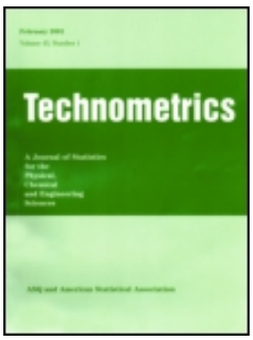

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://www.tandfonline.com/loi/utch20

A Distribution-Free Multivariate Control Chart

Nan Chen, Xuemin Zi & Changliang Zou

To cite this article: Nan Chen, Xuemin Zi & Changliang Zou (2016) A Distribution-Free Multivariate Control Chart, Technometrics, 58:4, 448-459, DOI: 10.1080/00401706.2015.1049750

To link to this article: http://dx.doi.org/10.1080/00401706.2015.1049750

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

A Distribution-Free Multivariate Control Chart

Nan Chen

Department of Industrial & Systems Engineering
National University of Singapore
Singapore
(isecn@nus.edu.sg)

Xuemin Zi

School of Science
Tianjin University of Technology and Education
Tianjin, China
(zi_xuemin@aliyun.com)

Changliang Zou

Institute of Statistics and LPMC
Nankai University
Tianjin, China
(nk.chlzou@gmail.com)

Monitoring multivariate quality variables or data streams remains an important and challenging problem in statistical process control (SPC). Although the multivariate SPC has been extensively studied in the literature, designing distribution-free control schemes are still challenging and yet to be addressed well. This article develops a new nonparametric methodology for monitoring location parameters when only a small reference dataset is available. The key idea is to construct a series of conditionally distribution-free test statistics in the sense that their distributions are free of the underlying distribution given the empirical distribution functions. The conditional probability that the charting statistic exceeds the control limit at present given that there is no alarm before the current time point can be guaranteed to attain a specified false alarm rate. The success of the proposed method lies in the use of data-dependent control limits, which are determined based on the observations online rather than decided before monitoring. Our theoretical and numerical studies show that the proposed control chart is able to deliver satisfactory in-control run-length performance for any distributions with any dimension. It is also very efficient in detecting multivariate process shifts when the process distribution is heavy-tailed or skewed. Supplementary materials for this article are available online.

KEY WORDS: Conditionally distribution-free; Empirical distribution; Nonparametric procedure; Robustness; Self-starting; Statistical process control.

1. INTRODUCTION

In modern quality control, it is common to monitor several quality characteristics of a process simultaneously (Stoumbos et al. 2000; Woodall and Montgomery 2013). This is called multivariate statistical process control (MSPC) in the literature and it is the focus of this article. One of the tasks of MSPC is to detect the change in location parameters $\boldsymbol{\mu}$ of a multivariate process as quickly as possible. It is usually assumed that there are m_0 independent and identically distributed (iid) historical (reference) observations, $\mathbf{X}_{-m_0+1}, \ldots, \mathbf{X}_0 \in \mathbb{R}^p$, for some integer, $p \geq 1$, and the ith future observation, $\mathbf{X}_i = (X_{1i}, \ldots, X_{pi})^T$, is collected over time following the multivariate location change-point model

$$\mathbf{X}_i \overset{\text{iid}}{\sim} \begin{cases} F_0(\mathbf{x}; \boldsymbol{\mu}_0) & \text{for} \quad i = -m_0 + 1, \ldots, 0, 1, \ldots, \tau, \\ F_1(\mathbf{x}; \boldsymbol{\mu}_1), & \text{for} \quad i = \tau + 1, \ldots, \end{cases} \tag{1}$$

where τ is the unknown change point. The F_0 and F_1 are the in-control (IC) and out-of-control (OC) distribution functions, respectively, and are assumed to be continuous. In practice, they can be the same or different, but their location parameters $\boldsymbol{\mu}_0 = (\mu_{10}, \ldots, \mu_{p0})^T$ and $\boldsymbol{\mu}_1 = (\mu_{11}, \ldots, \mu_{p1})^T$ that we are interested in are assumed to be unequal.

Due to both simplicity and nice properties, the approaches based on the quadratic formulation of the test statistics, $(\mathbf{X}_i - \widehat{\boldsymbol{\mu}}_0)^T \widehat{\boldsymbol{\Sigma}}_0^{-1} (\mathbf{X}_i - \widehat{\boldsymbol{\mu}}_0)$, are popular and have been well studied in the literature, where $\widehat{\boldsymbol{\mu}}_0$ and $\widehat{\boldsymbol{\Sigma}}_0$ are, respectively, the mean vector and covariance matrix estimated from the IC reference sample of size m_0. Among others, the multivariate cumulative sum (CUSUM) (Crosier 1988; Pignatiello and Runger 1990), the multivariate exponentially weighted moving average (MEWMA; Lowry et al. 1992), the regression-adjusted control chart (Hawkins 1991), and variable-selection-based control charts (Zou and Qiu 2009) were developed. See also Qiu (2014, chap. 7) for a nice summary. However, in many applications, m_0 is not large and in some cases it is in fact rather small. There would be considerable uncertainty in the parameter estimation, which in turn would distort the IC or OC run-length distribution of the control charts (Champ, Jones-Farmer, and Rigdon 2005; Zantek 2006). In the situation where a sufficiently large reference dataset is unavailable, self-starting methods that handle sequential monitoring by simultaneously updating parameter estimates and checking for OC conditions would be of use. For MSPC, see Quesenberry (1997), Sullivan and Jones (2002), Zamba and Hawkins (2006), and Hawkins and Maboudou-Tchao (2007).

These MSPC research works are mostly based on the fundamental assumption that either the process data have multinormal

TECHNOMETRICS, NOVEMBER 2016, VOL. 58, NO. 4
DOI: 10.1080/00401706.2015.1049750
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/tech.

distributions or the IC distribution F_0 (even F_1) is known. However, it is well recognized that in many applications, the underlying process distribution is unknown and not multinormal, so that the statistical properties of commonly used charts, which were designed to perform best under the normal distribution (or F_0), could potentially be (highly) affected. Nonparametric or robust charts may be useful in such situations. In the last several years, univariate nonparametric control charts have attracted much attention from researchers and a nice overview of this topic was presented in Qiu (2014, chap. 8). Some effort has been devoted to robust MSPC: Liu (1995) proposed control schemes based on data-depth. Qiu and Hawkins (2001) suggested a computationally trivial nonparametric multivariate CUSUM procedure based on the antiranks of the measurement components. Stoumbos and Sullivan (2002) recommended the classical MEWMA chart because it is robust in the sense that the IC run-length distribution for a continuous nonnormal process is quite close to that for a multinormal process with the same control limit if the smoothing parameter λ is small. Qiu (2008) proposed a nonparametric multivariate CUSUM procedure based on log-linear modeling. Zou and Tsung (2011) developed a multivariate sign EWMA (MSEWMA) chart by adapting the spatial-sign test to online sequential monitoring, resulting in a nonparametric counterpart of the MEWMA chart. Boone and Chakraborti (2012) proposed a Shewhart-type control chart based on componentwise ranking procedure, and used asymptotic results to find threshold values.

Although these nonparametric monitoring methods are able to detect shifts regardless of the underlying distributions, they are not "*distribution-free*," in the sense that the charting procedures are not guaranteed to attain the nominal IC average run-length (ARL_0) without knowing the IC distribution F_0. When F_0 is unknown but sufficient reference samples (large m_0) are available, the IC run-length distributions of the aforementioned nonparametric charts, or even parametric charts (e.g., MEWMA or MCUSUM) can be designed to approximately achieve the nominal one by means of simulations through resampling from the m_0 IC historical samples directly (Zou and Tsung 2011). However, how large m_0 should be depends on the dimension p and is always difficult to quantify. Zou and Tsung (2011) used simulated examples to show that the performances of MSEWMA and MEWMA are similarly affected when m_0 is small and very large Phase I samples must be collected for both charts to perform as well as those with known parameters. At least a reference sample of size 4000 is required to make MSEWMA or MEWMA attain the nominal ARL_0 for a five-dimensional case ($p = 5$). However, gathering such a large size of reference samples may be costly in practice; it may not be feasible to wait for the accumulation of sufficiently large calibration samples, especially in short-run processes. Zou, Wang, and Tsung (2012) developed a multivariate self-starting methodology for monitoring location parameters by using spatial-ranks, termed as the SREWMA chart. This method only requires a small m_0 to have satisfactory IC run-length distribution. However, its distribution-free property holds exactly only over a limited class of distributions (elliptical distributions). See also Holland and Hawkins (2014) for a related approach based on change-point models.

The main objective of this article is to develop a *distribution-free* method for online monitoring of multivariate processes or data streams with any dimension p when only a few historical observations are available. In other words, we require that the charting procedure should always attain the nominal IC run-length distribution regardless the distribution type or dimension of F_0, even when F_0 is unknown or m_0 is small. There are several challenges associated with this task. First, the insufficient reference sample prevents us from using resampling approaches or some estimation of distribution functions to find the control limits as done in Sun and Tsung (2003) and Qiu (2008). Second, it is challenging to design a distribution-free scheme for high-dimensional cases. Many charts, such as those proposed by Stoumbos and Sullivan (2002), Zou and Tsung (2011), and Zou, Wang, and Tsung (2012) are only approximately distribution-free in a low-dimensional situation. Third, the components in $\mathbf{X}_i$ are often correlated. If the components are independent, we could use multiple univariate distribution-free control charts (e.g., Zhou et al. 2009; Zou and Tsung 2010) to monitor each component, respectively. Such a joint scheme with multiple independent charts can achieve the nominal IC run-length distribution by adjusting false alarm rates of each individual chart. However, the unknown between-component correlation structure makes this method infeasible.

