulation results about the numerical performance of the proposed EWMA-P chart (4)-(5) and EWMA-Q chart (8)-(9). For comparison purposes, the following six alternative charts are also considered in the simulation studies.

- The self-starting multivariate EWMA chart suggested by Sullivan and Jones (2002), denoted as SS-MEWMA. This chart first defines a multivariate EWMA statistic as follows:

$$\mathbf{E}_{n,ss} = \lambda_{ss}(\mathbf{X}_n - \widehat{\mu}^{(n-1)}) + (1-\lambda_{ss})\mathbf{E}_{n-1,ss}, \qquad \text{for } n \geq 1,$$

  where $\mathbf{E}_{0,ss} = \mathbf{0}$, and $\lambda_{ss} \in (0,1]$ is a weighting parameter. Then, its charting statistic is defined to be

$$\sqrt{Q_1^{-1}\left[F_{p,m_0+n-p-1}\left(\frac{m_0+n-1}{p(m_0+n-2)}\mathbf{E}'_{n,ss}\widehat{\Sigma}^{-1}_{\mathbf{E}_{n,ss}}\mathbf{E}_{n,ss}\right)\right]},$$

  where $Q_1(\cdot)$ is the cdf of the χ^2_1 distribution, $F_{p,m_0+n-p-1}$ is the cdf of the $F(p, m_0+n-p-1)$ distribution, and $\widehat{\Sigma}_{\mathbf{E}_{n,ss}} = [\lambda_{ss}/(2-\lambda_{ss})]\widehat{\gamma}^{(n-1)}(0)$.

- The self-starting version of the cumulative sum of T chart suggested by Crosier (1988), denoted as SS-COT. This method applies the conventional upward CUSUM chart (cf., Qiu 2014, Chapter 4) to the sequence $\{T_n, n \geq 1\}$, where $T_n^2 = (\mathbf{X}_n - \widehat{\boldsymbol{\mu}}^{(n-1)})' \left[\widehat{\boldsymbol{\gamma}}^{(n-1)}(0)\right]^{-1} (\boldsymbol{X}_n - \widehat{\boldsymbol{\mu}}^{(n-1)})$ is the Hotelling's T-square statistic.

Pseudo Code : Proposed Robust Process Monitoring Scheme

Initial Estimation:: Obtain the initial estimates $\widehat{\boldsymbol{\mu}}^{(0)}$, $\{\widehat{\gamma}^{(0)}(s), 0 \le s \le b_{max}\}$ and $\{\widehat{F}_j^{(0)}(x), 1 \le j \le p\}$ from the initial IC data $\mathcal{X}_{IC}^{(0)}$, as discussed in subsection 2.1.

for $n = 1, 2, \ldots,$ **do**

Collect observation at the current time n.

if $n = 1$ **then**

Define the standardized observation to be $\mathbf{X}_1^* = [\widehat{\gamma}^{(0)}(0)]^{-1/2} \left(\mathbf{X}_1 - \widehat{\boldsymbol{\mu}}^{(0)}\right)$.

else

Define $\mathbf{W}_n = (\mathbf{X}'_{n-b}, \mathbf{X}'_{n-b+1}, \ldots, \mathbf{X}'_n)'$, where $b = \min(n-1, b_{max})$.

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

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

where $\widehat{\mathbf{e}}_{n-1} = [(\mathbf{X}_{n-b} - \widehat{\boldsymbol{\mu}}^{(n-1)})', (\mathbf{X}_{n-b+1} - \widehat{\boldsymbol{\mu}}^{(n-1)})', \ldots, (\mathbf{X}_{n-1} - \widehat{\boldsymbol{\mu}}^{(n-1)})']'$,
$\widehat{\mathbf{D}}_n = \widehat{\gamma}^{(n-1)}(0) - \widehat{\boldsymbol{\Sigma}}'_{n-1,n} \widehat{\boldsymbol{\Sigma}}^{-1}_{n-1,n-1} \widehat{\boldsymbol{\Sigma}}_{n-1,n}$, and $\widehat{\boldsymbol{\Sigma}}_{n-1,n-1}$ and $\widehat{\boldsymbol{\Sigma}}_{n-1,n}$ are estimates of $\mathrm{Cov}(\widetilde{\mathbf{W}}_{n-1}, \widetilde{\mathbf{W}}_{n-1})$ and $\mathrm{Cov}(\widetilde{\mathbf{W}}_{n-1}, \mathbf{X}_n)$, respectively, in which
$\widetilde{\mathbf{W}}_{n-1} = (\mathbf{X}'_{n-b}, \mathbf{X}'_{n-b+1}, \ldots, \mathbf{X}'_{n-1})'$ and both estimates can be obtained from $\{\widehat{\gamma}^{(n-1)}(s), 0 \le s \le b_{max}\}$ defined in (10).

end if

Transform the decorrelated data by either (3) or (7)-(8), and then apply the corresponding control chart (4)-(5) or (8)-(9).

if A shift is detected **then**

Give a signal.

else

Combine $\mathbf{X}_n$ with the IC data $\mathcal{X}_{IC}^{(n-1)}$ at the previous time point and
update the estimates of the IC quantities by (10).

Let n = n + 1.

end if

end for

- The self-starting version of the multivariate control chart discussed in Mei (2010), denoted as SS-SCUSUM. Its charting statistic is the sum of p CUSUM statistics for monitoring the p individual quality variables. More specifically, the charting statistic of SS-SCUSUM is defined

to be

$$M_n = \sum_{j=1}^{p} \max\{C_{nj}^+, -C_{nj}^-\},$$

where

$$\begin{aligned} C_{nj}^+ &= \max\left[0, C_{n-1,j}^+ + (X_n - \widehat{\mu}_j^{(n-1)})/\widehat{\gamma}_{jj}^{(n-1)}(0) - k\right], \\ C_{nj}^- &= \min\left[0, C_{n-1,j}^- + (X_n - \widehat{\mu}_j^{(n-1)})/\widehat{\gamma}_{jj}^{(n-1)}(0) + k\right], \end{aligned}$$

$C_{0j}^+ = C_{0j}^- = 0$, for $j = 1, 2, \ldots, p$, $k > 0$ is the allowance constant, $\widehat{\mu}_j^{(n-1)}$ is the jth element of $\widehat{\boldsymbol{\mu}}^{(n-1)}$, and $\widehat{\gamma}_{jj}^{(n-1)}(0)$ is the (j,j)th element of $\widehat{\boldsymbol{\gamma}}^{(n-1)}(0)$.

- The self-starting version of the nonparametric multivariate CUSUM chart suggested in Qiu (2008), denoted as NP-CUSUM. To use this method, the original process observation $\mathbf{X}_n$ should first be standardized to $\mathbf{Y}_n = \left[\widehat{\boldsymbol{\gamma}}^{(n-1)}(0)\right]^{-1/2}(\boldsymbol{X}_n - \widehat{\boldsymbol{\mu}}^{(n-1)})$. Then, the multivariate nonparametric chart based on data categorization suggested in Qiu (2008) is applied to the standardized data for online process monitoring.

- The nonparametric multivariate EWMA chart suggested by Zou et al. (2012), denoted as SR-EWMA. This chart is based on spatial ranks of the process observations. This method applies the conventional multivaraite EWMA chart to the sequence $\{\mathbf{R}_n(\mathbf{X}_n), n \geq 1\}$, where $\mathbf{R}_n(\mathbf{x})$ is the empirical spatial rank function that can first be estimated from the initial IC data and then get updated over time.

- The multivariate nonparametric self-starting CUSUM chart suggested by Qiu and Xie (2022), denoted as QX-CUSUM. This chart first decorrelates the observed data as described in Section 2.1, and then applies the multivariate nonparametric chart based on data categorization (cf., Qiu 2008) to the decorrelated data for online process monitoring.

Regarding the IC process distribution and serial data correlation, the four different scenarios listed in Table 1 when $p = 3$ are considered. Scenario I is the conventional case considered in the SPC literature with the i.i.d. process observations and the standard normal IC process distribution. Scenario II considers a case when the IC distributions of some quality variables are non-normal, but the process observations are still i.i.d. Scenario III is constructed based on Scenario II, except that the three quality variables have serial correlation and follow AR(1), MA(2) and ARMA(2,1) time series models, respectively. Scenario IV is constructed based on Scenario III, except that the three quality variables are mutually correlated besides their own serial correlation.

