Journal of Quality Technology
A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: https://www.tandfonline.com/loi/ujqt20

Robust Multivariate Control Chart Based on Goodness-of-Fit Test

Chen Zhang, Nan Chen & Changliang Zou

To cite this article: Chen Zhang, Nan Chen & Changliang Zou (2016) Robust Multivariate Control Chart Based on Goodness-of-Fit Test, Journal of Quality Technology, 48:2, 139-161, DOI: 10.1080/00224065.2016.11918156

To link to this article: https://doi.org/10.1080/00224065.2016.11918156

Robust Multivariate Control Chart Based on Goodness-of-Fit Test

CHEN ZHANG and NAN CHEN

Department of Industrial and Systems Engineering, National University of Singapore

CHANGLIANG ZOU

Institute of Statistics and LPMC, Nankai University

This paper proposes a distribution-free multivariate statistical process control (MSPC) chart to detect general distributional changes in multivariate process variables. The chart is deployed based on a multivariate goodness-of-fit test, which is extendible to high-dimensional observations. The chart also employs *data-dependent* control limits, which are computed on line along with the charting statistics, to ensure satisfactory and robust charting performance of the proposed method. Through theoretical and numerical analyses, we have shown that the proposed chart is exactly *distribution-free* and able to operate with an unknown in-control (IC) distribution or limited reference samples. The chart also has robust IC performance as well as satisfactory out-of-control (OC) detection power for general process changes without any assumption of the process distribution. A real-data example in semiconductor production processes is presented to demonstrate the application and effectiveness of our method.

Key Words: Distribution Free; Empirical Distribution; Multivariate Statistical Process Control; Nonparametric; Self-Starting.

1. Introduction

Modern manufacturing processes usually involve several related quality variables and demand effective multivariate statistical process control (MSPC) to improve their competitive advantages. MSPC originates from the Hotelling's T^2 chart, which monitors the mean vector of multiple process variables following the multivariate normal distribution. Since then, MSPC charts have been shown to be more effective in monitoring correlated process variables than multiple univariate SPC charts and, hence, have attracted significant attention. Subsequent developments include multivariate cumulative sum (CUSUM) charts (Crosier (1988), Healy (1987), Pignatiello (1990)) and multivariate exponentially weighted moving averag (EWMA) charts (Lowry et al. (1992), Runger and Prabhu (1996)) to improve the charting performance of detecting small mean shifts.

While many MSPC charts are designed or perform best to detect mean shifts of quality variables in a process, it has been well acknowledged that "changes in the process mean are occasionally accompanied with or might be masked by an unsuspected change in the process variability" (Zamba and Hawkins (2009)). As a consequence, several charts have been proposed to monitor the process variance or the covariance matrix, including the generalized variance method (Montgomery and Wadsworth (1972)), the generalized likelihood ratio (GLR) method (Alt (1985), Hawkins and Maboudou-Tchao (2008)), and the penalized likelihood method (Yeh et al. (2004)).

Recently, charts that can monitor both the mean vector and the covariance matrix simultaneously have received increasing attention in the literature. They can be classified into two categories. The first

Ms. Zhang is a doctoral student in the Department of Industrial and Systems Engineering. Her email address is zhang chen@u.nus.edu.

Dr. Chen is an Assistant Professor in the Department of Industrial and Systems Engineering. His email is isecn@nus.edu .sg. He is the corresponding author.

Dr. Zou is a Professor in the Institute of Statistics and LPMC. His email is chlzou@yahoo.com.cn.

category uses two separate charting statistics to detect shifts in the process mean vector and the covariance matrix, respectively, and combines them into a single statistic by appropriate transformations. Examples in this category include Chen et al. (2005), Yeh et al. (2005), Yeh and Lin (2002), Khoo (2004), Reynolds and Cho (2006), and Maboudou-Tchao and Hawkins (2011). The second category constructs a single charting statistic directly that can respond to both mean shifts and covariance matrix shifts effectively. For instance, Hawkins (1991) proposed a CUSUM chart based on regression-adjusted variables; Zamba and Hawkins (2009) formulated a change-point detection model based on the GLR test; Zhang et al. (2010) employed the EWMA strategy on the GLR statistic to achieve fast responses to different types of shifts.

Despite their significance in MSPC, the aforementioned methods often need to assume that process variables follow the multivariate normal distribution or some other *known* distributions, at least when the process is in control. Unfortunately, these distributional assumptions are frequently violated in practice, especially when the dimension p is large, because many data exhibit features (e.g., heavy tails or skewness) that are distinct from conveniently assumed distributions. More important, many charts fail to perform well when the distributional assumptions are not satisfied (Qiu 2008; Woodall 2000). To address this problem, control charts based on nonparametric statistics can be useful. These nonparametric charts have robust performance in different distributions in the presence of outliners. Qiu and Hawkins (2008) developed a CUSUM chart based on the anti-ranks among different dimensions within each sample. Qiu (2008) proposed a CUSUM chart based on a log-linear model. Zou and Tsung (2011) introduced the concept of spatial signs to MSPC, and obtained a distributionally robust multivariate sign EWMA chart. Holland and Hawkins (2014) proposed a quarantined change-point model based on directional rank test statistics. Liu (1995) and Liu et al. (2004) proposed several charts based on data depth. Recently, there are also several developments that formulate MSPC as a classification problem to determine whether the observed samples belong to the in-control (IC) "class" or out-of-control (OC) "class". They use different classification methods, which do not rely on any distributional assumptions as well (see Sun and Tsung (2003), He and Wang (2007), Sukchotrat et al. (2009), Hwang et al. (2007), Deng et al. (2012), for examples).