To address these challenges, this article proposes a new control chart that is in principle distribution-free for any distributions with any dimension p. The key idea is to construct a series of distribution-free test statistics conditioning on the empirical distribution functions at each observation time. Essentially, the proposed chart uses data-dependent limits that are determined online and depend on the current and past samples. Those control limits maintain the conditional false alarm probability (the probability that the charting statistic exceeds the control limit conditioning on that there is no alarm before the current time point) to a specified value at each time point. As a result, the IC run-length distribution of the proposed chart is *always* identical to the Geometric distribution regardless the IC distribution F_0. The proposed control chart is also sensitive to the process change and has good performance compared to the existing works. The proposed control chart is particularly useful in start-up or short-run situations because it does not require either knowledge of F_0 or unrealistically large reference samples to have robust performance.

The remainder of this article is organized as follows: our proposed methodology is described in detail in Section 2; the method is demonstrated in Section 3 using real-data examples from manufacturing processes; its numerical performance is thoroughly investigated in Section 4; finally several remarks in Section 5 conclude the article.

2. METHODOLOGY

Our proposed methodology is described in four parts. In Section 2.1, a distribution-free two-sample test and its properties are discussed. A distribution-free multivariate charting scheme is developed in Section 2.2. Its algorithm and computational issues are addressed in Section 2.3. Section 2.4 provides some practical guidelines for the design of the proposed method. Proofs of propositions and theorems in this section are included in supplementary Section S1. In addition, some potential extensions are discussed in supplementary Section S2.

2.1 A Distribution-Free Two-Sample Location Test

The monitoring problem (1) is closely related to nonparametric statistical tests of hypotheses for the two-sample location problem in the context of multivariate statistical analysis. Hence, to facilitate the derivation of the proposed charting statistic, we start by assuming that two independent samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_m\}$, $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ are distributed as $F_0(\mathbf{t}; \boldsymbol{\mu}_0)$ and $F_0(\mathbf{t}; \boldsymbol{\mu}_1)$, respectively. We want to test the null hypothesis, H_0, that $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$ against H_1 that $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$.

We are interested in the distribution-free tests for such location problem. In other words, we require that the distribution of the test statistic does not depend on F_0. When F_0 is any univariate distribution, distribution-free tests have been developed based on ranks and signs derived from the ordering of the data (e.g., Chakraborti, Van Der Laan, and Bakir 2001; Ross, Tasoulis, and Adams 2011). However, when F_0 is multivariate, a natural ordering of the data does not exist. Some tests based on spatial sign or spatial rank (see Oja 2010) are only distribution-free under certain models and assumptions; see Zou and Tsung (2011) and Zou, Wang, and Tsung (2012) for related discussion. On the other hand, a distribution-free test based on multivariate empirical distributions was proposed by Bickel (1969), but it has been demonstrated that it does not perform satisfactorily when p is large or the sample size is small, due to the so-called curse of dimensionality. To this end, we propose an intuitively appealing test caters to the two-sample location problem under the framework laid out by Bickel (1969).

The null hypothesis $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$ is equivalent to $\mu_{j0} = \mu_{j1}$ for $j = 1, \ldots, p$. So, it is natural to consider the Wilcoxon rank-sum test for each component, say

$$T_j = \frac{\sum_{i=1}^{m} R_{ji} - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}}, \tag{2}$$

where R_{ji} is the rank of X_{ji} among the pooled sample $\{X_{j1}, \ldots, X_{jm}, Y_{j1}, \ldots, Y_{jn}\}$. As long as $\mu_{j0} \neq \mu_{j1}$, $|T_j|$ would become large. Next, combine all the information of T_j for $j = 1, \ldots, p$ together, by using the "max" or "$\sum$" operator. For example, consider $T = \sum_j T_j^2$. A large value of T would lead to the rejection of H_0.

Unfortunately, the null distribution of T depends on F_0, and is not available in practice due to its dependency on the associations (correlations) between the variables. To make the test based on T distribution-free, we use an important conditionally distribution-free property of T. In more details, under null hypothesis, $\mathbf{X}_i$ and $\mathbf{Y}_j$ are identically distributed. Let $\hat{F}_{m+n}(\mathbf{t})$ denote the empirical cumulative distribution function (E.C.D.F.) of pooled samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ for $\mathbf{t} \in \mathbb{R}^p$, that is, $\hat{F}_{m+n}(\mathbf{t}) = (m+n)^{-1}[\sum_{i=1}^{m} I(\mathbf{X}_i \leq \mathbf{t}) + \sum_{i=1}^{n} I(\mathbf{Y}_i \leq \mathbf{t})]$, where the inequality is understood as componentwise version. We shall refer to conditional probabilities, expectations, and law of a random variable given $\hat{F}_{m+n}$, as $\Pr(\cdot \mid \hat{F}_{m+n})$, $E(\cdot \mid \hat{F}_{m+n})$, and $\mathcal{L}(\cdot \mid \hat{F}_{m+n})$, respectively. Formally, we have the following result.

Proposition 1. Under the null hypothesis, $\mathcal{L}\left(T \mid \hat{F}_{m+n}\right)$ does not depend on F_0.

Because $\mathcal{L}(T \mid \hat{F}_{m+n})$ is free from F_0 under the null hypothesis, at least in principle its distribution can be tabled. Hence, we are able to find $C_{\hat{F}}(\alpha)$ so that

$$\Pr(T \geq C_{\hat{F}}(\alpha) \mid \hat{F}_{m+n}) \leq \alpha, \quad \text{under } H_0, \tag{3}$$

where the equality can be approximately achieved provided that there are enough distinct values of T. In fact, conditioning on $\hat{F}_{m+n}$, $C_{\hat{F}}(\alpha)$ is constant and T is still random. A computational procedure (discussed later) can be used to find $C_{\hat{F}}(\alpha)$ such that (3) is valid. Since the construction of T only depends on the ranks, we expect that there would be C_{m+n}^{m} distinct values of T conditioning on $\hat{F}_{m+n}$ (e.g., when $m = n = 50$, $C_{m+n}^{m} = 1.0 \times 10^{29}$). As a result, $\Pr(T \geq C_{\hat{F}}(\alpha) \mid \hat{F}_{m+n})$ can be set *equal* to α with negligible discrepancy, although one need to know F_0 to quantify such discrepancy exactly.

Here, we want to emphasize that different from conventional tests that have constant cut-off values, $C_{\hat{F}}(\alpha)$ is a random variable depending on $\hat{F}_{m+n}$. To make this point clear, we use the subscript $\hat{F}$ to highlight the dependence, which is the key factor in constructing the distribution-free test. From the proof of Proposition 1, we know under H_0

$$\mathcal{L}\left(T \mid \hat{F}_{m+n}\right) = \mathcal{L}\left(T^* \mid \hat{F}_{m+n}\right),$$

where the notation with "*" represents the corresponding counterpart of T calculated using a random permutation of the pooled samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$. $C_{\hat{F}}(\alpha)$ can thus be approximated by a permutation procedure. The effectiveness of this test can be seen from the following result.

Proposition 2. Assume there exists $C_{\hat{F}}(\alpha)$ so that (3) with equality holds. Then

i. Under H_0, $\Pr(T \geq C_{\hat{F}}(\alpha)) = \alpha$.
ii. Under H_1, $\Pr(T \geq C_{\hat{F}}(\alpha)) \to 1$ as $m, n \to \infty$, if $\mu_{k0} \neq \mu_{k1}$ and $\Pr(X_k > Y_k) \neq 1/2$ for some k.

The first part of this proposition indicates the test is distribution-free under the null hypothesis. Even though the equality in (3) might not hold exactly for some $C_{\hat{F}}(\alpha)$, the practical influence is minimal as shown in Section 4. It is also noted that the test statistic T is not distribution-free as its distribution depends on F_0. However, the test *procedure* is unconditionally distribution-free in the sense that Part (i) can be made valid, regardless of the choice of F_0, by using data-dependent threshold $C_{\hat{F}}(\alpha)$. The second part establishes the consistency of the test, which is a basic requirement of any meaningful test procedure. In the context of MSPC, m or n is often small, but the test is still able to deliver good detection power comparing with Hotelling's T^2 and other multivariate nonparametric location tests on nonnormal distributions (see some evidence in Section 4 in the context of MSPC).

In the proposed test statistic, the information of $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$ is integrated from each individual marginally without taking the correlations between variables (accordingly T_j's) into account. Certainly, we could construct a Mahalanobis-distance-based test by incorporating the covariance matrix of T_j. However, when the sample size is small, as p increases, classical Mahalanobis distance may not work well because the contamination bias in estimating the covariance matrix grows rapidly with p. This drawback has been revealed by some recent study on high-dimensional tests (Bai and Saranadasa 1996; Feng et al. 2013). When p and n are comparable in certain sense, having the inverse

of the sample covariance matrix in constructing tests would be no longer beneficial. Note that the "small n, large p" situation is particularly common in our monitoring problem since we want to start the control chart with only limited reference samples. To this consideration, we focus on the test statistic T in the following because of its simplicity and effectiveness, especially in high-dimensional problems.

2.2 A Distribution-Free Multivariate Control Chart

This section discusses how to effectively sequentialize the distribution-free test proposed in Section 2.1 to construct a distribution-free control chart. For illustration, we focus on an EWMA construction due to its simplicity. Nevertheless, the idea presented in this section can be applied to many other appropriate charting schemes. Some alternatives are briefly discussed in the supplementary Section S2.