Table 1: Four Scenarios considered in the simulation study.

Scenarios	Process Observations: $\mathbf{X}_n = (X_{n1}, X_{n2}, X_{n3})'$ where $n \geq 1$
I	$\mathbf{X}_n \overset{i.i.d.}{\sim} N_3(\mathbf{0}, I_{3\times 3})$
II	$X_{n1} = \epsilon_{n1}$, $X_{n2} = \epsilon_{n2}$, $X_{n3} = \epsilon_{n3}$, where $\{\epsilon_{n1}\} \overset{i.i.d.}{\sim} N(0,1)$ $\{\epsilon_{n2}\} \overset{i.i.d.}{\sim}$ standardized version of the t_3 distribution $\{\epsilon_{n3}\} \overset{i.i.d.}{\sim}$ standardized version of the χ^2_3 distribution
III	$X_{n1} = 0.2X_{n-1,1} + \epsilon_{n1}$, for $n \geq 1$, where $X_{01} = 0$, $X_{n2} = \epsilon_{n2} + 0.8\epsilon_{n-1,2} + 0.6\epsilon_{n-2,2}$ for $n \geq 3$, where $X_{12} = X_{22} = 0$, $X_{n3} = 0.3X_{n-1,3} + 0.1X_{n-2,3} + \epsilon_{n3} - 0.5\epsilon_{n-1,3}$, for $n \geq 3$ where $X_{13} = X_{23} = 0$, $\{\epsilon_{n1}\}$, $\{\epsilon_{n2}\}$, $\{\epsilon_{n3}\}$ are generated in the same way as that in Scenario II.
IV	$X_{n1} = 0.2X_{n-1,1} + \epsilon_{n1}$, for $n \geq 1$, where $X_{01} = 0$, $X_{n2} = 0.1X_{n1} + \epsilon_{n2} + 0.8\epsilon_{n-1,2} + 0.6\epsilon_{n-2,2}$ for $n \geq 3$, where $X_{12} = X_{22} = 0$, $X_{n3} = 0.1X_{n1} + 0.2X_{n2} + \epsilon_{n3}$, $\{\epsilon_{n1}\}$, $\{\epsilon_{n2}\}$, $\{\epsilon_{n3}\}$ are generated in the same way as that in Scenario II.

3.1 IC performance

We first evaluate the IC performance of the related control charts. In the simulation study, the nominal ARL_0 value is fixed at 200 for all charts. The allowance constants of SS-COT, SS-SCUSUM, NP-CUSUM and QX-CUSUM are chosen to be $\sqrt{p}$, 0.1, 0.1 and 0.1, respectively, and the weighting parameters of the charts SS-MEWMA, SR-EWMA, EWMA-P and EWMA-Q are all chosen to be 0.05. The b_{max} of the charts EWMA-P and EWMA-Q is chosen to be 10. The control limits of the charts SS-MEWMA, SS-COT, SS-SCUSUM are determined by simulation based on the normality assumption, and the control limits of NP-CUSUM, SR-EWMA, and QX-CUSUM are computed as discussed in Qiu (2008), Zou et al. (2012), and Qiu and Xie (2022), respectively.

To compute the actual ARL_0 value of a given chart, an IC dataset of size m_0 is first generated. Then, a control chart is applied to a sequence of 2,000 IC process observations for online process monitoring, and its RL value is recorded. The online process monitoring is then repeated for 1,000 times, and the average of the 1,000 RL values is used as an estimate of the actual conditional ARL_0 value of the chart, 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 the computation 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 sample size m_0 is first fixed at 500. The results of the estimated actual ARL_0 values of different charts are shown in Table 2. From the table, it can be seen that i) the three competing charts SS-MEWMA, SS-COT and SS-SCUSUM all perform well in Scenario I when process observations are i.i.d. and normally distributed, but their performance is quite poor in other scenarios when their assumptions are violated, ii) the nonparametric charts NP-CUSUM and SR-EWMA have a reasonably good performance in Scenarios I and II when the process observations are independent at different observation times, but are unreliable in Scenarios III and IV when the process observations are serially correlated, and iii) the chart QX-CUSUM and the two proposed charts EWMA-P and EWMA-Q have a reasonably good performance in all scenarios considered.

Table 2: Estimated actual ARL_0 values and their standard errors (in parentheses) of the eight control charts when $p = 3$, $m_0 = 500$, and the nominal ARL_0 values of all charts are fixed at 200.

Scenarios	SS-MEWMA	SS-COT	SS-CUSUM	NP-CUSUM	SR-EWMA	QX-CUSUM	EWMA-P	EWMA-Q
I	200(2.73)	188(5.60)	200(3.66)	208(1.75)	194(2.12)	218(2.70)	197(5.76)	190(5.39)
II	184(4.46)	122(5.24)	206(8.79)	199(1.84)	191(1.83)	206(3.21)	191(5.96)	193(5.73)
III	60.1(0.79)	88.6(2.97)	152(4.54)	71.7(0.77)	55.3(0.48)	203(2.84)	197(6.31)	191(6.03)
IV	55.3(0.74)	90.5(3.04)	120(3.51)	54.3(0.65)	50.4(0.38)	196(2.76)	184(5.62)	188(5.89)

The IC performance of the eight charts discussed above could be affected by the initial 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, and 1,000, and all other setups remain the same as those in the example of Table 2. The estimated actual ARL_0 values of the eight charts in such cases are presented in Figure 2. From the figure, we can have the following conclusions. First, the IC performance of all charts improves in Scenario I when m_0 increases. Second, the IC performance of SS-COT is unsatisfactory in Scenario II even when m_0 is large, while the IC performance of all other charts is reasonably good in that scenario. Third, the IC performance of the charts SS-MEWMA, NP-CUSUM, SS-CUSUM, SS-DOT and SR-EWMA is unsatisfactory in Scenarios III and IV even when m_0 is large, while the IC performance of QX-CUSUM, EWMA-Q and EWMA-P improves in these two scenarios when m_0 increases. From this example, it can be seen that the IC performance of the proposed charts EWMA-Q and EWMA-P is quite reliable when $m_0 \geq 400$ in all cases considered since their actual ARL_0 values are within 10% of the nominal

ARL_0 value of 200 in such cases.