Even though these nonparametric charts perform robustly in detecting changes across different types of data distributions, they are not *distribution free*. By distribution free, we mean without knowing the exact IC distribution or requiring a sufficiently large size of IC samples, a chart can attain the specified IC run-length distribution or at least IC average run length (ARL_0). Unfortunately, unlike many univariate nonparametric charts (see Zou and Tsung (2010), Chakraborti et al. (2001), for examples and reviews), which are both robust and distribution free, it is challenging to design a *distribution-free* MSPC scheme based on these conventional constructions. For example, the multivariate sign EWMA chart of Zou and Tsung (2011) requires at least 4,000 IC samples to attain a specified ARL_0 when monitoring five-dimensional data with an unknown distribution. Not surprisingly, it requires increasingly more IC samples as the dimension increases. The similar problem exists in other multivariate nonparametric charts (Zou et al. (2012)). However, in many industrial applications, knowledge of the IC distribution or a large group of IC samples is infeasible or challenging to obtain. In these cases, the inaccurate estimation of the IC distribution (or its parameters) from limited samples can significantly compromise the charting performance (Jones et al. (2001)).

To (partially) mitigate the effects of limited IC samples, especially at the start-up stage, various self-starting control charts have been proposed. In essence, self-starting charts sequentially monitor the process and update the IC distribution parameters using the newly observed samples if they are deemed to be in control. They also adjust the control limits such that the conditional false-alarm rate meets the specification even at the early stage with limited samples. For instance, Hawkins and Maboudou-Tchao (2007) and Maboudou-Tchao and Hawkins (2011) proposed two EWMA charts with self-starting features to monitor the mean vector and the covariance matrix. Zamba and Hawkins (2009) used a change-point model for MSPC. However, these three methods perform satisfactorily only when the data follow the multivariate normal distribution. Later, Zou et al. (2012) proposed a self-starting nonparametric approach based on spatial ranks. It has robust IC performance and satisfactory OC performance. It even enjoys the *distribution-free* property when the data come from the elliptical distribution family. However, its performance for other distributions is not guaranteed. See also Holland and Hawkins (2014) for a related approach based on change-point models. To the

best of our knowledge, there lacks an MSPC chart that has satisfactory performance regardless of the data distribution type and dimension, even with limited IC samples.

To fill in the research gap, this paper aims to develop a nonparamemtric MSPC chart with both objectives in mind. First, the chart should be able to detect general changes in the multivariate distribution effectively, including changes in the mean vector, the covariance matrix, and the distribution shapes. Second, the chart should be *distribution free* with robust performance even with limited IC (reference) samples. There are several challenges associated with both objectives. To detect general distributional changes, it is insufficient to monitor each marginal distribution separately and combine the information together. Monitoring marginal distributions only fails to detect the changes in the correlation structure. On the other hand, when the dimension p is large, it is prohibiting to estimate the distribution or even its covariance matrix well due to the curse of dimensionality and contamination noise (Feng et al. (2013)). Moreover, existing charts are not able to ensure the distribution-free property under general MSPC settings.

In this paper, we propose a new MSPC chart based on a goodness-of-fit (GoF) test. Instead of monitoring marginal distributions separately or monitoring the entire joint distribution directly, we construct the charting statistic through a series of bivariate GoF tests. Each test is designed to detect general distributional changes in the bivariate distribution of a selected pair of process variables. Then, by integrating the change information of a collection of bivariate distributions, the proposed chart is able to detect a much broader category of changes and is computationally efficient. For different distributional changes, the optimal pair selection mechanisms are different. We also provide some guidelines on how to select the pairs for different purposes. The bottom line is that, through the standard pair selection, the chart is omnipotent for general distributional changes. In addition, to achieve robust IC performance with unknown distributional information or limited IC reference samples, we develop a novel *data-dependent* control limit scheme. This scheme, along with each monitoring statistic, determines the control limit at each step on line. Leveraging on the permutation principle, the scheme can deliver *distribution-free* IC performance, meaning that the IC run-length distribution can meet the specification regardless of the IC distribution. We also design many numerical studies to demonstrate the satisfactory performance of our proposed chart.

The remainder of this paper is organized as follows: Section 2 introduces the multivariate goodness-of-fit test for MSPC. Section 3 presents the distribution-free chart based on the GoF test; Section 4 evaluates the charting performance and compares it with some alternative approaches; Section 5 provides some practical guidelines on the optimal pair selection for different distributional changes. Section 6 demonstrates the proposal using a real-data example from a semiconductor production process; Finally Section 7 concludes this paper with remarks. Some technical details are provided in Appendix.