Recall the change point model (1) with individual observations. When the nth observation is collected, we can construct a distribution-free location test. Denote $\mathcal{X}_{k,j}^{n} = \{X_{jk}, \ldots, X_{jn}\}$. Following the proposal in Section 2.1, a charting statistic can be constructed as $T_n(w, \lambda) = \sum_{j=1}^{p} T_{jn}^2(w, \lambda)$, where

$$T_{jn}(w, \lambda) = \sum_{i=n-w+1}^{n} (1-\lambda)^{n-i} \frac{R_{jni} - w(m_0+n+1)/2}{\sqrt{w(m_0+n+1)(m_0+n-w)/12}},$$

where w is the window size, λ is the smoothing parameter, R_{jni} is the rank of X_{ji} among the sample $\mathcal{X}_{-m_0+1,j}^{n}$. $T_{jn}(w, \lambda)$ is essentially a weighted version of the two-sample Wilcoxon rank-sum statistic (2) for testing the equality of the locations of the sample $\mathcal{X}_{-m_0+1,j}^{n-w}$ and $\mathcal{X}_{n-w+1,j}^{n}$. Different rank observations in $T_{jn}(w, \lambda)$ are weighted as in an EWMA chart, that is, the more recent observations receive more weight, and the weight decays exponentially over time. Once there is a location change occurs in the jth component at (unknown) change point τ, $|T_{jn}(w, \lambda)|$ would become large once $n > \tau$ and accordingly a large value of T_n would trigger an alarm. Different from traditional use of EWMA, our $T_{jn}(w, \lambda)$ is a truncated one due to the use of w. In practice, we may choose an integer w so that $(1-\lambda)^w \approx 0$. Throughout this article, we set w as the smallest integer that satisfies $(1-\lambda)^w \le 0.05$. This adjustment hardly affects the performance of the proposed chart but can alleviate the computation burden to a large degree. See Proposition 3 and (6). As a side note, this test statistic automatically standardizes each variable due to the rank-based test so that the proposed method is invariant under scale transformation. In other word, we can use this charting scheme without making any preliminary standardization even if the scales of components differ much.

To construct a distribution-free control chart, it is critical to determine the control limits so that a charting scheme has satisfactory run-length behavior without making any distributional assumption. As recognized in the literature, it is often insufficient to summarize run-length behavior by ARL, especially for self-starting control charts (Hawkins and Maboudou-Tchao 2007; Zou, Wang, and Tsung 2012). The IC run-length distribution is usually considered to be satisfactory if it is close to the geometric distribution (Hawkins and Olwell 1998), which corresponds to the run-length distribution of a conventional Shewhart chart. Otherwise, a chart often attains the ARL_0 at a cost of elevated probability of false alarms with short runs, which would in turn hurt an operator's confidence in valid alarms.

To this end, we want to find the control limits so that the conditional probability that the charting statistic exceeds the control limit at present given that there is no alarm before the current time point is a prespecified constant. Although Proposition 2 reveals that the test with T_n is distribution free with data-dependent cut-off threshold, extending it to sequential testing seems not trivial. This is because the chart statistic T_n is correlated with previous values T_k, $k < n$. Hence the conditional false alarm probability is difficult to be controlled. To this end, we propose to determine the control limits $H_n(\alpha)$ by solving the following equations:

$$\begin{aligned} &\Pr(T_1(w, \lambda) > H_1(\alpha) \mid \hat{F}_1) = \alpha, \\ &\Pr(T_n(w, \lambda) > H_n(\alpha) \mid T_i(w, \lambda) < H_i(\alpha), \\ &1 \le i < n, \hat{F}_n) = \alpha \ \text{ for } n > 1, \end{aligned} \tag{4}$$

where α is the prespecified false alarm rate and $\hat{F}_n(\mathbf{t}) = (m_0+n)^{-1} \sum_{i=-m_0+1}^{n} I(\mathbf{X}_i \le \mathbf{t})$. Based on $H_n(\alpha)$, we can formally define the following charting procedure, termed as distribution-free multivariate EWMA chart (abbreviated as DFEWMA), with the run-length

$$\text{RL} = \min\{n;\ T_n(w, \lambda) \ge H_n(\alpha),\ n \ge 1\}. \tag{5}$$

Similar to the threshold $C_{\hat{F}}(\alpha)$ in two-sample location test, the control limits in our procedure are data-dependent, and are determined online along with the process observations (see Figure 2 for illustration of chart operations). That is, the control limits vary in each run even for monitoring the observations from the same F_0. At time n, the control limit $H_n(\alpha)$ is determined right after we observe the value of $\mathbf{X}_n$ (actually depending on $\hat{F}_n$ rather than the observations themselves). Consequently, this chart does not need any distributional assumption and thus does not suffer from wrong assumptions. This is the key success factor of designing a distribution-free multivariate control chart without any distributional assumption. Theorem 1 summarizes the distribution-free property of the proposed charting procedure as follows.

Theorem 1. Under the IC model, $\Pr(\text{RL} = n) = \alpha(1-\alpha)^{n-1}$ for any $n \ge 1$ and any continuous F_0.

According to Theorem 1, the IC run-length distribution of DFEWMA is exactly the Geometric distribution and the ARL_0 of the proposed chart is $1/\alpha$.

Remark 1. The proposed charting procedure employs data-dependent control limits, varying with index n and observations $\hat{F}_n$. Despite the similarities on the surface, they are *fundamentally different* from the approach of using dynamic control limits, which is originally proposed by Margavio et al. (1995), and subsequently formalized and used by, for example, Hawkins, Qiu, and Kang (2003), Zamba and Hawkins (2006), and Zou and Tsung (2010). In dynamic control limits, the sequence of control limits is determined once before monitoring and used in different runs. In other words, the sequence remain unchanged for the same IC distribution or distribution type. This feature

significantly distinguishes it from our data-dependent limits, which vary from one run to another even if the data follow the same IC distribution. Recently, Chatterjee and Qiu (2009) proposed a bootstrap method to determine the control limits for CUSUM charts. Compared with dynamic control limits as in, for example, Hawkins, Qiu, and Kang (2003), it does not require the knowledge of F_0. Instead, it can effectively determine the control limits through Markov chain theory and bootstrap sampling from m_0 in control samples. As a result, it can be more generally applied and valid for all F_0 as long as m_0 is sufficiently large. Nevertheless, the limits based on the bootstrap only depend on the m_0 IC samples. As long as the same IC samples are used for reference in Phase II monitoring, the same control limits are expected. Although the bootstrap procedure does not explicitly require to know F_0 to work well, the limits obtained might not be accurate enough to deliver exact IC run-length distribution when p is high and m_0 is small. Our approach is also different from adaptive control charts (Tsung and Wang 2010). Theoretical and numerical comparisons with some adaptive charts can be found in supplementary Section S4.

2.3 Algorithm and Computational Issues

The key to construct the distribution-free chart using the proposed procedure is to find the control limits $H_n(\alpha)$ satisfying (4). Due to the intricacy of the conditional probability (4), it seems impossible to solve $H_n(\alpha)$ analytically. In addition, it is infeasible to enumerate all possible values of $T_n(w,\lambda)$ to obtain $H_n(\alpha)$ though $T_n(w,\lambda)$ is (conditionally) distribution-free. For efficient implementation, we suggest an algorithm to approximate $H_n(\alpha)$. Before proceeding, we mention the following useful result.

Proposition 3. For each j, $T_{jn}(w,\lambda)$ is independent of $T_{jk}(w,\lambda)$, $k \le n-w$.

Using this proposition, we can simplify (4) as

$$\Pr(T_n(w,\lambda) > H_n(\alpha) \mid T_i(w,\lambda) < H_i(\alpha), \max\{1, n-w+1\} \le i < n, \hat{F}_n) = \alpha. \quad (6)$$

This formulation would reduce the computation in estimating $H_n(\alpha)$ for large n since the probability is conditioning on the event that the recent w T_i's do not exceed the control limits rather than all $n-1$ previous T_i's. Accordingly, we suggest the following algorithm.

i. For $n=1$, generate a random permutation of $\{-m_0+1,\ldots,1\}$, say $\{i_1,\ldots,i_{m_0+1}\}$. Obtain the corresponding $T_1^v(w,\lambda)$ based on permuted samples $\{\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_{m_0+1}}\}$. Repeat this procedure b times to obtain $T_1^1(w,\lambda),\ldots,T_1^b(w,\lambda)$, and find the threshold value, $\hat{H}_1(\alpha)$, by the $(1-\alpha)$ empirical quantile from samples $T_1^v(w,\lambda)$, $v=1,\ldots,b$.

ii. For $n>1$, similarly generate a random permutation and calculate the test statistics $T_k^v(w,\lambda)$, $\max\{1,n-w+1\} \le k < n$. If $T_k^v(w,\lambda) < H_k(\alpha)$, calculate $T_n^v(w,\lambda)$; otherwise, discard this permutation. Repeat this procedure until b samples $T_n^1(w,\lambda),\ldots,T_n^b(w,\lambda)$ are obtained. Find the limit $\hat{H}_n(\alpha)$ as the $(1-\alpha)$ empirical quantile from samples $T_n^v(w,\lambda)$, $v=1,\ldots,b$.