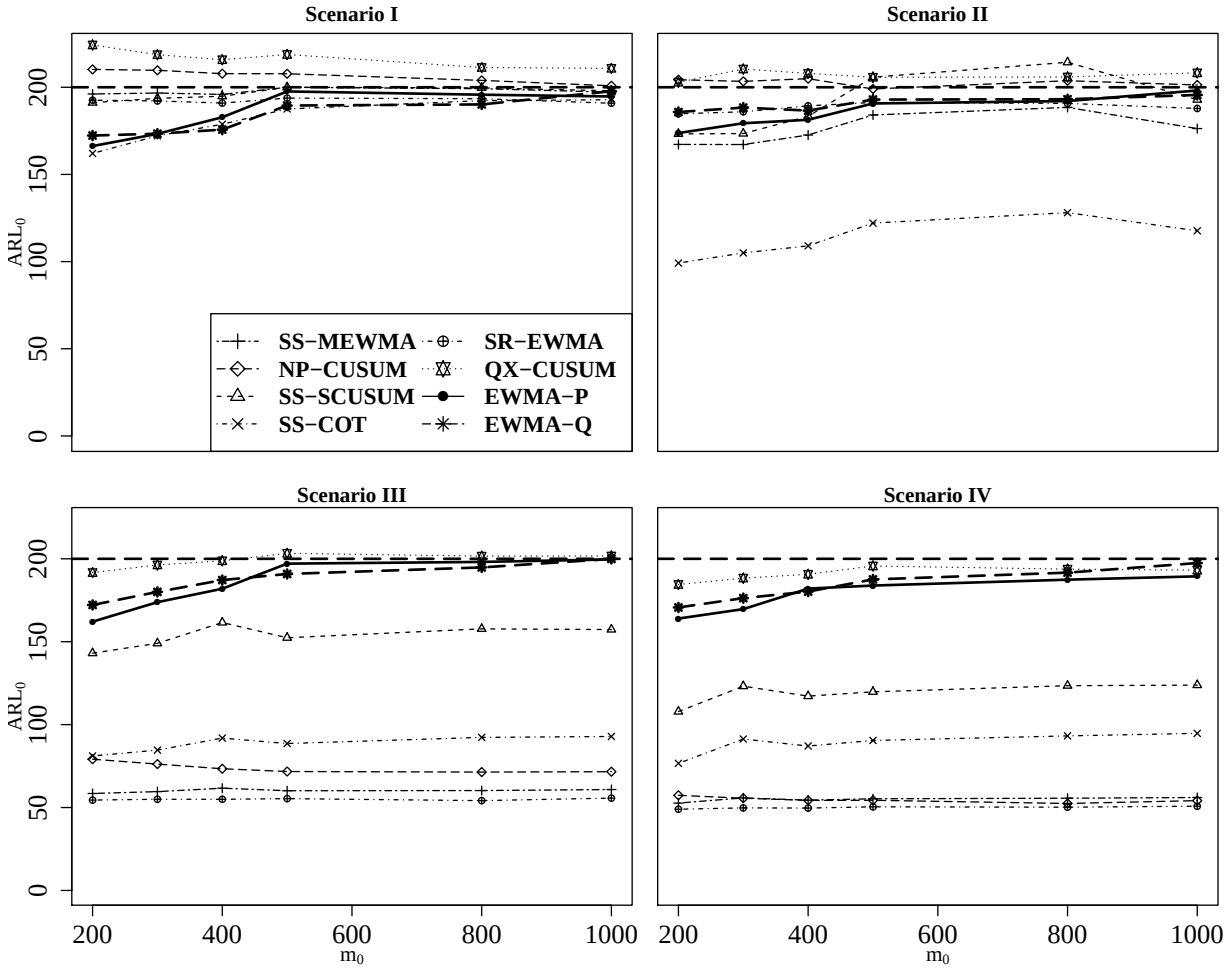

Figure 2: Estimated ARL_0 values of eight different control charts when their nominal ARL_0 values are fixed at 200, and the IC sample size m_0 changes among $\{200, 300, 400, 500, 800, 1000\}$.

To study the impact of the dimensionality p on the IC performance of the proposed charts EWMA-P and EWMA-Q, we consider an extension of Scenario I to a p-dimensional case, where p can change among $\{3, 5, 7\}$ and all other setups are kept the same as those in Figure 2. The estimated ARL_0 values of EWMA-P and EWMA-Q are presented in Figure 3. From the figure, it can be seen that the necessary IC sample size m_0 should be larger to have a reliable IC performance of the two charts when p is larger.

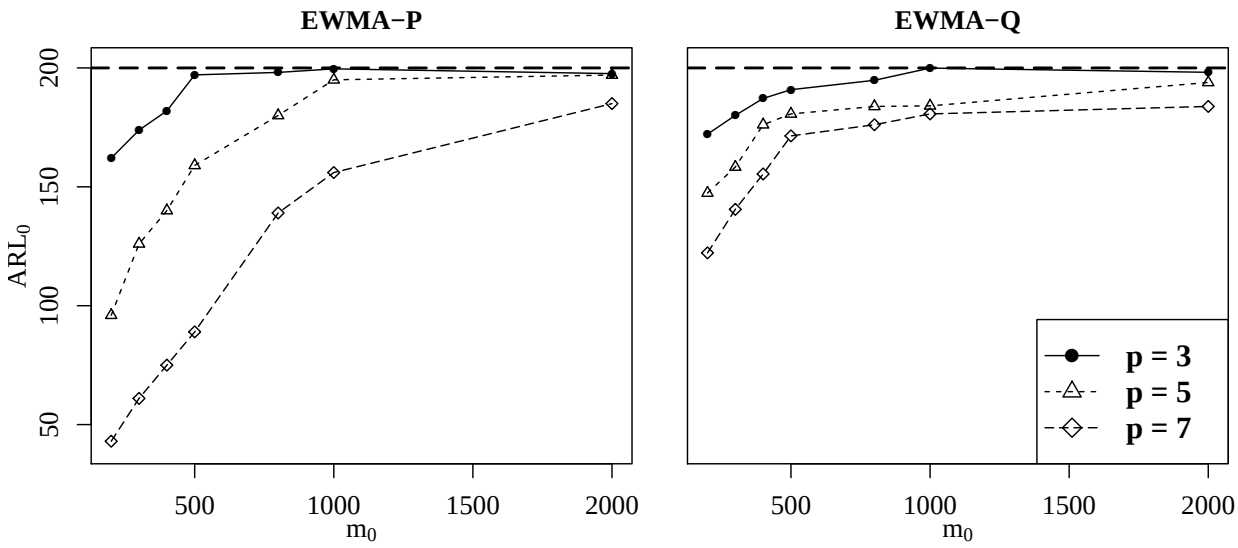

Figure 3: Estimated ARL_0 values of the charts EWMA-P and EWMA-Q when their nominal ARL_0 values are fixed at 200, and the IC sample size m_0 changes among $\{200, 300, 400, 500, 800, 1000, 2000\}$, and p changes among $\{3, 5, 7\}$.

3.2 OC performance

Next, we study the OC performance of the related control charts in cases when $p = 3$, $m_0 = 500$ and the nominal ARL_0 values of all charts are fixed at 200. To this end, it is assumed that a mean shift occurs at the beginning of process monitoring, and two cases of shifts are considered in this example. The shifted mean is $\boldsymbol{\mu} + \delta(1, 1, 1)'$ in Case I, and $\boldsymbol{\mu} + (0.5, \delta, \delta)'$ in Case II, where δ denotes the shift size and changes among $0, \pm 0.25, \pm 0.5, \pm 0.75$, and ± 1. In the above two cases, all three quality variables shift in the same size in Case I, and in different sizes in Case II. To make the comparison among different charts as fair as possible, their control limits have been adjusted properly so that their actual ARL_0 values all equal to the nominal ARL_0 value of 200. Also, for detecting a given shift, the weighting parameters of the four EWMA charts SS-MEWMA, SR-EWMA, EWMA-P, EWMA-Q and the allowance constants of the four CUSUM charts SS-COT, SS-SCUSM, NP-CUSUM, QX-CUSUM have all been searched such that the OC average run length ARL_1 of each chart reaches the minimum. Such a minimal ARL_1 value is called the optimal ARL_1 value, and the associated value of the procedure parameter is called the optimal procedure parameter value hereafter. So, the optimal OC performance of the related charts is compared here. It should be pointed out that the optimal ARL_1 values are considered here to make the comparison as fair as possible. Otherwise, the OC performance of a CUSUM chart with its allowance constant

chosen to be 0.1 may not be comparable to the OC performance of an EWMA chart with its weighting parameter chosen to be 0.1, for instance. The computed optimal ARL_1 values of the eight charts for detecting shifts in Cases I and II are presented in Figures 4 and 5, respectively.

From Figures 4 and 5, we can have the following conclusions. i) The charts SS-MEWMA, SS-SCUSUM and SS-COT perform well in Scenario I when process observations are i.i.d. and normally distributed. But, they are less effective in Scenarios II-IV when the normality and/or the data independence assumptions are violated. ii) The nonparametric charts NP-CUSUM and SR-EWMA have a good performance in Scenarios I and II when the process observations are independent at different observation times. But, they perform poorly in Scenarios III and IV when there is serial correlation in the observed data. iii) The EWMA-P chart has a reasonably good performance in Figure 4 where all quality variables have the same shift. But, its performance is not satisfactory in some cases shown in Figure 5 when the second and third quality variables have negative shifts or no shifts while the first quality variable has a positive shift of 0.5. iv) The charts QX-CUSUM and EWMA-Q both have a reasonably good performance in all cases considered, and EWMA-Q outperforms QX-CUSUM in most cases considered. Therefore, this example confirms that the EWMA-P chart could be ineffective for detecting certain shifts, as discussed in Proposition 2 and Examples 1 and 2, and the EWMA-Q chart should be considered if no prior information is available about a future shift direction. In addition, EWMA-Q usually performs better than QX-CUSUM since the latter is constructed based on data categorization that would result in information loss as discussed in Section 1.