2. Goodness-of-Fit Test

In this section, we first review the univariate GoF test. Subsequently, we propose a new GoF test for multivariate distributions.

2.1. A Powerful Univariate Goodness-of-Fit Test

Let $\mathcal{S}_n = \{X_1, \ldots, X_n\}$ be independent and identically distributed (i.i.d) samples from a distribution with cumulative distribution function (CDF) $F(t)$. It is of interest to test whether X follows a specific distribution $F_0(t)$. It is equivalent to test the following hypothesis:

$$\begin{aligned} H_0 &: F(t) = F_0(t), \quad \text{forall } t \in (-\infty, \infty), \\ H_1 &: F(t) \neq F_0(t), \quad \text{for some } t \in (-\infty, \infty). \end{aligned} \tag{1}$$

Without strong assumptions on $F_0(t)$ or $F(t)$, many nonparametric tests have been proposed, including the Kolmogorov–Smirnov test, the Anderson–Darling test, and the Cramér-von Mises test (see Conover (1999) for an overview and references). Despite the differences in the original accounts of these tests, Zhang (2002) proposed a new testing framework to include them as special cases. Moreover, based on this framework, Zhang (2002) also proposed a new class of powerful GoF tests based on nonparametric likelihood ratio (NLR) statistics.

In more details, the original null hypothesis can be transformed to the null hypothesis $H_0^\tau : F(\tau) = F_0(\tau)$, for all $\tau \in (-\infty, \infty)$. The later can be formulated as testing the success rate of a binomial distribution, i.e., $H_0'^\tau : p \equiv P(X \leq \tau) = F_0(\tau)$. This problem is well studied and can be readily solved by

the likelihood-ratio test given the samples $\mathcal{S}_n$, i.e.,

$$L(\tau) = n\Bigg\{\hat{F}_n(\tau)\ln\frac{\hat{F}_n(\tau)}{F_0(\tau)} + (1-\hat{F}_n(\tau))\ln\frac{1-\hat{F}_n(\tau)}{1-F_0(\tau)}\Bigg\}, \quad (2)$$

where $\hat{F}_n(\tau) = n^{-1}\sum_{i=1}^{n}\mathcal{I}(X_i \le \tau)$ is the empirical distribution function (ECDF) based on $\mathcal{S}_n$. $\mathcal{I}(\cdot)$ is the indicator function, which equals one when the condition is true and zero otherwise.

According to the construction, $L(\tau)$ is always nonnegative. Moreover, when H_0^τ is false, $L(\tau)$ is expected to be large. As a consequence, we can test H_0 in (1) by aggregating the information at all $\tau \in (-\infty,\infty)$. Zhang (2002) recommended $Z = \int_{-\infty}^{\infty} L(\tau)\mathrm{d}w(\tau)$, where $w(\tau)$ is a pre-specified weight function. With a properly chosen $w(\tau)$, Z could be a very powerful test statistic compared with many existing methods. One $w(\tau)$ proposed by Zhang (2002) is $\mathrm{d}w(\tau) = [\hat{F}_n(\tau)(1-\hat{F}_n(\tau))]^{-1}\mathrm{d}\hat{F}_n(\tau)$. More importantly, Z has the same null distribution not depending on $F_0(t)$, making it a useful nonparametric and robust test statistic.

2.2. A New Multivariate Goodness-of-Fit Test

Although the univariate GoF test has been proven to be effective and powerful, its direct extension to multivariate distributions is challenging. This is because estimating the multivariate ECDF function, especially when the dimension p is large, is prohibiting with limited samples. In this part, we propose an alternative construction of the multivariate GoF test based on NLR statistics.

Let $\mathcal{S}_n = \{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$, $\mathbf{X} \in \mathbb{R}^p$ be i.i.d. samples from a p-dimensional multivariate distribution with CDF $F(\mathbf{t}), \mathbf{t} \in \mathbb{R}^p$. The variable in the jth dimension, $X_{i,j}$, has the marginal CDF $F_j(t), t \in \mathbb{R}$. A natural and simple idea is to construct p univariate GoF tests for each dimension separately, and combine the test statistics together in a meaningful way. However, this construction ignores the relationship of these p variables. While it might be useful for detecting changes in marginal distributions, it is ineffective in detecting changes in the correlation structure. As a result, we need to strike a balance between the detection capability and the distribution dimensionality. Our idea is to construct a testing statistic for the p-dimensional distribution based on a series of *bivariate* GoF tests. Each bivariate test is designed for two selected dimensions, to detect changes in both marginal distributions and their correlation. Therefore, collectively, the proposed test is able to detect a much larger class of deviations from the null distribution. In the following, we first elaborate on how to construct the bivariate GoF test and later discuss how to combine the bivariate test results together.