Note that this is a permutation test that is different from the usual bootstrap method in the sense that it is exactly distribution-free. As long as n is not too small so that we have enough distinct values of $T_n^*(w,\lambda)$ and b is sufficiently large, we can get the exact cut-off value of the conditional distribution. In computing each $H_n(\alpha)$, about $b = 5p/\alpha$ replications should be enough to obtain reliable approximations based on our experience. Even though some permutations are discarded in (ii), causing overhead in computation, the discarding probability is at most $1/(1-\alpha)^w - 1$. When $\alpha = 0.005$, $w = 28$ as in simulation settings, the discarding probability is 0.15.

Remark 2. Even though the control limit is estimated as the sample quantiles of b permuted samples, it has minimal influence on the practical performance. This is because even $H_n(\alpha)$ might not be exact, in most cases the decision on whether T_n is larger than $H_n(\alpha)$ is correct. Only when T_n falls in the tiny neighborhood of $H_n(\alpha)$, which has very small probability at the distribution tails, the decision might be incorrect due to inexact $\hat{H}_n(\alpha)$. As a result, the effect of the inexactness of the test level is negligible on the monitoring outcome. This effect is further reduced when larger b, larger m_0, or larger w is used.

Fast implementation is important and some computational issues deserve our careful examination. For the proposed chart, computing the charting statistic $T_n(w,\lambda)$ is trivial because at any time point n, it involves at most $p(n-1)$ ordering operations (i.e., comparing X_{jn} with $X_{j,-m_0+1},\ldots,X_{j,n-1}$ for $j=1,\ldots,p$). However, it is more computationally intensive to determine the $H_n(\alpha)$, which we analyze as follows. After a new observation $\mathbf{X}_n$ is observed, the ranks $\mathcal{R}_n = (\mathbf{R}_{-m_0+1},\ldots,\mathbf{R}_n)$ are updated and recorded, where $\mathbf{R}_i = (R_{1i},\ldots,R_{pi})^T$. In each permutation, $\mathcal{R}_n^*$ would be obtained directly from $\mathcal{R}_n$. Note that in calculating $T_k^*(w,\lambda)$ for $k = \max\{1,n-w+1\},\ldots,n-1$, the ranks R_{jki}^* are different from those in $\mathcal{R}_n^*$ because R_{jki}^* is the rank of X_{ji}^* in the sample $\mathcal{X}_{-m_0+1,j}^{k*}$ rather than in the sample $\mathcal{X}_{-m_0+1,j}^{n*}$. However, R_{jki}^* can be obtained with $O(w^2)$ computation and storage given $\mathcal{R}_n^*$ for all k,i. Also, the calculation of $T_k^*(w,\lambda)$ for $k = \max\{1,n-w+1\},\ldots,n-1$ can be performed in a recursive manner and only $O(w)$ computation is required in each update. In addition, generating a random permutation with size n requires $O(n)$ computation. Hence, at each time point, the total computational complexity is $O(bpw^2+bn)$. This complexity is only linear in n,b,p, and quadratic in w. In practice, the control limits often stabilize around a fixed value (see Figure 2 for examples). It is possible to use the constant stable limit for future monitoring samples to reduce the computation effort.

In comparison with any existing works, the DFEWMA chart requires a considerable amount of computing time because our control limits are obtained online. We would like to emphasize that such computation burden should not be considered as a "curse" but a "blessing." As we explained before, it seems impossible to construct a distribution-free multivariate chart (in a strict and general sense) with a traditional concept of SPC that the control limit is fixed and determined before monitoring. Today's computing power has improved dramatically and it is computationally feasible to implement the DFEWMA chart. For instance, when $b = 10{,}000$, $p = 30$, and $w = 28$, it takes

about 1 sec to finish the procedure at $n = 500$, using an Intel i7-2630 CPU and Matlab. With the help of parallel computing platform, the proposed procedure is promising for online monitoring of high-dimensional data streams. The computer codes implementing the proposed scheme in Fortran and Matlab are available from the authors upon request.

2.4 Practical Guidelines

On the m_0: Although the DFEWMA chart is a self-starting scheme in the sense that it can be implemented at the start-up of a process, we suggest that a practitioner should gather a modest number of observations through a Phase-I study and prior knowledge to obtain at least an initial verification that the process is actually stable before starting the formal DFEWMA chart. This is because a too small m_0 would result in a severe "masking-effect" if a short-run change occurs. Our empirical results show that to obtain a satisfactory monitoring performance it may require at least 50–100 IC observations (of course the more the better) before the change actually occurs; this number also depends on the dimensionality, the number of shifted components and the magnitudes of shifts.

On the start-up: When n is small, say $n < w$, some historical observations are included in the charting statistics. This may reduce the detection power when the shifts occur at the beginning. For early detection in such cases, we may set $w = \max\{5, \min\{w_1, n\}\}$, where w_1 is the window size satisfying $(1-\lambda)^{w_1} \approx 0$ as mentioned before. The constant 5 is suggested due to the following reason. The possible values of R_{jni} for very small n, say $n = 1$ or 2, are very limited when m_0 is small, which prevents the $H_n(\alpha)$ achieving a desired α if we set $w = \min\{w_1, n\}$. To resolve this problem, we suggest setting the window size to be at least 5 to produce sufficient distinct values of $T_n^*(w, \lambda)$. Based on our simulation study, using this adjustment is not only able to avoid "masking effect" in a short-run process but can also produce the desired IC run-length distributions. Similar treatment has been also suggested by Zou and Tsung (2010).

On choosing λ: Unlike the MEWMA and MSEWMA charts in which the choice of λ should be chosen to balance the robustness to nonnormality and the detection ability to various shift magnitudes (see Zou and Tsung 2011), the DFEWMA chart is distribution-free under IC with any weight. In general, the choice of λ is very similar to the choice of the smoothing parameter in traditional EWMA-type control charts. A smaller λ leads to a quicker detection of smaller shifts; See Section 4 for some simulation results. Typical choice of λ includes 0.05, 0.1, and 0.2. To obtain more robust protection against various shift sizes, an adaptive version of DFEWMA may be developed similar to the adaptive EWMA chart (Capizzi and Masarotto 2003). Also, a control chart based on change-point detection formulas is generally more robust against unknown shift magnitude (see Holland and Hawkins 2014) (see a brief discussion in supplementary Section S2). These studies are beyond the scope of this article but could be subjects of future research.

On post-signal diagnostic: In monitoring complex systems, apart from quick detection of abnormal changes of system performance and key parameters, a diagnostic aid to locate the change point in the process and to isolate responsible factors has become increasingly critical in a variety of applications. It often increases process throughout, decreases learning time, and reduces per unit production costs.

An estimate of the change point based on the nonparametric test $T_n(w, 0)$ is proposed to assist the diagnosis in our DFEWMA chart. We assume that the chart signals at kth observation, that is, there are m_0 historical IC observations and k future observations, and a shift occurred after the τth sample ($0 \le \tau < k$) as illustrated by model (1). Given an estimate of change-point v, we can calculate the test statistic $T_k(k - v, 0)$. Actually, the $T_k(k - v, 0)$ can intuitively be understood as a test statistic for testing if the sample $\{\mathbf{X}_{-m_0}, \ldots, \mathbf{X}_v\}$ has the same distribution as $\{\mathbf{X}_{v+1}, \ldots, \mathbf{X}_k\}$. Then, our suggested estimator of the change-point, τ, is given by

$$\widehat{\tau} = \underset{0 \le v < k}{\arg\max}\, T_k(k - v, 0). \tag{7}$$

After the shift location $\widehat{\tau}$ is estimated, we have partitioned the observations into two samples. The diagnosis framework proposed by Zou, Jiang, and Tsung (2011) can be readily applied.

3. REAL DATA APPLICATIONS

In this section, we demonstrate the proposed methodology using several real data examples. We focus on a real dataset from a semiconductor manufacturing process with *high*-dimensional process variables. Applications to well-studied data with lower dimensions, for example, data from aluminum electrolytic capacitor process (Zou and Tsung 2011), and data from aluminum smelter process (Qiu and Hawkins 2001), are included in supplementary Section S5.

The semiconductor manufacturing process under study is under constant surveillance via the monitoring of signals/variables collected from sensors at many measurement points. The dataset contains 1567 vector observations, and is publicly available in the UC Irvine Machine Learning Repository (*http://archive.ics.uci.edu/ml/datasets/SECOM*). The data were collected from July 2008 to October 2008 by a computerized system that automatically manages the process. Each observation consists of 591 continuous measurements from sensors. A categorical label, quality (± 1), indicating whether a single production is a conforming yield through house-line testing is also provided in this dataset. In this dataset, 1463 observations belong to the conforming group, and the remaining 104 observations are nonconforming. The goal of the data analysis is mainly to model and monitor production quality based on those sensor measurements without additional house-line testing.