The optimal values of the weighting parameters λ_1 and λ_2 at which the optimal ARL_1 values of the charts EWMA-P and EWMA-Q shown in Figures 4 and 5 are reached are presented in Figure 6. From the figure, it can be seen that the optimal values of λ_1 and λ_2 are larger for detecting larger shifts, which is consistent with the general guidelines on the selection of a weighting parameter in the literature that a larger value of the weighting parameter should be used for detecting a larger shift (cf., Qiu 2014, Chapter 5).

Next, we study the possible impact of the parameter b_{max} on the performance of the proposed charts EWMA-P and EWMA-Q. In the same setup as that in the example of Figure 4, let b_{max} change among $\{1, 5, 10, 20\}$. The calculated optimal ARL_1 values of EWMA-P and EWMA-Q are shown in Figure 7. From the plots in the figure, it can be seen that (i) the performance of both EWMA-P and EWMA-Q do not change much in Scenarios I and II when b_{max} changes, and (ii)

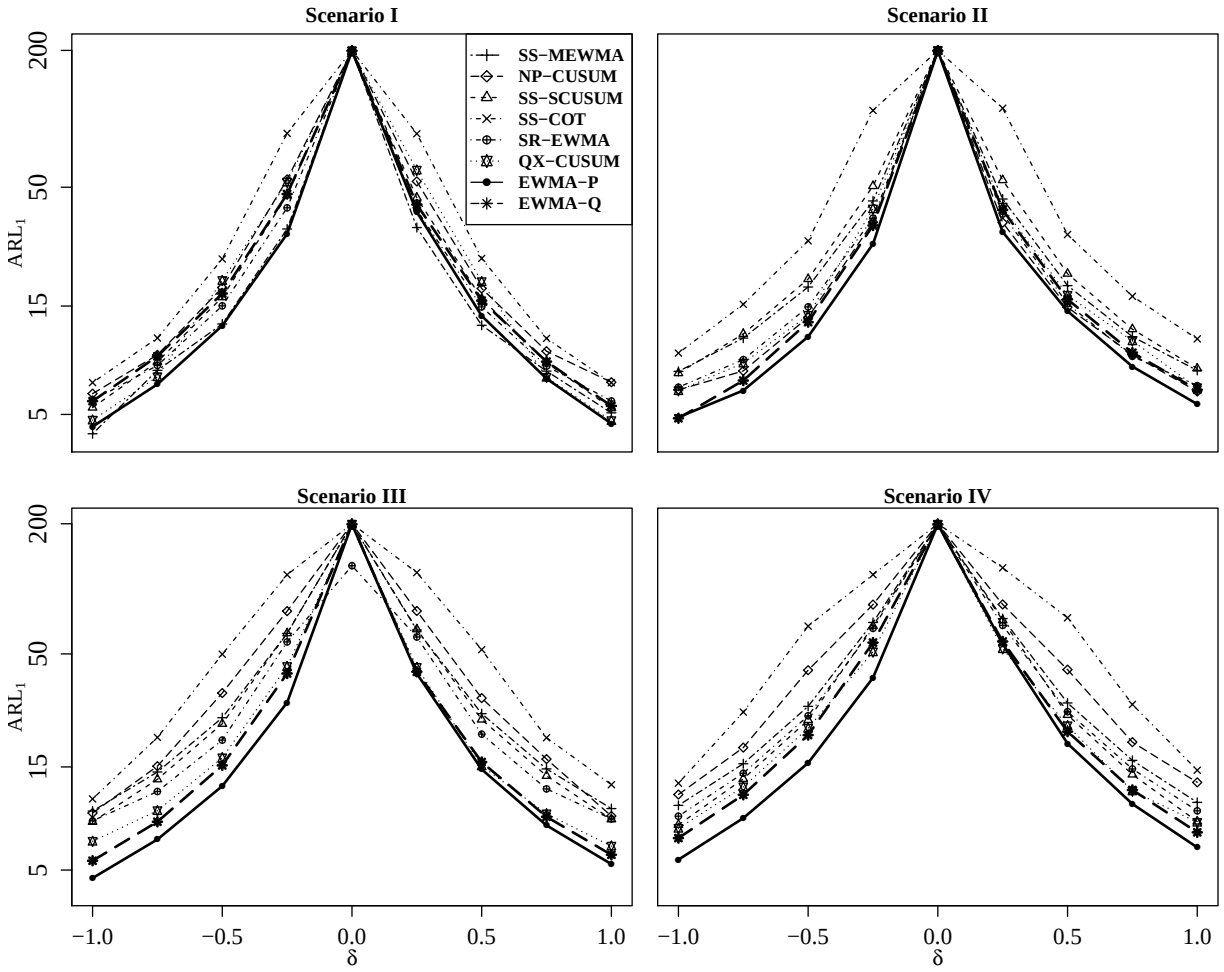

Figure 4: Optimal ARL_1 values of the eight control charts for detecting shifts in Case I, when the nominal ARL_0 values of all charts are fixed at 200, $p = 3$, $m_0 = 500$, and all quality variables have the same shift of size δ that changes among 0, ± 0.25, ± 0.5, ± 0.75, and ± 1.

the performance of both EWMA-P and EWMA-Q is unsatisfactory in Scenarios III and IV when $b_{max} = 1$, and their performance does not change much when $b_{max} \geq 5$. This example shows that the proposed charts EWMA-P and EWMA-Q would perform stably when the assumed time range of autocorrelation b_{max} is chosen to be 5 or larger in all scenarios considered.

In all previous examples, the mean shifts are focused. Next, let us consider an example in which both the mean and the variance of each quality variable have shifts. More specifically, it is assumed that each quality variable has a mean shift with a fixed size of 0.1, and a variance shift with size η that changes from 0 to 0.6 with a step 0.1. All other setups are kept to be the same as those in the example of Figures 4 and 5. The computed optimal ARL_1 values of the eight charts are presented in Figure 8. From the figure, we can have the following conclusions. (i) The two