To begin with, we consider the joint distribution of $[X_{i,j}, X_{i,k}]$, with CDF $F_{jk}(\boldsymbol{\tau})$, to illustrate the idea, where $\boldsymbol{\tau} = [\tau_1,\tau_2] \in \mathbb{R}^2$ is a bivariate point on the real plane. Similar to the univariate NLR construction, for any given $\boldsymbol{\tau}$, the set of samples $\mathcal{S}_n$ are partitioned into four disjoint regions according to whether $X_{i,j}$ ($X_{i,k}$) is smaller than τ_1 (τ_2) or not (illustrated in Figure 1). Under $H_0^{\boldsymbol{\tau}}$, the numbers of samples $[X_{i,j}, X_{i,k}]$ in each of the four regions follow a multinomial distribution, with probabilities $P_{0,jk}^r(\boldsymbol{\tau}), r = 1,2,3,4$ shown in Figure 1, where $P_{0,jk}^1(\boldsymbol{\tau}) = P(X_{i,j} \le \tau_1, X_{i,k} \le \tau_2)$, $P_{0,jk}^2(\boldsymbol{\tau}) = P(X_{i,j} > \tau_1, X_{i,k} \le \tau_2)$, $P_{0,jk}^3(\boldsymbol{\tau}) = P(X_{i,j} \le \tau_1, X_{i,k} > \tau_2)$, and $P_{0,jk}^4(\boldsymbol{\tau}) = P(X_{i,j} > \tau_1, X_{i,k} > \tau_2)$ under F_0. As a consequence, testing the hypothesis $H_0 : F_{jk}(\boldsymbol{\tau}) = F_{0,jk}(\boldsymbol{\tau})$ for all $\boldsymbol{\tau} \in \mathbb{R}^2$, is equivalent to testing the probabilities of the multinomial distribution. In more detail, we use $\hat{P}_{jk}^r(\boldsymbol{\tau})$ to denote the counterpart of $P_{0,jk}^r(\boldsymbol{\tau})$ for $r = 1,2,3,4$ based on the ECDF $\hat{F}_{jk}(\boldsymbol{\tau}) = n^{-1}\sum_{i=1}^{n}\mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} \le \tau_2)$. Following the likelihood ratio principle, the bivariate NLR statistic can be expressed as

$$T_{jk}(\boldsymbol{\tau}) = n\sum_{r=1}^{4}\hat{P}_{jk}^r(\boldsymbol{\tau})\ln\frac{\hat{P}_{jk}^r(\boldsymbol{\tau})}{P_{0,jk}^r(\boldsymbol{\tau})}. \quad (3)$$

Again $T_{jk}(\boldsymbol{\tau})$ is nonnegative and expected to be large when $H_0^{\boldsymbol{\tau}}$ is false. Then $T_{jk}(\boldsymbol{\tau})$ at different $\boldsymbol{\tau}$ can be aggregated by maximization or integration, e.g., $Z_{jk} = \int_{\boldsymbol{\tau}} T_{jk}(\boldsymbol{\tau})\mathrm{d}w_{jk}(\boldsymbol{\tau})$. Analogue to the choice of $w(t)$ in the univariate case (Zhang (2002), Zou and Tsung (2010)), we use $\mathrm{d}w_{jk}(\boldsymbol{\tau}) = [\prod_{r=1}^{4}\hat{P}_{jk}^r(\boldsymbol{\tau})]^{-1/4}\mathrm{d}\hat{F}_{jk}(\boldsymbol{\tau})$ in this paper. In fact, the determinant of the Fisher information matrix in estimating $(P_{0,jk}^1, P_{0,jk}^2, P_{0,jk}^3, P_{0,jk}^4)$ is simply $\prod_{r=1}^{4}\hat{P}_{0,jk}^r(\boldsymbol{\tau})$, to have higher off-center weights, which leads to

$$Z_{jk} = \sum_{i=1}^{n}\Bigg[\prod_{r=1}^{4}\hat{P}_{jk}^r(\mathbf{X}_{i,jk})\Bigg]^{-1/4} \times \sum_{r=1}^{4}\hat{P}_{jk}^r(\mathbf{X}_{i,jk})\ln\frac{\hat{P}_{jk}^r(\mathbf{X}_{i,jk})}{P_{0,jk}^r(\mathbf{X}_{i,jk})}, \quad (4)$$

where $\mathbf{X}_{i,jk} = [X_{i,j}, X_{i,k}]$ is the subvector of the ith sample. We want to stress that different weight

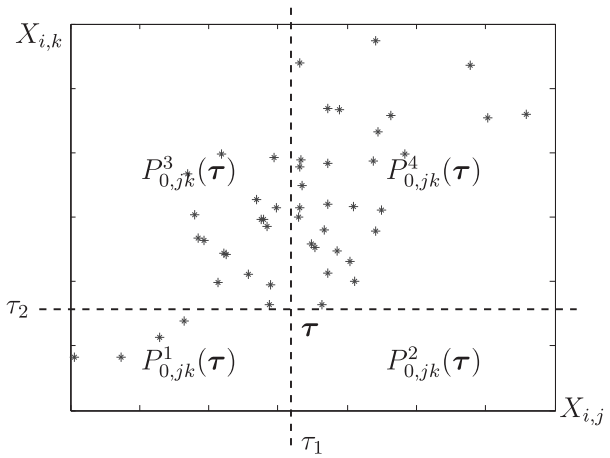


FIGURE 1. Illustration of Bivariate Nonparamemtric Likelihood-Ratio Test.

functions are also possible. Our choice here can balance the detection of different types of shifts. Other weights and their corresponding performance are also discussed in the supplementary material available upon request.