It is desirable to set up an online detection system to automatically monitor the production process to guarantee its quality using MSPC. The sample correlation matrix of this data (not reported here) demonstrates that the variables have considerable interrelationships and suggests that MSPC is likely to be more effective. As a preprocessing, the missing values, which account for a negligible fraction of the dataset, were imputed by corresponding mean values. In addition, 117 variables with constant values in all observations are removed and 474 remaining data streams are used for further analysis (denoted by V1, V2, …, V474, respectively). Typically in such a high-dimensional monitoring situation, assuming normality is rather

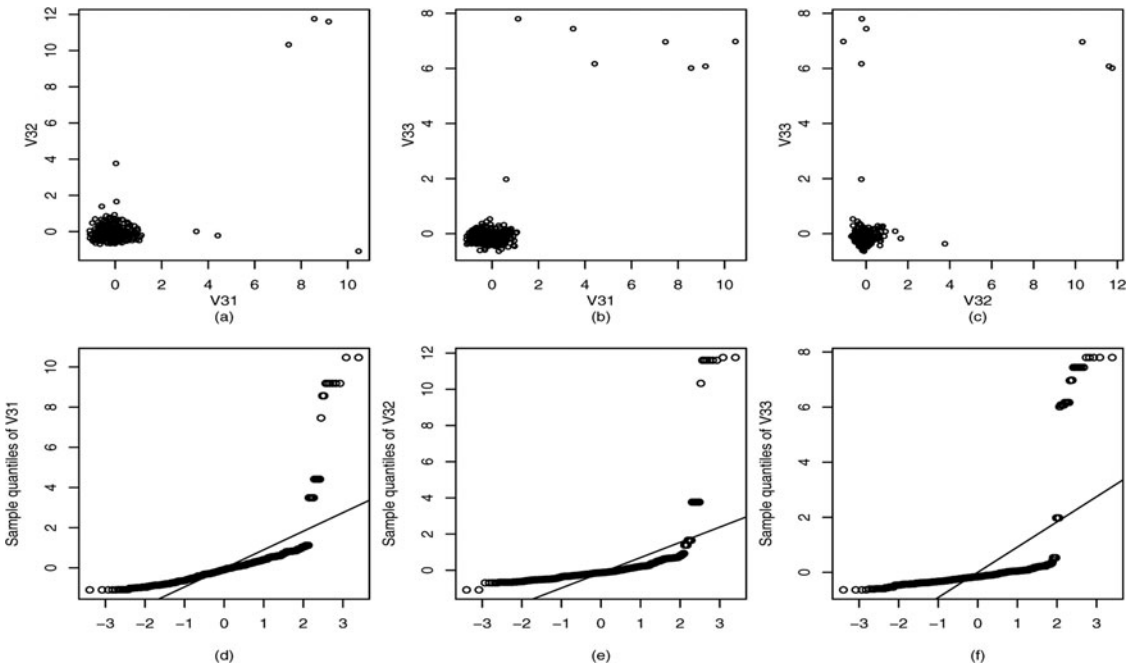

Figure 1. (a)–(c): The scatterplots of three stream observations; (d)–(f): the normal Q-Q plots for V31, V32, and V33, respectively.

risky. Figure 1(a)–1(c) shows the scatterplots of the raw data for the measurements of V31–V33, based on the in-control samples. The joint distribution of each pair of variables are far from bivariate normal. The normal Q-Q plots for these three distributions are shown in Figure 1(d)–1(f), which clearly indicate that the marginals are not normal either. There are also quite a few extreme observations in each stream. All these characteristics collectively suggest that the multivariate normality assumption is invalid and thus we could expect our distribution-free chart to be more robust and powerful than normal-based approaches for this dataset. To reduce the strong correlation among variables, we first randomly select 800 IC samples to perform the principal component analysis (PCA). The projection coefficients corresponding to the first 90 principal components, which roughly explained 80% of the variance, are obtained. We apply the projection to the remaining 663 IC samples and 104 OC samples, and use them to evaluate the monitoring performance.

To evaluate the IC performance, we use the $m_0 = 663$ IC (projected) samples as reference. A reference sample of this size is apparently *insufficient* to determine fully the IC distribution or its parameters since $p = 90$ is quite large. Therefore, self-starting control charts would be more desirable in this situation. We choose $\lambda = 0.05$ and in-control $\text{ARL}_0 = 200$. To get a rough picture of the IC performance, we use resampling method to obtain the IC run-length distribution. To be more specific, the Phase-II observations are independently drawn from 663 conforming observations with replacement. Then the DFEWMA chart is applied to those observations and the corresponding run-lengths are recorded. Repeat this procedure 1000 times and the IC ARL, run-length standard deviation (SDRL), and early false alarm rate $P(\text{RL} \leq 30)$ (FAR) of DFEWMA are 204, 226, 0.153, respectively, close to the nominal ones. This demonstrates DFEWMA has robust performance in monitoring high-dimensional nonnormal observations.

To demonstrate the OC performance of DFEWMA chart, we obtain the corresponding OC ARL by independently drawing from 104 nonconforming observations with replacement. In practice, it is rare to see the changes occur at the very beginning of the monitoring. Hence, we artificially set $m_0 = 600$ historical observations. During the Phase-II monitoring stage, the first 63 observations are also IC samples and the subsequent observations are sampled from 104 OC samples. In other words, we are evaluating the steady state ARL with $\tau = 63$. The resulting SSARL is 9.87, which reflects that the DFEWMA is able to quickly detect the changes in high-dimensional data streams. An example of the operation of DFEWMA control chart in one of the 1000 replications is presented in Figure 2, along with its data-dependent control limits.

4. PERFORMANCE ASSESSMENT

We present some simulation results in this section regarding the performance of the proposed DFEWMA chart and compare

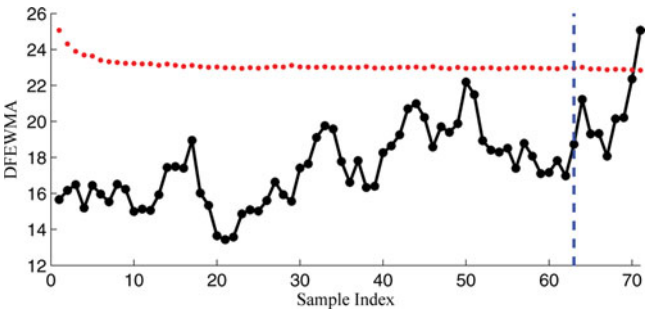

Figure 2. Operations of DFEWMA chart when changes occur at $\tau = 63$ (blue dashed line). The black dot represents the $T_n(w, \lambda)$ calculated at each step, and the red dot represents $H_n(\alpha)$.

Table 1. IC performance of the DFEWMA, SREWMA, SSEWMA, MSEWMA, MEWMA, and RTC charts with 10-dimensional multivariate normal observations

		$\lambda = 0.1$			$\lambda = 0.05$		
m_0	Method	ARL_0	SDRL	FAR	ARL_0	SDRL	FAR
	Geometric	200	200	0.140	200	200	0.140
50	DFEWMA	201	194	0.133	202	197	0.137
	SREWMA	201	221	0.151	201	199	0.093
	SSEWMA	202	186	0.102	198	181	0.062
	MSEWMA	56.3	50.7	0.381	53.6	42.3	0.322
	MEWMA	27.0	22.4	0.701	30.1	20.5	0.640
	RTC	8.61	3.04	0.948	—	—	—
100	DFEWMA	197	193	0.135	200	194	0.128
	SREWMA	200	211	0.120	197	193	0.074
	SSEWMA	199	188	0.101	199	176	0.063
	MSEWMA	87.0	78.5	0.224	78.7	63.5	0.182
	MEWMA	54.0	47.6	0.393	52.1	39.4	0.344
	RTC	27.7	30.8	0.546	—	—	—

it with some other procedures in the literature. All the results in this section are obtained from 10,000 replications unless indicated otherwise.

Comparing the DFEWMA procedure with alternative nonparametric methods turned out to be difficult due to the lack of an obviously comparable method. This is because most of the approaches in the literature were designed for the cases where sufficient historical observations are available to accurately estimate some parameters of or the entire IC distribution. See Zou and Tsung (2011) for some discussions and reviews. To our best knowledge, the SREWMA chart proposed by Zou, Wang, and Tsung (2012) is the only natural benchmark because it is a self-starting nonparametric scheme. We also consider the parametric self-starting multivariate EWMA (denoted as SSEWMA) control chart proposed by Hawkins and Maboudou-Tchao (2007), which was designed under the normality assumption. To demonstrate the usefulness of DFEWMA, MSEWMA chart (Zou and Tsung 2011), classical MEWMA chart, and the chart based real time contrast (RTC, Deng, Runger, and Tuv 2012) are also included for comparison.

Following the robustness analysis of Zou and Tsung (2011), we consider the following distributions: (i) multinormal; (ii) multivariate t with ζ degrees of freedom, denoted as $t_{p,\zeta}$; (iii) multivariate gamma with shape parameter ζ and scale parameter 1, denoted as $\text{Gam}_{p,\zeta}$. For each distribution, the covariance matrix $\mathbf{\Sigma}_0 = (\sigma_{ij})$ is chosen to be $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.5^{|i-j|}$, for $i, j = 1, 2, \ldots, p$. The shift of size δ in the first $[p/5]$ components is used as OC models, that is, $\mathbf{X}_i + \delta\mathbf{e}_1$ with $\mathbf{e}_1 = (\underbrace{1, \ldots, 1}_{1,\ldots,[p/5]}, 0, \ldots, 0)^T$. A low-dimensional case with $p = 10$ and a high-dimensional case with $p = 30$ are compared for each distribution case. For illustration purpose, we only present the results when the ARL_0 is set to 200. We have also conducted some other simulations with various shift types, different ARL_0's (370 and 500), and the results show that the general conclusions given below do not change. Due to space limit, some of the simulation results are in supplementary Section S6.2, together with results when other shift patterns and distributions are compared (supplementary Section S6.3).