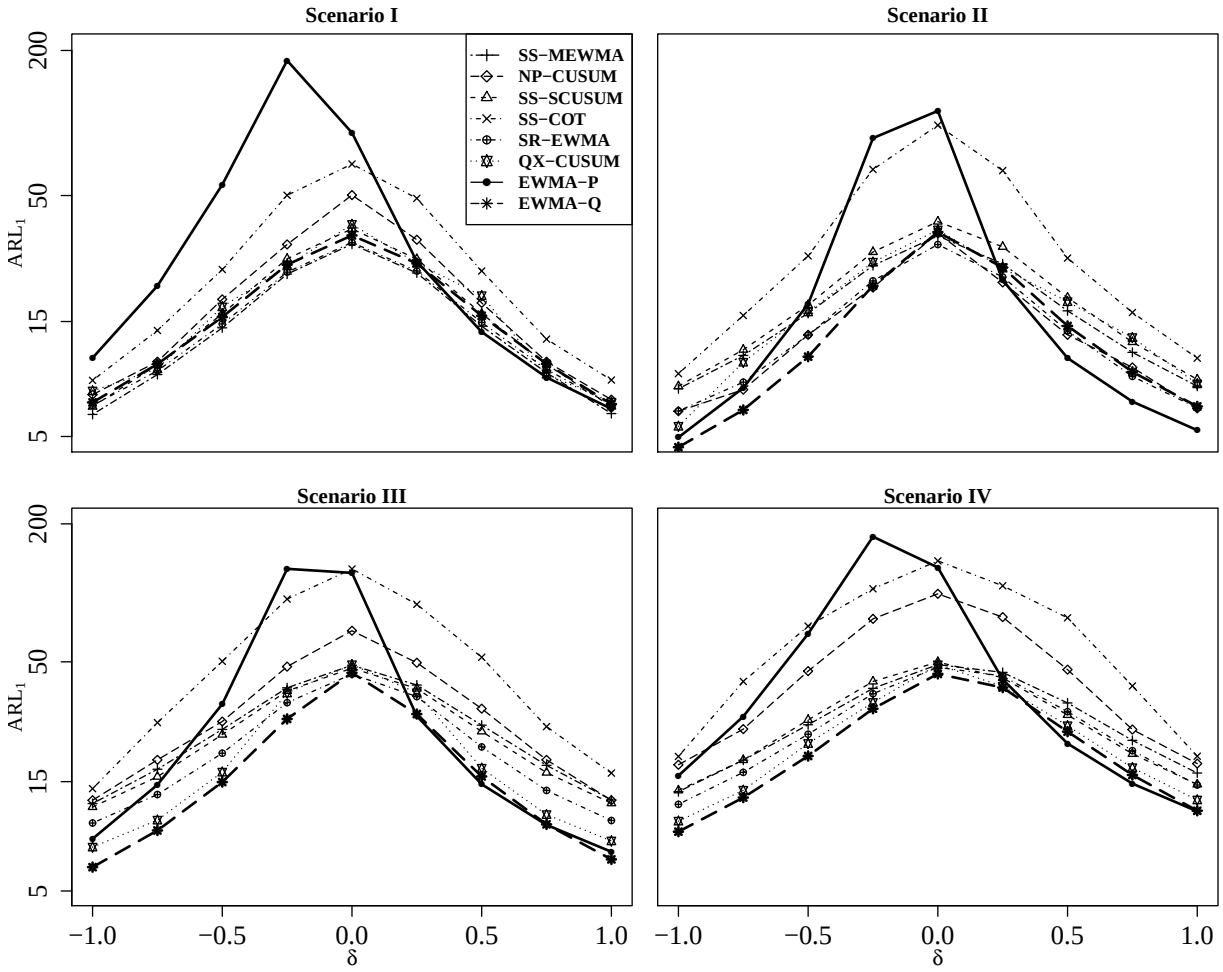

Figure 5: Optimal ARL_1 values of the eight control charts for detecting shifts in Case II, when the nominal ARL_0 values of all charts are fixed at 200, $p = 3$, $m_0 = 500$, the first quality variable has the shift of size 0.5, and the second and third quality variables have the same shift of size δ that changes among 0, ± 0.25, ± 0.5, ± 0.75, and ± 1.

nonparametric control charts NP-CUSUM and QX-CUSUM are not sensitive to the increase of the variance shift, and the nonparametric chart SR-EWMA is not very sensitive either. (ii) The charts SS-COT, SS-SCUSUM and SS-MEWMA perform well in Scenario I when their normality and "data independence" assumptions are valid, and are less effective to the increase of the variance shift in Scenarios II–IV when one or both of these assumptions are invalid. (iii) EWMA-Q performs better than EWMA-P in all cases considered, and EWMA-Q has the best performance among all eight charts in Scenarios III and IV when the IC process observations are serially correlated. This example shows that the proposed charts EWMA-P and EWMA-Q have a decent performance in comparison with its peers when both the process mean and the process variance have shifts.

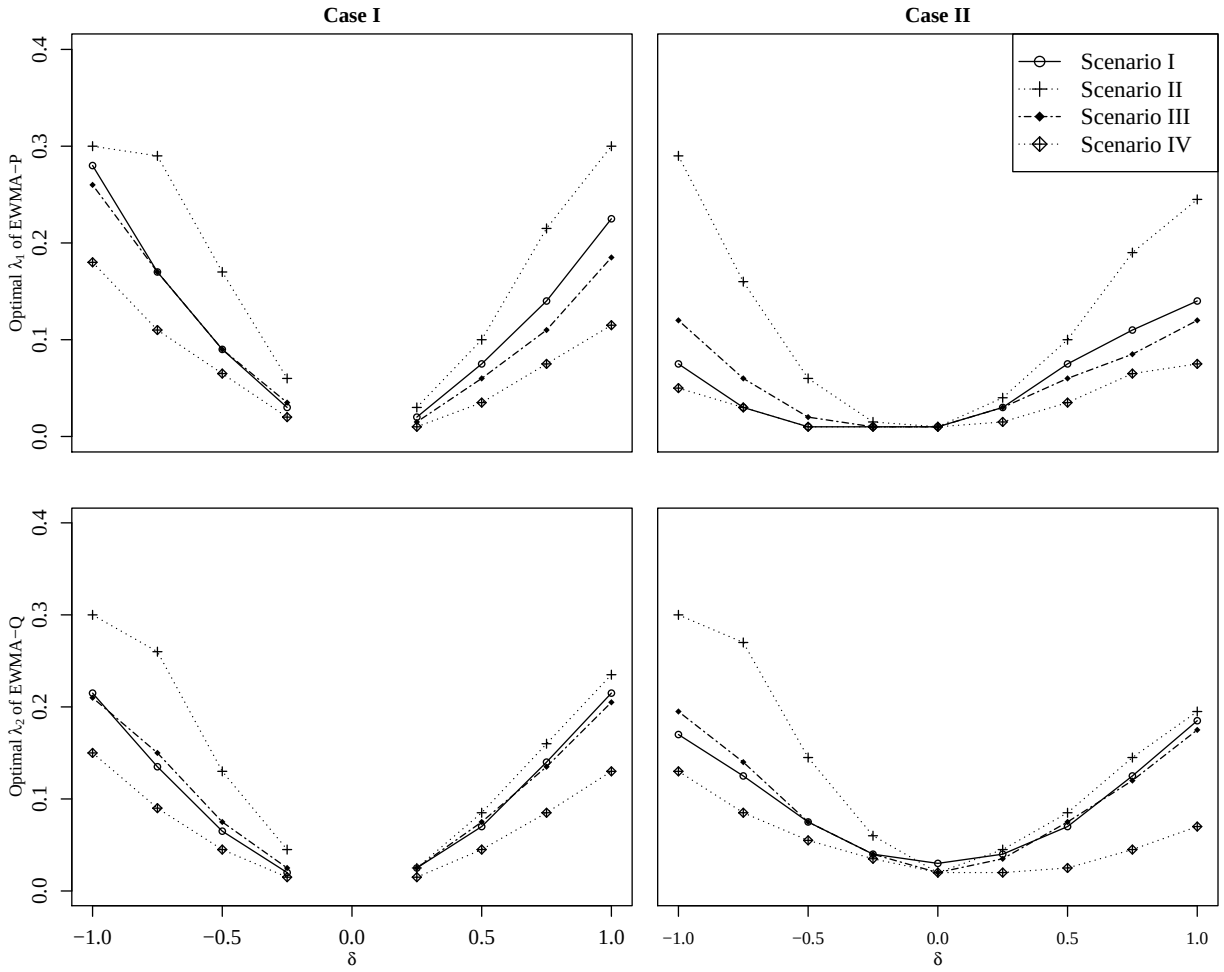

Figure 6: Optimal values of the weighting parameters λ_1 and λ_2 of the proposed charts EWMA-P and EWMA-Q in cases considered in Figures 4 and 5.

4 A Real-Data Application

In this section, we demonstrate the application of the proposed control charts by using a real dataset from a semiconductor manufacturing process. The dataset is available at the UC Irvine Machine Learning Repository with the link http://archive.ics.uci.edu/ml/datasets/SECOM. The dataset was collected from July 2008 to October 2008 by a computerized system that automatically manages a semiconductor manufacturing process consisting of a series of steps. At each step, observations of some quality variables, including film thickness, film uniformity, and electronic resistance, are collected by sensors at many key measurement points. So, many quality variables are measured at different steps of the manufacturing of a product. In this application, it is important to monitor the observed multivariate data streams collected from the production process for quality control