From the bivariate GoF test for the j,kth dimension of $\mathbf{X}$, we can construct a GoF test for the entire p-dimensional distribution. In particular, we select a set of pairs $\mathcal{P} = \{(j_1, k_1), \ldots, (j_q, k_q)\}$, where $1 \le j_i, k_i \le p, j_i \ne k_i, i = 1, \ldots, q$. Note that $Z_{jk} \ge 0$ and is large when H_0 is false. We can aggregate the bivariate test results together by $Z = \sum_{(j,k)\in\mathcal{P}} Z_{jk}$. As a result, Z becomes large when the bivariate distribution of any pair belonging to $\mathcal{P}$ shifts from that specified by $F_0(\mathbf{t})$. As a result, Z is able to detect changes not only in marginal distributions, but also in certain correlations. Even though Z might not cover all the correlations of the p dimensions, the undercoverage can be mitigated by an appropriate design of $\mathcal{P}$. Some guidelines are provided in Section 5 on how to construct $\mathcal{P}$ for different purposes.

Unlike the univariate GoF test, the null distribution of Z depends on $F_0(\mathbf{t})$ and $\mathcal{P}$. In fact, without estimating and utilizing the covariance matrix of $\mathbf{X}$, the test statistic is not affine invariant. In this sense, the proposed multivariate GoF test statistic Z is not *distribution free*. In other words, we cannot find a *constant* c such that under H_0, $P(Z > c) \le \alpha$ for all p-dimensional distributions $F_0(\mathbf{t})$. Nevertheless, in this one-sample test problem, we can always find a cut-off value $c(\alpha, F_0, n)$ such that $P(Z > c(\alpha, F_0, n)) \le \alpha$ under a certain $F_0(\mathbf{t})$. Clear from the notation, the cut-off value depends on F_0 and the sample size n. $c(\alpha, F_0, n)$ can always be approximated through sampling from F_0.

3. A Distribution-Free Multivariate Control Chart

Based on the multivariate GoF test developed in Section 2.2, we can construct the MSPC chart to sequentially monitor the distributional changes in $\mathbf{X}$. We follow the conventional change-point formulation of the MSPC problem. In particular, we assume that there are m_0 i.i.d. reference observations (or IC samples interchangeably) when the process is in control, i.e., $\mathbf{X}_{-m_0+1}, \ldots, \mathbf{X}_0 \in \mathbb{R}^p$. In the monitoring stage, the subsequent ith observation, $\mathbf{X}_i = (X_{1i}, \ldots, X_{pi})^T$, is collected over time following the change-point model

$$\mathbf{X}_i \overset{\text{i.i.d.}}{\sim} \begin{cases} F_0(\mathbf{t}) & \text{for } i = -m_0+1, \ldots, 0, 1, \ldots, \xi, \\ F_1(\mathbf{t}), & \text{for } i = \xi+1, \ldots, \end{cases} \tag{5}$$

where ξ is the unknown change point. F_0 and F_1 are the IC and OC distribution functions, respectively, and assumed to be continuous. Our aim is to construct a robust charting procedure based on the multivariate GoF test to detect the change point ξ as early and accurately as possible, without strong assumptions on either F_0 or F_1. When F_0 is unknown and m_0 is relatively small, the test statistic based on Eq. (4) is not directly applicable for two main reasons. First, the probabilities $P^r_{0,jk}(\boldsymbol{\tau})$, $r = 1, \ldots, 4$ are unknown. Second, the distribution of the test statistic Z and the corresponding cut-off value $c(\alpha, F_0, n)$ cannot be accurately obtained through sampling. Both challenges need to be addressed to deploy the charting scheme based on the multivariate GoF test.

3.1. Construction of Charting Statistics

We first sequentialize the test procedure in Section 2.2 when m_0 is limited. In such cases, we can estimate $P^r_{0,jk}(\boldsymbol{\tau})$, $r = 1, \ldots, 4$ based on the m_0 IC reference samples and update the estimates sequentially in a self-starting way. Particularly, when the chart has not signalled an OC alarm up to the nth sample, the previous $n-1$ samples can be considered as IC samples. Together with the m_0 reference samples, they can be used to estimate the IC distribution by the ECDF function $\hat{F}^n_{0,jk}(\boldsymbol{\tau}) = (m_0 + n - 1)^{-1} \sum_{i=-m_0+1}^{n-1} \mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} \le \tau_2)$, and the correspondingly probabilities in each region (see Figure 1) by frequencies, denoted as $\hat{P}^{n,r}_{0,jk}(\boldsymbol{\tau})$, $r = 1, \ldots, 4$.