4.1 In-Control Comparison

First, we study IC run-length distribution of the DFEWMA chart. The control chart performance is summarized using ARL_0, standard deviation of the run-length (SDRL) and FAR = $\Pr(RL \le 30)$, the false alarm rate during the first 30 observations, for each chart. Here the IC run-length distribution is considered to be satisfactory if it is close to the geometric distribution or more generally its variation is less than that of a geometric distribution. Note that when the run-length distribution is geometric, the SDRL is equal to ARL_0 and FAR is 0.140. We summarize the results of the control charts discussed before for 10-dimensional multivariate normal observations in Table 1, where cases with different combinations of λ and m_0 are compared. From this table, it can be seen that (i) the three self-starting charts, DFEWMA, SREWMA, and SSEWMA have satisfactory IC performance. Note that the control limits of both SREWMA and SSEWMA are obtained under the normality assumption and thus their appealing results in this table are not surprising; (ii) when the sample size of the IC dataset is small, the actual IC ARL, SDRL, and FAR of MSEWMA, MEWMA, and RTC are all quite far away from the nominal levels because their control limits are obtained under the assumption that m_0 is sufficiently large.

Next, we compare the results of the nonnormal distributions, $t_{p,5}$ and $\text{Gam}_{p,3}$ that are summarized in Table 2. As demonstrated in Table 1, the MSEWMA, MEWMA, and RTC perform unsatisfactorily when m_0 is small even for normal distributions, so we do not report their results in Table 2. We can see that the DFEWMA is still robust to the heavy-tailed and skewed distributions regardless of the values of m_0, p, and λ. Its IC ARLs and SDRLs are always (approximately due to simulation error) equal to 200. Of course, this can be understood because the IC run-length distribution of DFEWMA is exactly geometric as claimed in Theorem 1. In comparison, the SREWMA and SSEWMA usually have large biases in the IC ARL and the deviation becomes more pronounced as the dimension increases. Although the SREWMA performs noticeably better than SSEWMA, it is unable to achieve desired ARL_0 either because the SREWMA is not exactly distribution free.

Table 2. IC performance of the DFEWMA, SREWMA, and SSEWMA charts with multivariate $t_{p,5}$ and $\text{Gam}_{p,3}$ observations when $m_0 = 100$

		$\lambda = 0.1$			$\lambda = 0.05$		
p	Method	ARL_0	SDRL	FAR	ARL_0	SDRL	FAR
	Geometric	200	200	0.140	200	200	0.140
				Multivariate t_5 distribution			
10	DFEWMA	197	196	0.140	201	197	0.134
	SREWMA	169	176	0.142	188	183	0.081
	SSEWMA	81.8	74.4	0.272	124	110	0.130
30	DFEWMA	195	197	0.140	201	196	0.127
	SREWMA	129	169	0.271	165	187	0.122
	SSEWMA	50.3	41.2	0.400	87.7	68.5	0.163
				Multivariate Gam_3 distribution			
10	DFEWMA	199	201	0.137	201	205	0.144
	SREWMA	177	180	0.144	196	189	0.080
	SSEWMA	113	108	0.191	159	147	0.101
30	DFEWMA	198	189	0.137	196	199	0.152
	SREWMA	160	216	0.253	185	215	0.121
	SSEWMA	112	102	0.174	152	134	0.073

4.2 Out-of-Control Comparison

In this subsection, the DFEWMA and other considered charts are compared in terms of OC ARL. For an OC ARL comparison, we consider the steady-state ARL (SSARL). To evaluate the SSARL behavior of each chart, any replication in which a signal occurs before the $(\tau + 1)$th observation is discarded. $\tau = 25$ is fixed in all cases.

4.2.1 Comparisons of DFEWMA, MSEWMA, and MEWMA. Although the major benefit of DFEWMA is in self-starting situations when m_0 is small, it is also important to compare the DFEWMA with MSEWMA and MEWMA, which can provide better understanding of the performance of DFEWMA. To this end, we first compare three charts by assuming that m_0 is sufficiently large (equivalently, the IC parameters are known), in this case one hundred thousand. In such situations, the DFEWMA chart would essentially reduce to its theoretical counterpart (S.2) as described in supplementary Section S2. Figure 3 compares ARL of three charts with the normal, $t_{p,5}$, and $\text{Gam}_{p,3}$ observations for $p = 30$. Again, $\lambda = 0.05$ and 0.1 are considered. To give a relatively fair comparison, we consider the MSEWMA and MEWMA charts with the same value of λ as DFEWMA but their control limit are adjusted to make the IC ARLs equal to the nominal one. Note that such charting schemes with adjustments are only for comparison use in our simulations but not applicable in practical applications since the data distribution is usually unknown.

It can be readily seen that except for very large shifts, the three charts have similar performance for the normal distribution. The DFEWMA and MSEWMA charts have similar performance and both charts are more efficient in detecting the small and moderate shifts than MEWMA for the nonnormal distributions, while the MEWMA is more efficient in detecting large shifts, such as $\delta \geq 3.0$. This is understandable because the DFEWMA and MSEWMA charts, which are essentially based on ranks or signs rather than distances, share a similar drawback as those rank-based charts for univariate processes. That is, even though the shift is quite large, the ranks of the observations may not be able to grow as fast as the distance. In addition, as discussed by Zou and Tsung (2011), the MSEWMA chart is actually designed for the class of distributions with elliptical directions, which the $\text{Gam}_{p,\zeta}$ does not belongs to. Hence, it may not be as efficient as DFEWMA for this distribution. A similar Figure with $p = 10$ is included in supplementary Section S6.1, from which a similar conclusion can be reached.

Next, we compare these three charts when m_0 is not large in Table 3. We consider two combinations of (m, p), (100, 10) and (200, 30). Both the control limits of MSEWMA and MEWMA are again adjusted given m_0, p, and distributions. Besides the ARLs, the corresponding SDRLs are also included in this table to give a broader picture of the run-length distribution. The advantage of DFEWMA is quite clear when $p = 30$: it performs better than the other two charts for small or moderate shifts, and the difference is quite remarkable in many cases. These results indicate that even though the MSEWMA and MEWMA can attain the nominal ARL_0 by adjusting their control limits (may not be feasible in practice), their abilities to detect changes are largely compromised. In contrast, the DFEWMA chart is generally much more efficient when m_0/p is not large.

4.2.2 Comparisons of DFEWMA, SREWMA, SSEWMA, and RTC. Now, we turn to compare OC ARL values of DFEWMA, SREWMA, SSEWMA, and RTC. Again, for fair comparisons, we adjust the control limits of SREWMA, SSEWMA, and RTC as done in Section 4.2.1 to make the three charts have the same ARL_0s. Tables 4 and 5 summarize the comparison when $p = 10$ and 30, respectively. More results can be found in supplementary Section S6.2. We observe that the SSEWMA chart has superior efficiency for normally distributed data as we would expect, since the distribution (multinormal) is the one it assumes. The SREWMA, RTC, and DFEWMA charts also have quite satisfactory performance when δ is not too large and the difference among the three charts is small, even when p is large. In addition, we can see that the SREWMA and DFEWMA charts are much more efficient in detecting the small and moderate shifts than the SSEWMA chart when the distribution is nonnormal, especially when $p = 30$. As the

Table 3. OC ARL comparison of the DFEWMA, MSEWMA, and MEWMA charts when m_0 is not sufficiently large; numbers in parentheses are SDRL values

			$\lambda = 0.1$			$\lambda = 0.05$		
	(m_0, p)	δ	DFEWMA	MSEWMA	MEWMA	DFEWMA	MSEWMA	MEWMA
Normal	(100, 10)	0.5	69.3(90.9)	42.2(39.3)	39.7(39.2)	44.5(46.1)	37.1(28.0)	34.8(27.9)
		1.0	14.3(9.08)	13.5(6.69)	11.7(6.41)	14.4(6.67)	15.2(6.42)	13.3(6.25)
		2.0	6.84(2.11)	6.66(1.99)	4.84(1.76)	7.75(2.50)	8.32(2.48)	6.11(2.17)
		4.0	5.45(1.46)	4.82(1.15)	2.42(0.72)	6.53(1.88)	6.15(1.57)	3.12(0.95)
	(200, 30)	0.5	22.7(17.7)	33.7(25.8)	33.6(27.0)	22.4(13.0)	31.6(18.3)	30.4(18.6)
		1.0	8.47(3.49)	11.4(4.82)	10.5(4.73)	9.55(3.69)	13.4(4.76)	12.2(4.72)
		2.0	4.65(1.35)	5.70(1.53)	4.43(1.40)	5.71(1.68)	7.27(1.88)	5.70(1.72)
		4.0	3.91(1.00)	3.95(0.85)	2.25(0.61)	4.76(1.43)	5.06(1.13)	2.97(0.77)
t_5	(100, 10)	0.5	101(137)	46.0(43.4)	97.6(116)	64.8(85.2)	40.9(32.3)	63.2(69.7)
		1.0	18.2(14.0)	14.8(8.17)	28.4(27.7)	17.3(9.02)	16.4(7.51)	21.4(12.6)
		2.0	7.73(2.64)	7.23(2.42)	8.33(3.50)	9.04(2.92)	8.96(2.91)	8.81(3.37)
		4.0	5.80(1.53)	5.07(1.29)	3.63(1.09)	6.80(1.99)	6.46(1.72)	4.25(1.32)
	(200, 30)	0.5	37.7(45.5)	37.7(30.7)	119(125)	28.0(20.2)	33.7(20.6)	64.9(55.1)
		1.0	10.4(4.51)	12.6(5.74)	36.1(29.6)	11.3(4.64)	14.4(5.53)	21.8(10.3)
		2.0	5.36(1.63)	6.19(1.90)	9.26(3.21)	6.26(1.97)	7.82(2.21)	8.96(2.74)
		4.0	4.14(1.07)	4.20(0.98)	3.89(0.98)	5.05(1.40)	5.37(1.27)	4.35(1.12)
Gam_3	(100, 10)	0.5	56.8(90.8)	49.9(51.5)	71.1(91.3)	36.5(55.7)	59.2(56.6)	76.2(98.6)
		1.0	12.1(5.49)	14.6(8.54)	21.5(21.1)	12.7(5.33)	22.1(12.5)	26.7(21.4)
		2.0	6.39(1.90)	6.90(2.15)	6.95(2.96)	7.51(2.33)	10.5(3.69)	10.5(4.82)
		4.0	4.94(1.35)	4.94(1.22)	3.19(1.04)	6.22(1.79)	6.93(1.89)	4.91(1.76)
	(200, 30)	0.5	19.0(13.3)	33.0(25.9)	42.1(39.0)	18.1(8.24)	30.1(17.2)	33.4(23.4)
		1.0	7.40(2.48)	11.2(4.77)	12.7(6.80)	8.57(2.70)	13.0(4.59)	13.6(5.70)
		2.0	4.43(1.16)	5.60(1.51)	5.10(1.70)	5.39(1.58)	7.06(1.86)	6.20(1.97)
		4.0	3.52(0.95)	3.91(0.86)	2.52(0.72)	4.42(1.23)	4.99(1.13)	3.18(0.87)