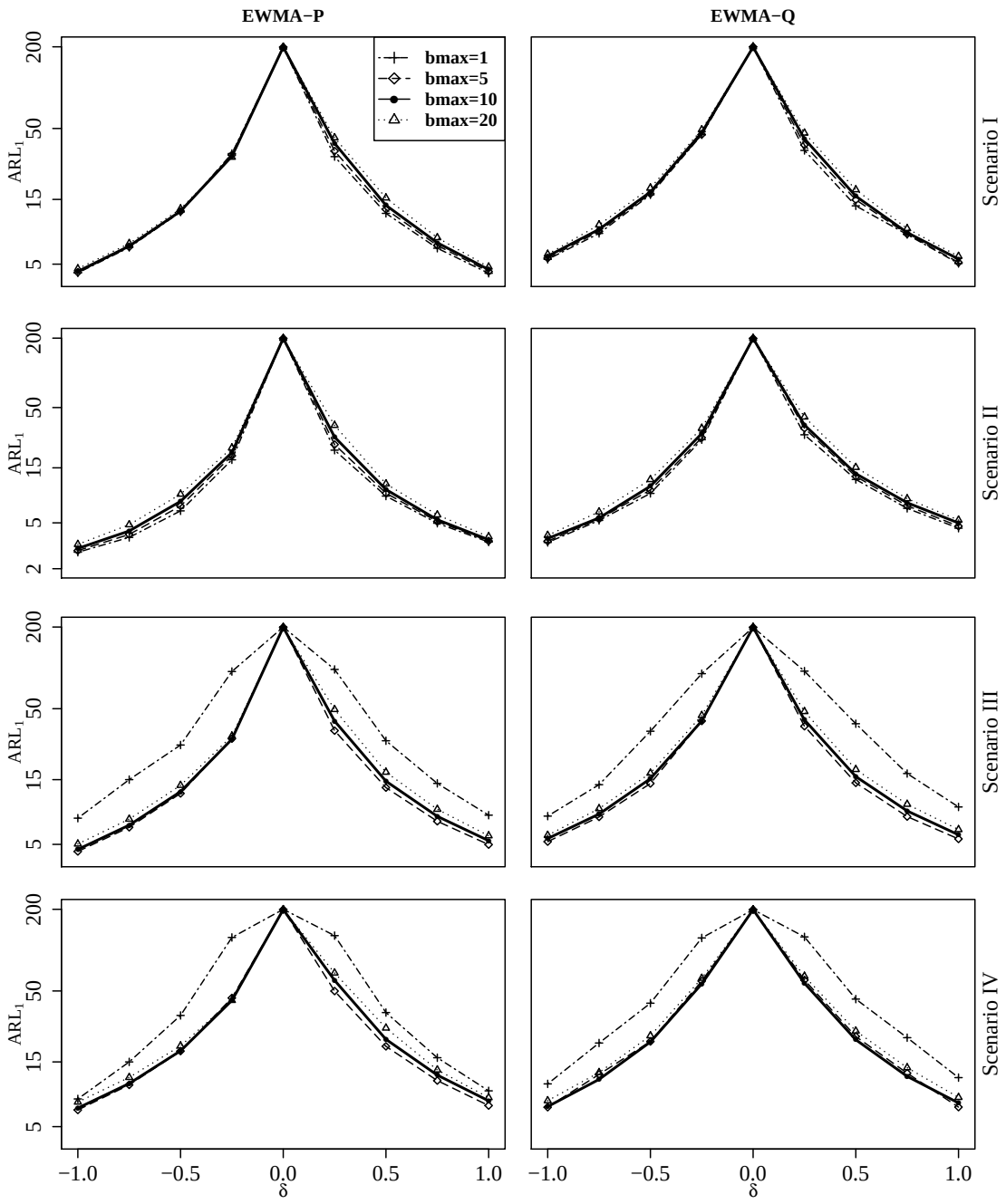

Figure 7: Optimal ARL_1 values of the proposed charts EWMA-P and EWMA-Q when the nominal ARL_0 values are fixed at 200, $p = 3$, $m_0 = 500$, all quality variables have the same shift size δ that changes among 0, ± 0.25, ± 0.5, ± 0.75, and ± 1, and the assumed time range of autocorrelation b_{max} changes among $\{1, 5, 10, 20\}$.

purposes. This dataset has a total of 590 quality variables and 1,567 observations of these variables. A total of 600 observations of three specific quality variables are selected here to demonstrate the application of the proposed control charts, which are shown in Figure 9. From the plots of the

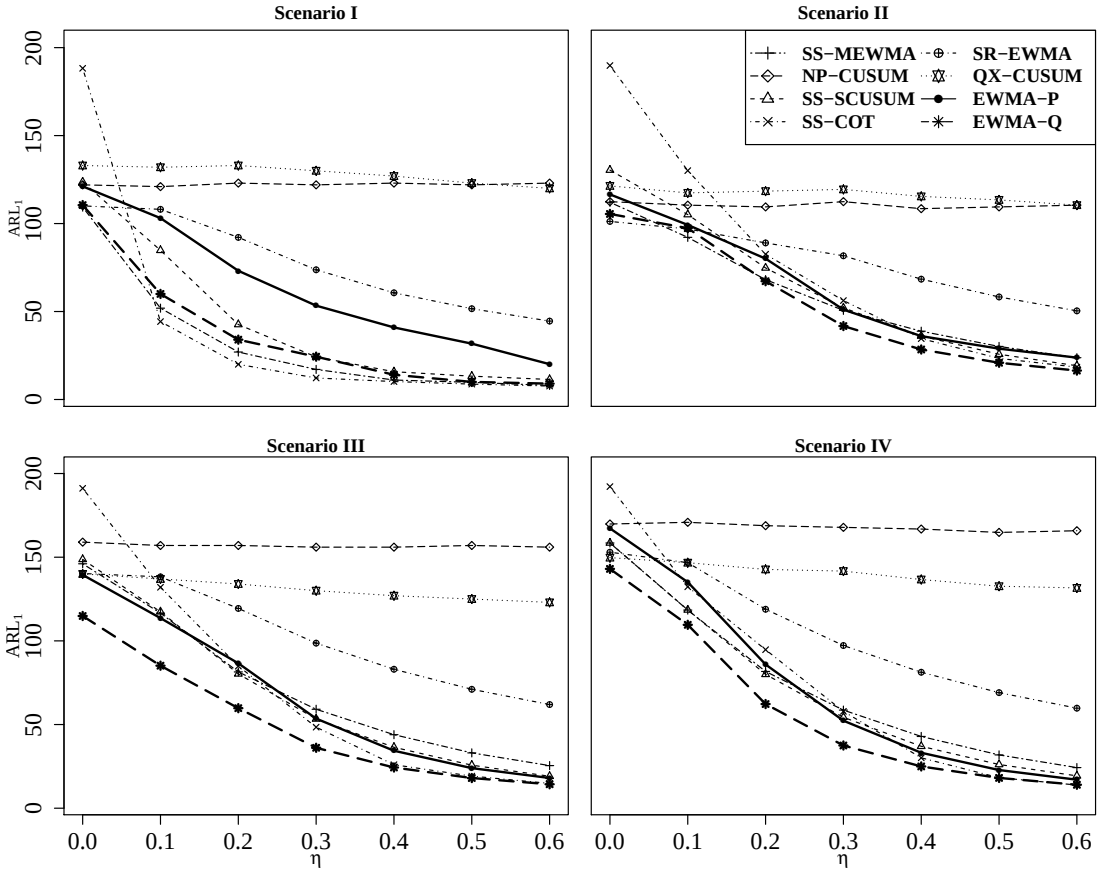

Figure 8: Optimal ARL_1 values of the eight control charts when the nominal ARL_0 values are fixed at 200, $p = 3$, $m_0 = 500$, and all quality variables have the same mean shift of size 0.1 and the same variance shift of size η that changes among $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$.

figure, it seems that the first 500 observations of each variable are quite stable. Thus, they are used as the initial IC data. The remaining 100 observations are used for online process monitoring. In Figure 9, the two types of observations are separated by a vertical line in each plot.