In addition, to improve the detection performance on small shifts and to reduce the computational load, we adopt the window limited exponential weighting strategy (see Zou and Tsung (2010), Stoumbos and Sullivan (2002), for examples). Essentially, we consider the w most recent samples as potential OC samples and weight them exponentially in computing the counterparts of $\hat{P}^r_{jk}(\boldsymbol{\tau})$ in Eq.(2). Equivalently, they can be estimated from the weighted ECDF $\hat{F}^n_{jk}(\boldsymbol{\tau} \mid \lambda, w) = a^{-1}_{\lambda,w} \sum_{i=n-w+1}^{n} (1-\lambda)^{n-i} \mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} \le \tau_2)$, where $a_{\lambda,w} = \sum_{n-w+1}^{n}(1-\lambda)^{n-i} = \lambda^{-1}[1-(1-\lambda)^w]$. For notation simplicity, we denote them at the nth sample as $\hat{P}^{n,r}_{jk}(\boldsymbol{\tau}; \lambda, w)$ to highlight their dependence on λ, w. The revision of the testing procedure is schematically illustrated in Figure 2, and the key elements in the test are summarized below:

$$\hat{P}^{n,1}_{0,jk}(\boldsymbol{\tau}) = \sum_{i=-m_0+1}^{n-1} \frac{\mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} \le \tau_2)}{m_0+n-1},$$

$$\hat{P}^{n,1}_{jk}(\boldsymbol{\tau}; \lambda, w) = \sum_{i=n-w+1}^{n} \frac{(1-\lambda)^{n-i}}{\lambda^{-1}[1-(1-\lambda)^w]} \times \mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} \le \tau_2)$$

$$\hat{P}^{n,2}_{0,jk}(\boldsymbol{\tau}) = \sum_{i=-m_0+1}^{n-1} \frac{\mathcal{I}(X_{i,j} > \tau_1, X_{i,k} \le \tau_2)}{m_0+n-1},$$

$$\hat{P}^{n,2}_{jk}(\boldsymbol{\tau}; \lambda, w) = \sum_{i=n-w+1}^{n} \frac{(1-\lambda)^{n-i}}{\lambda^{-1}[1-(1-\lambda)^w]} \times \mathcal{I}(X_{i,j} > \tau_1, X_{i,k} \le \tau_2)$$

$$\hat{P}^{n,3}_{0,jk}(\boldsymbol{\tau}) = \sum_{i=-m_0+1}^{n-1} \frac{\mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} > \tau_2)}{m_0+n-1},$$

$$\hat{P}^{n,3}_{jk}(\boldsymbol{\tau}; \lambda, w) = \sum_{i=n-w+1}^{n} \frac{(1-\lambda)^{n-i}}{\lambda^{-1}[1-(1-\lambda)^w]} \times \mathcal{I}(X_{i,j} \le \tau_1, X_{i,k} > \tau_2)$$

$$\hat{P}^{n,4}_{0,jk}(\boldsymbol{\tau}) = \sum_{i=-m_0+1}^{n-1} \frac{\mathcal{I}(X_{i,j} > \tau_1, X_{i,k} > \tau_2)}{m_0+n-1},$$

$$\hat{P}^{n,4}_{jk}(\boldsymbol{\tau}; \lambda, w) = \sum_{i=n-w+1}^{n} \frac{(1-\lambda)^{n-i}}{\lambda^{-1}[1-(1-\lambda)^w]} \times \mathcal{I}(X_{i,j} > \tau_1, X_{i,k} > \tau_2).$$

In practice, $\lambda = 0.1$ or 0.05 is the common choice in EWMA-type charts and w is chosen such that $(1-\lambda)^w$ is small, say 0.05. This window-limited statistic barely influences the detection performance, but can significantly reduce the computational load (see the discussion and analysis later). After accommodating these changes, the test statistic of Eq. (4) at the nth sample becomes

$$Z^n_{jk}(\lambda, w) = \sum_{i=n-w+1}^{n} \frac{(1-\lambda)^{n-i}}{\left[\prod_{r=1}^{4} \hat{P}^{n,r}_{jk}(\mathbf{X}_{i,jk}; \lambda, w)\right]^{1/4}} \times \sum_{r=1}^{4} \hat{P}^{n,r}_{jk}(\mathbf{X}_{i,jk}; \lambda, w) \ln \frac{\hat{P}^{n,r}_{jk}(\mathbf{X}_{i,jk}; \lambda, w)}{\hat{P}^{n,r}_{0,jk}(\mathbf{X}_{i,jk})}.$$

By aggregating the statistic for all the pairs in $\mathcal{P}$, the charting statistic becomes

$$Z_n(\lambda, w) = \sum_{(j,k)\in\mathcal{P}} Z^n_{jk}(\lambda, w). \tag{6}$$

Even though Eq. (6) appears to be more complicated than Eq. (4) in Section 2.2, they have similar computational complexity (see the discussion in Appendix). For notation simplicity, we simply use Z_n when there is no confusion.

3.2. Data-Dependent Control Limits

The charting statistic of Eq. (6) indicates that, when there are distributional changes from the IC distribution F_0, Z_n is expected to be large. To make the chart operational, we need to determine the control limits such that the chart has satisfactory IC and OC performances.