effectiveness of spatial-rank-based method relies more or less on certain symmetric assumptions (Oja 2010), the advantage of DFEWMA over SREWMA is more prominent for $\text{Gam}_{p,\zeta}$. We can also note that RTC performs reasonably well, especially in detecting large shifts. However, DFEWMA in general is better than RTC in detecting small shifts when $\lambda = 0.05$.

We conducted other simulations with various correlation structures, p and ARL_0. These simulation results (partially reported in supplementary Section S6.3) show that the DFEWMA chart works well in terms of its IC and OC ARLs in other cases as well. Of course, it should be emphasized that the OC ARL comparison depends on the correlation structure and shift

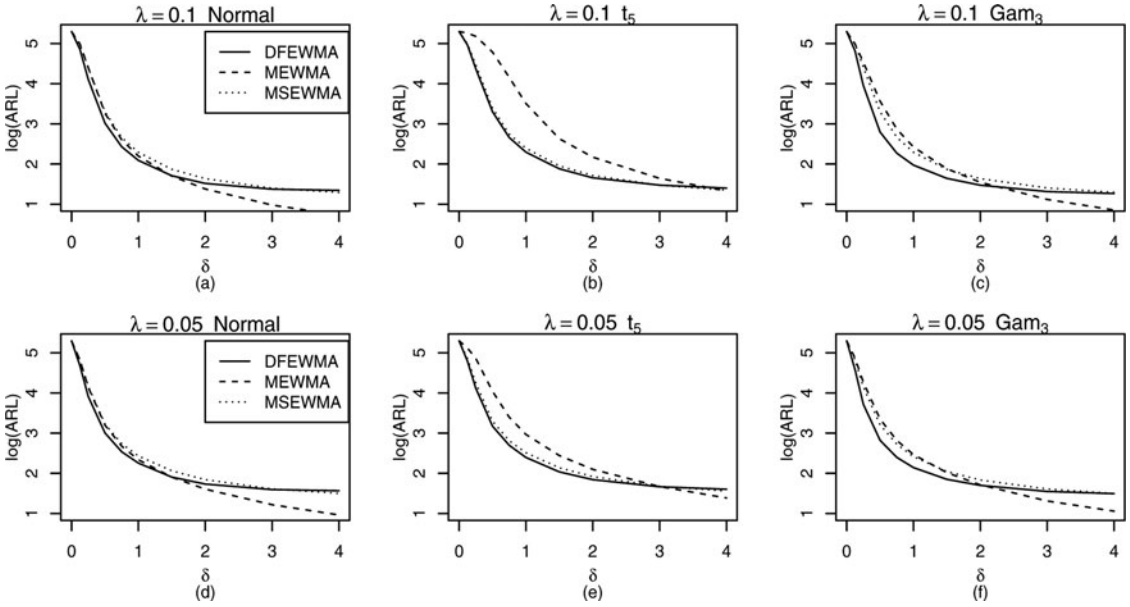

Figure 3. OC ARL comparison of the DFEWMA, MSEWMA, and MEWMA charts using $\lambda = 0.05$ and 0.1 when $p = 30$ and $m_0 = 100{,}000$.

Table 4. OC ARL comparison of the DFEWMA, SREWMA, SSEWMA, and RTC charts with $m_0 = 50$ and $p = 10$; numbers in parentheses are SDRL values

		$\lambda = 0.1$			$\lambda = 0.05$			
	δ	DFEWMA	SREWMA	SSEWMA	DFEWMA	SREWMA	SSEWMA	RTC
Normal	0.5	86.4(115)	76.0(127)	72.8(115)	60.4(79.3)	52.8(79.7)	50.9(68.8)	71.2(90.1)
	1.0	16.1(16.6)	12.6(8.19)	13.4(10.7)	16.0(8.68)	13.7(7.10)	14.0(7.35)	15.7(13.3)
	2.0	6.97(2.08)	5.51(1.98)	5.00(1.93)	8.23(2.55)	6.80(2.49)	6.20(2.32)	6.61(1.89)
	4.0	5.56(1.44)	3.61(1.05)	2.58(0.82)	6.75(1.93)	4.61(1.48)	3.36(1.14)	5.07(1.22)
t_5	0.5	117(148)	89.5 (142)	140(158)	88.4(111)	64.7(100)	101(131)	86.7(116)
	1.0	23.8(30.2)	16.8 (20.5)	54.6(96.4)	19.7(13.0)	16.2(13.1)	24.0(28.3)	19.6(18.9)
	2.0	8.25(2.79)	6.25 (2.60)	8.49(4.15)	9.27(3.24)	7.51(2.98)	8.73(3.61)	7.29(2.26)
	4.0	5.83(1.61)	3.85 (1.21)	3.75(1.18)	7.04(2.10)	4.86(1.63)	4.37(1.47)	5.49(1.47)
Gam_3	0.5	78.2(122)	68.3(99.7)	78.1(114)	49.7(75.2)	90.9(134)	107(140)	77.8(107)
	1.0	13.6(9.40)	17.1(11.7)	20.9(18.9)	13.9(6.56)	18.9(25.4)	32.0(55.3)	13.3(12.0)
	2.0	6.61(1.98)	7.69(2.88)	8.04(3.44)	7.99(2.53)	6.40(2.42)	7.39(3.67)	6.67(1.83)
	4.0	5.13(1.40)	4.94(1.55)	4.08(1.43)	6.27(1.86)	3.94(1.14)	3.40(1.13)	5.32(1.34)

Table 5. OC ARL comparison of the DFEWMA, SREWMA, SSEWMA, and RTC charts with $m_0 = 100$ and $p = 30$; numbers in parentheses are SDRL values

		$\lambda = 0.1$			$\lambda = 0.05$			
	δ	DFEWMA	SREWMA	SSEWMA	DFEWMA	SREWMA	SSEWMA	RTC
Normal	0.5	28.3(32.9)	44.7(99.6)	50.5(74.6)	24.3(15.3)	28.7(35.1)	35.8(36.9)	33.1(32.3)
	1.0	8.53(3.65)	9.40(4.80)	10.7(5.45)	10.0(3.69)	10.9(4.89)	12.0(4.98)	8.96(3.96)
	2.0	4.79(1.36)	4.62(1.62)	4.50(1.51)	5.80(1.69)	5.84(2.04)	5.78(1.87)	5.29(1.36)
	4.0	3.98(1.04)	3.02(0.90)	2.42(0.71)	4.84(1.35)	3.88(1.24)	3.18(0.94)	4.25(1.01)
t_5	0.5	47.6(71.8)	71.3 (138)	140(143)	31.9(36.6)	42.2(59.8)	81.1(97.4)	51.6(66.0)
	1.0	11.1(5.38)	12.6 (8.59)	51.7(75.3)	11.8(4.62)	13.4(6.52)	20.2(12.3)	11.0(6.72)
	2.0	5.40(1.66)	5.45 (1.94)	8.80(4.64)	6.48(2.10)	6.70(2.37)	8.39(2.80)	6.10(1.66)
	4.0	4.15(1.15)	3.36 (0.99)	3.87(1.03)	5.06(1.46)	4.26(1.35)	4.32(1.24)	4.79(1.25)
Gam_3	0.5	21.9(21.2)	49.2(109)	63.0(86.9)	19.4(9.88)	31.9(36.9)	41.7(49.3)	33.0(34.0)
	1.0	7.55(2.62)	10.4(5.92)	13.7(10.6)	9.06(3.06)	11.6(5.27)	13.5(6.07)	7.72(2.05)
	2.0	4.49(1.23)	4.91(1.66)	5.18(1.83)	5.57(1.61)	6.08(2.08)	6.23(2.11)	5.39(1.33)
	4.0	3.65(0.92)	3.13(0.88)	2.70(0.80)	4.46(1.22)	4.02(1.22)	3.40(1.05)	4.25(1.01)

directions. However, the number and variety of covariance matrices and shift directions are too large to allow a comprehensive, all-encompassing comparison. Our goal is to show the effectiveness, robustness, and sensitivity of the DFEWMA chart, and thus we choose a few representative and commonly used models for illustration. Overall, the effectiveness of DFEWMA is two-fold: the DFEWMA can always achieve the desired IC run-length distribution; with the same real ARL_0, the DFEWMA is quite sensitive to small to medium process shifts in nonnormal observations. Therefore, although requiring more computational effort, the DFEWMA is a reasonable alternative for nonmultinormal processes if we take its efficiency, convenience, and robustness into account.