For the IC data, the p-values of the Durbin–Watson test for testing serial correlation for the three quality variables are 1.79×10^{-3}, 4.76×10^{-4}, and 9.74×10^{-2}, respectively. Thus, there is a significant autocorrelation in the observed IC data of the first and second quality variables. The Augmented Dickey-Fuller (ADF) test for testing stationarity of the autocorrelation gives the p-values that are < 0.01 for all quality variables, implying that it is reasonable to assume the serial correlation in the observed IC data of the three quality variables to be stationary. Therefore, we can conclude from these tests that the IC data have a significant stationary serial correlation in this example. To check the normality assumption for the IC data, the Shapiro test is performed. Its

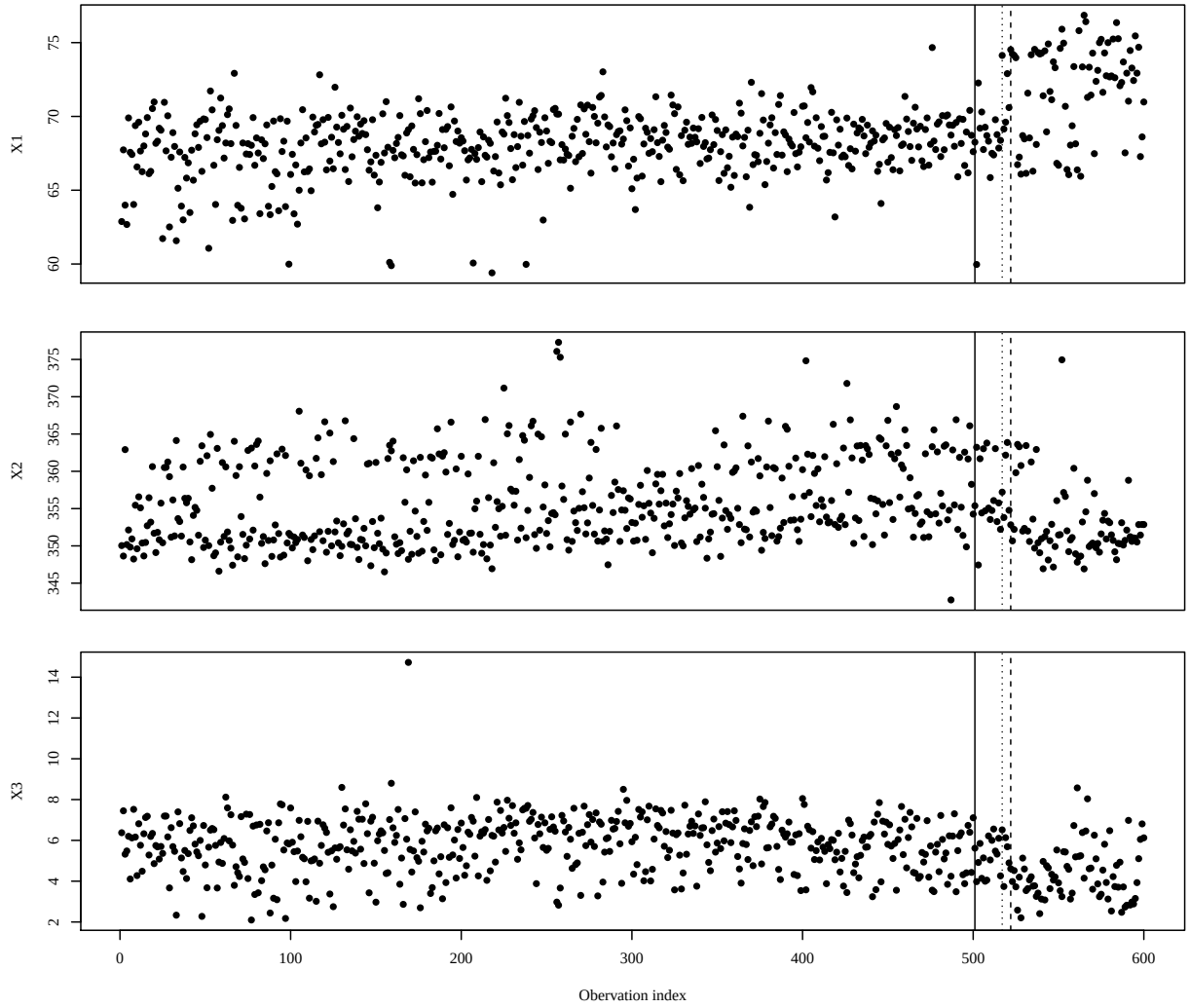

Figure 9: Original observations of three quality variables of a semiconductor manufacturing process. The solid vertical line in each plot separates the initial IC data and the data for online process monitoring, the dashed vertical line denotes the first signal time by the proposed chart EWMA-Q, and the dotted vertical line denotes the detected shift location by a change-point detection approach.

p-values for the three quality variables are 1.82×10^{-12}, 3.37×10^{-13} and 5.46×10^{-11}, respectively, implying that the IC process distribution is significantly different from a normal distribution.

Next, we discuss online process monitoring starting from the 501^{st} observation time. In all control charts, their nominal ARL_0 values are fixed at 200, and their control limits are computed in the same way as that in the simulation study discussed in Section 3. The charting statistics of the

eight charts are presented in Figure 10, where the dashed horizontal lines denote the related control limits. From the figure, it can be seen that the seven charts SS-MEWMA, SS-COT, SS-SCUSUM, SR-EWMA, QX-CUSUM, EWMA-P and EWMA-Q give their first signals at the $523^{rd}, 524^{th}, 536^{th}$, $523^{th}, 523^{th}$, 539^{th} and 522^{th} observation times, respectively, while the chart NP-CUSUM gives signals many times at the beginning of online process monitoring. Because the "data independence" assumption required by the chart NP-CUSUM is violated in this example as discussed above, its result may not be reliable. To check whether the signals given by the related control charts are reliable, the change-point detection approach based on the generalized maximum likelihood estimation (cf., Qiu 2014, Section 7.5) is applied to the transformed data under online monitoring (i.e., $\{\mathbf{Z}_n = (Z_{n1}, Z_{n2}, \ldots, Z_{np})', n = 501, 502, \ldots, 600\}$). The detected change-point is at $n = 517$. Then, the Hotelling's T^2 test is used to check whether the means of the two groups of transformed data with the observation times in [501,516] and [517,600], respectively, are significantly different, and the resulting p-value is 1.568×10^{-3}. So, we conclude that the two group means are significantly different, and a process mean shift at the time $n = 517$ is then confirmed which is denoted by the dotted vertical lines in Figure 9. Therefore, among the eight control charts, the signal by NP-CUSUM may not be reliable as explained above and the first signal by EWMA-Q, which is denoted by the dashed vertical lines in Figure 9, is the earliest among the remaining seven charts in this example. As a side note, from Figure 9, it seems that the shifts of the three quality variables have different signs (i.e., the first quality variable has a positive shift while the second and third ones have negative shifts), which explains why the chart EWMA-P is ineffective in this example, as demonstrated by the simulation results shown in Figure 5.

5 Concluding Remarks

Sequential monitoring of multivariate processes when their IC distributions cannot be described well by parametric forms is an important and challenging research problem. There have been some existing discussions on this topic. One existing approach is to use the ordering information among process observations or data categorization for constructing nonparametric control charts. Because much information in the original process observations would be lost by considering data ordering or data categorization, the effectiveness of these nonparametric charts would be compromised. Another approach is to use the conventional control charts with their procedure parameters chosen

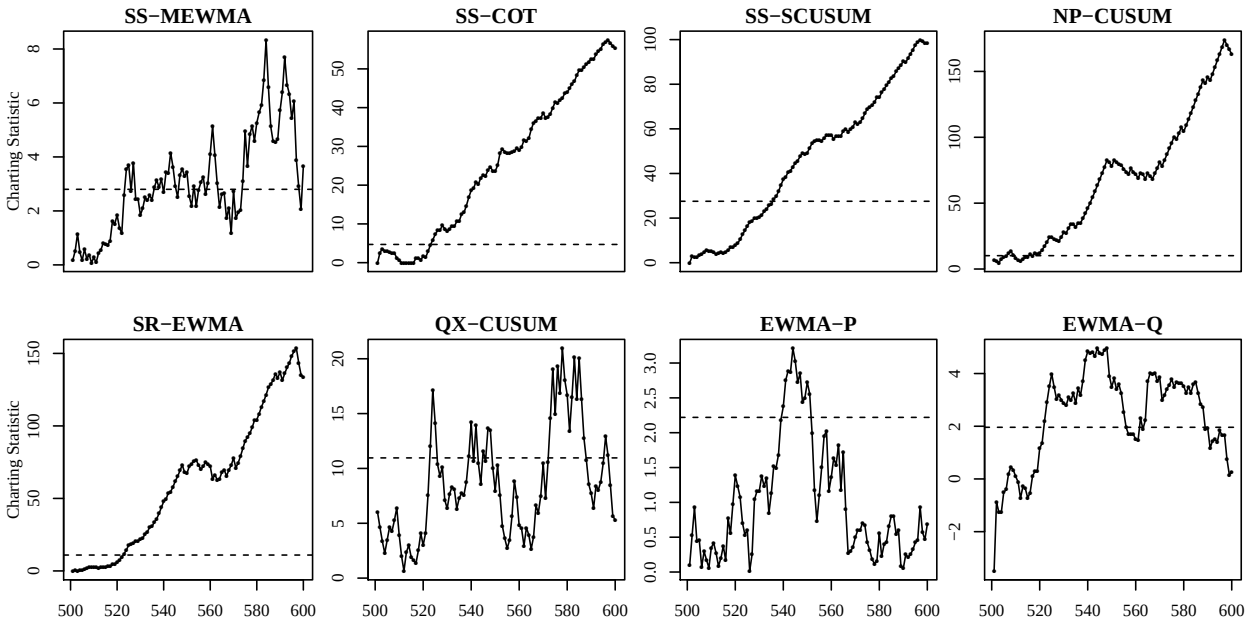

Figure 10: Control charts for monitoring the last 100 process observations shown in Figure 9. In each plot, the horizontal dotted line denotes the control limit of the related control chart.