Unfortunately, without knowing F_0, it is difficult to find the limit $c(\alpha, F_0, n)$ such that $P(Z_n >$

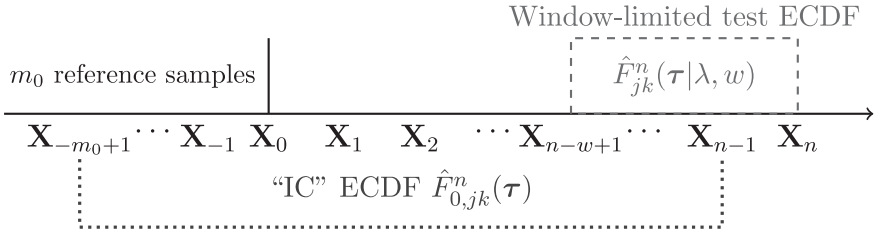


FIGURE 2. Illustration of the EWMA-Based On-Line GoF Test.

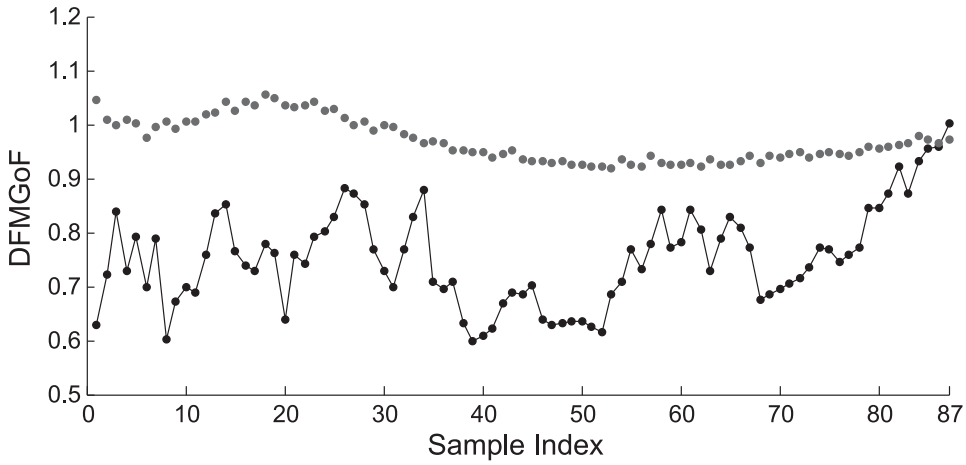


FIGURE 3. An Example of DFMGoF Chart. The black dots represent the charting statistic Z_n and the red dots represent the control limit $H_n(\alpha, \mathcal{F}_{m_0+1})$. The chart triggered an OC alarm at $n = 87$.

$c(\alpha, F_0, n)) \leq \alpha$ when all $\mathbf{X}_i \sim F_0$, $i = -m_0 + 1, \ldots, 0, \ldots, n$. On the other hand, we observe that the charting statistic of Eq. (6) depends on the samples $\{\mathbf{X}_{-m_0+1}, \ldots, \mathbf{X}_0, \ldots, \mathbf{X}_n\}$ only through the ECDF, denoted by $\mathcal{F}_{m_0+n}$. Equivalently speaking, the conditional distribution $P(Z_n \mid \mathcal{F}_{m_0+n})$ is free from F_0. This conditional distribution-free property provides an alternative avenue to determine the control limits. In fact, because $P(Z_n \mid \mathcal{F}_{m_0+n})$ is free from F_0, we can always find a quantity $H_n(\alpha, \mathcal{F}_{m_0+n})$ such that $P(Z_n > H_n(\alpha, \mathcal{F}_{m_0+n}) \mid \mathcal{F}_{m_0+n}) \leq \alpha$. Note that $H_n(\alpha, \mathcal{F}_{m_0+n})$ depends on $\mathcal{F}_{m_0+n}$ only. Hence it is a random variable but has a fixed realization given $\mathcal{F}_{m_0+n}$. This implies that, in different independent runs of the control chart, we have different $\mathcal{F}_{m_0+n}$ and correspondingly different realizations of $H_n(\alpha, \mathcal{F}_{m_0+n})$. That means the value of $H_n(\alpha, \mathcal{F}_{m_0+n})$ depends on the samples $\{\mathbf{X}_{-m_0+1}, \ldots, \mathbf{X}_0, \ldots, \mathbf{X}_n\}$ at each step, therefore given its name *data-dependent* limits.