5. CONCLUDING REMARKS

As indicated in Section 1, there lacks a general methodology to monitor multivariate data streams robustly for all data distributions. Existing MSPC methods often need to know the IC distribution or require a sufficiently large reference sample before monitoring. Since we rarely have such foreknowledge in real applications and an inappropriate assumption or estimation may lead to poor run-length performance, it is difficult to apply existing methods in many situations. To this end, we propose a multivariate distribution-free control scheme, DFEWMA. This self-starting scheme integrates a powerful distribution-free test with EWMA process monitoring. We suggest to use data-dependent control limits, which are determined based on the observations online. Compared with both parametric and nonparametric schemes, DFEWMA not only has much more robust IC performance (can always achieve the desired IC run-length distribution), but also can detect small and moderate shifts in location parameters for skewed or heavy-tailed multivariate distributions effectively.

In the future research, there are some valuable directions discussed as follows. First of all, the idea of using data-dependent control limits can also be extended to other cases as well. For example, increasing attentions have been paid to the problem of monitoring the occurrence rate of an adverse event with time-varying sample sizes in Phase II analysis (see Zhou et al. 2012). In this problem, the observation at time point t is the count of an event, which is usually assumed to be Poisson distributed with the occurrence rate θn_t conditional on sample size n_t. Traditional control charts require perfect knowledge of sample sizes to determine the control limits before monitoring. We expect this impractical requirement to be resolved to certain degree by

using a sequence of online determined control limits. Moreover, it requires more research to extend our method to Phase I analysis, in which detection of outliers or change-points in a historical dataset and estimation of the baseline distribution would be of great interest. In addition, it is also interesting to consider design parameters, such as ω, λ, to be data-dependent to further improve the OC performance in detecting unknown shifts, as in other adaptive SPC works. Finally, the current version of the proposed scheme is designed to detect changes occurred in marginal distributions only. We believe that, after certain modifications, the proposed method should be able to monitor changes in the covariance structure by using bivariate empirical distributions.

SUPPLEMENTARY MATERIALS

Technical details: The PDF file provides the technical details as referred in the article. It also includes additional figures and tables from simulation studies (PDF file).

Source code: The zipped package contains Matlab codes for simulation studies, and data for case study (ZIP file).

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. Chen was partially supported by Singapore AcRF Tier 1 funding #R-266-000-078-112. Zi and Zou were supported by the NNSF of China Grants 11622104, 11431006, 11131002, 11371202, 11271205 and Foundation for the Author of National Excellent Doctoral Dissertation of PR China 201232. Zou is the corresponding author.

[Received September 2013. Revised May 2015.]

REFERENCES

Bai, Z., and Saranadasa, H. (1996), "Effect of High Dimension: By an Example of a Two Sample Problem," *Statistica Sinica*, 6, 311–329. [450]

Bickel, P. J. (1969), "A Distribution Free Version of the Smirnov Two Sample Test in the P-Variate Case," *The Annals of Mathematical Statistics*, 40, 1–23. [450]

Boone, J. M., and Chakraborti, S. (2012), "Two Simple Shewhart-Type Multivariate Nonparametric Control Charts," *Applied Stochastic Models in Business and Industry*, 28, 130–140. [449]

Capizzi, G., and Masarotto, G. (2003), "An Adaptive Exponentially Weighted Moving Average Control Chart," *Technometrics*, 45, 199–207. [453]

Chakraborti, S., Van Der Laan, P., and Bakir, S. T. (2001), "Nonparametric Control Charts: An Overview and Some Results," *Journal of Quality Technology*, 33, 304–315. [450]

Champ, C. W., Jones-Farmer, L. A., and Rigdon, S. E. (2005), "Properties of the T2 Control Chart When Parameters Are Estimated," *Technometrics*, 47, 437–445. [448]

Chatterjee, S., and Qiu, P. (2009), "Distribution Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits," *Annals of Applied Statistics*, 3, 349–369. [452]

Crosier, R. B. (1988), "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes," *Technometrics*, 30, 291–303. [448]

Deng, H., Runger, G. C., and Tuv, E. (2012), "System Monitoring With Real-Time Contrasts," *Journal of Quality Technology*, 44, 9–27. [455]

Feng, L., Zou, C., Wang, Z., and Chen, B. (2013), "Rank-Based Score Tests for High-Dimensional Regression Coefficients," *Electronic Journal of Statistics*, 7, 2131–2349. [450]

Hawkins, D. M. (1991), "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, 33, 61–75. [448]

Hawkins, D. M., and Maboudou-Tchao, E. M. (2007), "Self-Starting Multivariate Exponentially Weighted Moving Average Control Charting," *Technometrics*, 49, 199–209. [448,451,455]

Hawkins, D. M., and Olwell, D. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, Berlin: Springer Verlag. [451]

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003), "The Changepoint Model for Statistical Process Control," *Journal of Quality Technology*, 35, 355–366. [451]

Holland, M. D., and Hawkins, D. M. (2014), "A Control Chart Based on a Nonparametric Multivariate Change-Point Model," *Journal of Quality Technology*, 46, 63–77. [449,453]

Liu, R. Y. (1995), "Control Charts for Multivariate Processes," *Journal of the American Statistical Association*, 90, 1380–1387. [449]

Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992), "A Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, 34, 46–53. [448]

Margavio, T. M., Conerly, M. D., Woodall, W. H., and Drake, L. G. (1995), "Alarm Rates for Quality Control Charts," *Statistics & Probability Letters*, 24, 219–224. [451]

Oja, H. (2010), *Multivariate Nonparametric Methods with R*, New York: Springer. [450,457]

Pignatiello, J. J., and Runger, G. C. (1990), "Comparisons of Multivariate Cusum Charts," *Journal of Quality Technology*, 22, 173–186. [448]

Qiu, P. (2008), "Distribution-Free Multivariate Process Control Based on Log-Linear Modeling," *IIE Transactions*, 40, 664–677. [449]

——— (2014), *Introduction to Statistical Process Control*, Boca Raton, FL: CRC Press. [448]

Qiu, P., and Hawkins, D. (2001), "A Rank-Based Multivariate CUSUM Procedure," *Technometrics*, 43, 120–132. [449,453]

Quesenberry, C. P. (1997), *SPC Methods for Quality Improvement*, New York: Wiley. [448]

Ross, G. J., Tasoulis, D. K., and Adams, N. M. (2011), "Nonparametric Monitoring of Data Streams for Changes in Location and Scale," *Technometrics*, 53, 379–389. [450]

Stoumbos, Z. G., Marion, R. Reynolds, J., Ryan, T. P., and Woodall, W. H. (2000), "The State of Statistical Process Control as We Proceed into the 21st Century," *Journal of the American Statistical Association*, 95, 992–998. [448]

Stoumbos, Z. G., and Sullivan, J. H. (2002), "Robustness to Non-Normality of the Multivariate EWMA Control Chart," *Journal of Quality Technology*, 34, 260–276. [449]

Sullivan, J. H., and Jones, L. A. (2002), "A Self-Starting Control Chart for Multivariate Individual Observations," *Technometrics*, 44, 24–33. [448]

Sun, R., and Tsung, F. (2003), "A Kernel-Distance-Based Multivariate Control Chart Using Support Vector Methods," *International Journal of Production Research*, 41, 2975–2989. [449]

Tsung, F., and Wang, K. (2010), "Adaptive Charting Techniques: Literature Review and Extensions," in *Frontiers in Statistical Quality Control 9*, eds. Lenz, H. J., Wilrich, P.-Th., and Schmid, W., Berlin: Physica-Verlag, pp. 19–36. [452]

Woodall, W. H., and Montgomery, D. (2013), "Some Current Directions in the Theory and Application of Statistical Process Monitoring," *Journal of Quality Technology*, 46, 79–94. [448]

Zamba, K. D., and Hawkins, D. M. (2006), "A Multivariate Change-Point Model for Statistical Process Control," *Technometrics*, 48, 539–549. [448,451]

Zantek, P. F. (2006), "Design of Cumulative Sum Schemes for Start-up Processes and Short Runs," *Journal of Quality Technology*, 38, 365–375. [448]

Zhou, C., Zou, C., Zhang, Y., and Wang, Z. (2009), "Nonparametric Control Chart Based on Change-Point Model," *Statistical Papers*, 50, 13–28. [449]

Zhou, Q., Zou, C., Wang, Z., and Jiang, W. (2012), "Likelihood-Based EWMA Charts for Monitoring Poisson Count Data with Time-Varying Sample Sizes," *Journal of the American Statistical Association*, 107, 1049–1062. [458]

Zou, C., Jiang, W., and Tsung, F. (2011), "A Lasso-Based Diagnostic Framework for Multivariate Statistical Process Control," *Technometrics*, 53, 297–309. [453]

Zou, C., and Qiu, P. (2009), "Multivariate Statistical Process Control Using Lasso," *Journal of the American Statistical Association*, 104, 1586–1596. [448]

Zou, C., and Tsung, F. (2010), "Likelihood Ratio-Based Distribution-Free EWMA Control Charts," *Journal of Quality Technology*, 42, 174–196. [449,451,453]

——— (2011), "A Multivariate Sign EWMA Control Chart," *Technometrics*, 53, 84–97. [449,450,453,455,456]

Zou, C., Wang, Z., and Tsung, F. (2012), "A Spatial Rank-Based Multivariate EWMA Control Chart," *Naval Research Logistics*, 59, 91–110. [449,450,451,455]