properly. However, because it is often unknown in advance how different between the actual IC process distribution and a parametric (e.g., normal) distribution, implementation of this approach is difficult and the resulting control chart may not be effective for detecting relatively large shifts. In this paper, we made another research effort to solve this challenging process monitoring problem. In the previous sections, we have described a general framework for robust monitoring of multivariate processes with serially correlated data, based on data decorrelation, data transformation, and data integration. Two new control charts (i.e., EWMA-P and EWMA-Q) are suggested by using two different approaches of data integration. Because the new robust process monitoring framework is flexible, it should be able to provide a powerful tool for many process monitoring applications. Based on intensive numerical studies, we conclude that both charts are reliable to use, the chart EWMA-P may not be effective for detecting shifts close to the ones satisfying Equation (6), and the chart EWMA-Q should be effective for detecting any shifts. Thus, the chart EWMA-Q is recommended to use in practice if there is no prior information on the future shift direction.

For the proposed robust process monitoring framework, there are still some issues to address in the future. For instance, when the number of quality variables is large, the current process monitoring charts EWMA-P and EWMA-Q may not be effective when shifts occur only in a small number of quality variables. In such cases, some variable selection procedures might be helpful in

constructing the related control chart (cf., Zou and Qiu 2009). In the current version of the proposed robust process monitoring framework, it is assumed that possible serial correlation in the observed data is short-ranged and stationary. These assumptions should be reasonable in some applications. But, in some other applications, serial correlation may not be stationary and/or short-ranged. As a matter of fact, it has been confirmed that some processes in practice contain long-range serial correlation (cf., Beran 1992, Giraitis et al., 2012). In such cases, estimation of serial correlation becomes challenging. One possible approach is to use the covariance matrix function and estimate it by a kernel smoothing approach (cf., Xie and Qiu 2023). All these issues will be studied carefully in our future research.

Supplementary Materials

ComputerCodesAndData.zip: This zip file contains some computer codes to implement the proposed method and the real data used in the paper.
supplement.pdf: This supplementary file contains the proof of the result in Proposition 1.

Acknowledgments

The authors thank the editor, the associate editor, and four referees for many constructive comments and suggestions which improved the quality of the paper greatly. This research is supported in part by an NSF grant.

References

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

Bickel, P.J. and Levina, E (2008), "Covariance regularization by thresholding," *The Annals of statistics*, **36**, 2577–2604.

Borror, C.M., Montgomery, D.C. and Runger, G.C. (1999), "Robustness of the EWMA control chart to non-normality," *Journal of Quality Technology*, **31**, 309–316.

Brean, J. (1992), "Statistical methods for data with long-range dependence," *Statistical Science*, **4**, 404–416.

Capizzi, G., and Masarotto, G. (2008), "Practical design of generalized likelihood ratio control charts for autocorrelated data," *Technometrics*, **50**, 357–370.

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

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

Chakraborti, S., van der Laan, P. and van de Wiel, M. (2004), "A class of distribution-free control charts," *Journal of the Royal Statistical Society (Series C)*, **53**, 443–462.

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

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

Hawkins, D.M. (1987), "Self-starting cusums for location and scale," *The Statistician*, **36**, 299–315.

Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag.

Hawkins, D.M., Qiu, P., and Kang, C.W. (2003), "The changepoint model for statistical process control," *Journal of Quality Technology*, **35**, 355–366.

Higham, N.J. (1988), "Computing a nearest symmetric positive semidefinite matrix," *Linear Algebra and its Applications*, **103**, 103–118.

Krupskii, P., Harrou, F., Hering, A.S. and Sun, Y. (2020), "Copula-based monitoring schemes for non-Gaussian multivariate processes," *Journal of Quality Technology*, **52**, 219–234.

Li, J. (2021), "Nonparametric adaptive CUSUM chart for detecting arbitrary distributional changes," *Journal of Quality Technology*, **53**, 154–172.

Li, W., Pu, X., Tsung, F., and Xiang, D. (2017), "A robust self-starting spatial rank multivariate EWMA chart based on forward variable selection," *Computers & Industrial Engineering*, **103**, 116–130.

Li, W., and Qiu, P. (2020), "A general charting scheme for monitoring serially correlated data with short-memory dependence and nonparametric distributions," *IISE Transactions*, **52**, 61–74.

Giraitis, M., Koul, H., and Surgailis, D. (2012), *Large sample inference for long memory processes,* World Scientific.

Mei, Y. (2010), "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, **97**, 419–433.

Montgomery, D.C. (2012), *Introduction to Statistical Quality Control,* New York: John Wiley & Sons.

Page, E.S. (1954), "Continuous inspection scheme," *Biometrika*, **41**, 100–115.

Peña, D., and Prieto, F.J. (2001), "Multivariate outlier detection and robust covariance matrix estimation," *Technometrics*, **43**, 286–310.

Pourahmadi, M. (2013), *High-Dimensional Covariance Estimation: With High-Dimensional Data,* New York: John Wiley & Sons.

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

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

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

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

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

Qiu, P., and Xie, X. (2022), "Transparent sequential learning for statistical process control of serially correlated data," *Technometrics*, **64**, 487–501.

Roberts, S.V. (1959), "Control chart tests based on geometric moving averages," *Technometrics*, **1**, 239–250.

Shewhart, W.A. (1931), *Economic Control of Quality of Manufactured Product,* New York: D. Van Nostrand Company.

Stoumbos, Z., and Sulliva, J. (2002), "Robustness to non-normality of the multivariate EWMA control chart," *Journal of Quality Technology*, **34**, 260–276.

Sullivan, J.H., and Jones, L.A. (2002), "A self-starting control chart for multivariate individual observations," *Technometrics*, **44**, 24–33.

Testik, M.C., Runger, G.C., and Borror, C.M. (2003), "Robustness properties of multivariate EWMA control charts," *Quality and Reliability Engineering International*, **19**, 31–38.

Tucker, H.G. (1959), "A generalization of the Glivenko–Cantelli theorem," *The Annals of Mathematical Statistics*, **30**, 828–830.

Xie, X., and Qiu, P. (2023), "Control chart for dynamic process monitoring with an application to air pollution surveillance," *Annals of Applied Statistics*, **17**, 47–66.

Yu, G., Zou, C., and Wang, Z. (2012), "Outlier detection in functional observations with applications to profile monitoring," *Technometrics*, **54**, 308–318.

Zhang, C., Chen, N., and Zou, C. (2016), "Robust multivariate control chart based on goodness-of-fit test," *Journal of Quality Technology*, **48**, 139–161.

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

Zou, C., and Tsung, F. (2010), "Likelihood ratio-based distribution-free EWMA control charts," *Journal of Quality Technology*, **42**, 1–23.

Zou, C., and Tsung, F. (2011), "A multivariate sign EWMA control chart," *Technometrics*, **53**, 84–97.

Zou, C., Wang, Z., and Tsung, F. (2012), "A spatial rank-based multivariate EWMA control chart," *Naval Research Logistics*, **59**, 91–110.