In practice, we can design $H_n(\alpha, \mathcal{F}_{m_0+n})$ such that the chart has a prespecified ARL_0. However, as recognized in the literature, it is often insufficient to summarize the run-length behavior by ARL, especially for self-starting control charts (Hawkins and Maboudou-Tchao (2007), Zou and Tsung (2010)). It is more desired that the conditional false-alarm rate is controlled at each step such that the IC run length is geometrically distributed (Hawkins and Olwell (1998)). To achieve this ideal run-length distribution, we can adjust 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 is a prespecified constant α. Equivalently, $H_i(\alpha, \mathcal{F}_{m_0+i})$, $i = 1, 2, \ldots$, needs to satisfy

$$\begin{aligned} &P(Z_1 > H_1(\alpha, \mathcal{F}_{m_0+1}) \mid \mathcal{F}_{m_0+1}) = \alpha, \\ &P(Z_n > H_n(\alpha, \mathcal{F}_{m_0+n}) \mid \\ &\quad Z_i \leq H_i(\alpha, \mathcal{F}_{m_0+i}), 1 \leq i < n, \mathcal{F}_{m_0+n}) = \alpha \end{aligned} \quad (7)$$

for $n > 1$. Subsequently, we can formally define the following charting procedure, termed as distribution-free multivariate GoF chart (abbreviated as DFMGoF), with the run length

$$\text{RL} = \min\{n;\ Z_n \geq H_n(\alpha, \mathcal{F}_{m_0+n}),\ n \geq 1\}. \quad (8)$$

An example of the chart operation is shown in Figure 3 for illustration. Because of the conditional distribution-free property of $P(Z_n \mid \mathcal{F}_{m_0+n})$, the conditional false-alarm rate α in Eq. (7) holds regardless of the IC distribution F_0 or its dimension. As a result, our construction ensures that $P(\text{RL} = n) = (1-\alpha)^{n-1}\alpha$ exactly, and correspondingly $\text{ARL}_0 = 1/\alpha$. This result is remarkable as it does not require the distributional type or parameters of F_0 and can always ensure the desired IC run-length distribution (see results in Section 4.1). It can also operate with small m_0, which is crucial in short-run processes and mass-customerization applications.

Despite the existence of $H_n(\alpha, \mathcal{F}_{m_0+n})$, forall $n \geq 1$ theoretically, it is often analytically infeasible to find them to satisfy Eq. (7). To make the charting procedure practical and generally applicable, we propose a computational algorithm to find $H_n(\alpha, \mathcal{F}_{m_0+n})$, forall $n \geq 1$. The algorithm is based on the fundamental *permutation principle*. In more detail, if the pro-

cess has being in control until the nth sample, $\mathcal{S}_n = \{\mathbf{X}_{-m_0+1}, \ldots, \mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n\}$ are i.i.d. samples from the IC distribution F_0. As a result, any permuted $\mathcal{S}_n^\nu = \{\mathbf{X}_{v_{-m_0+1}}, \ldots, \mathbf{X}_{v_0}, \mathbf{X}_{v_1}, \ldots, \mathbf{X}_{v_n}\}$, where $\{v_{-m_0+1}, \ldots, v_0, v_1, \ldots, v_n\}$ is simply a random permutation of the index set $\{-m_0+1, \ldots, 0, 1, \ldots, n\}$, have the same distribution as $\mathcal{S}_n$. As a result, the charting statistic Z_n computed from $\mathcal{S}_n$ has the identical distribution as Z_n^ν from the permuted $\mathcal{S}_n^\nu$. By generating a large number of permuted samples $\mathcal{S}_n^\nu$ and computing their corresponding charting statistics Z_n^ν, for $\nu = 1, \ldots, b$, we are able to approximate the conditional distribution $P(Z_n \mid \mathcal{F}_{m_0+n})$ through sample approximation. In particular, the desired control limit can be approximated by the sample quantile. Because this computational procedure is valid regardless of the IC distribution F_0, the algorithm can always ensure the validity of Eq. (7). See Chen et al. (2015) for detailed theoretical analyses. The idea can be formalized into the following procedure:

(i) For $n = 1$, generate b permutations $\mathcal{S}_1^\nu, \nu = 1, \ldots, b$, independently from $\mathcal{S}_1$ and compute their corresponding charting statistics, $Z_1^\nu, \nu = 1, 2, \ldots, b$. Find the $(1-\alpha)$ sample quantile of $Z_1^\nu, \nu = 1, \ldots, b$ as $H_1(\alpha, \mathcal{F}_{m_0+1})$.

(ii) For $n > 1$ and a permutation $\mathcal{S}_n^\nu$ from $\mathcal{S}_n$, compute Z_i^ν for $\max\{1, n-w+1\} \le i \le n$. If $Z_i^\nu \le H_i(\alpha, \mathcal{F}_{m_0+i})$ for all $\max\{1, n-w+1\} \le i < n$, accept Z_n^ν as a valid permutation statistic. Otherwise, $\mathcal{S}_n^\nu$ is discarded and a new permutation is drawn. Repeat this procedure until b valid Z_n^ν are obtained and find the $(1-\alpha)$ sample quantile of $Z_n^\nu, \nu = 1, \ldots, b$ as $H_n(\alpha, \mathcal{F}_{m_0+n})$.

In (ii), we only need to estimate the quantile of $P(Z_n \mid Z_i \le H_i(\alpha, \mathcal{F}_{m_0+i}), \max\{1, n-w+1\} \le i < n, \mathcal{F}_{m_0+n})$ instead of the one in Eq. (7) to reduce the computational complexity, especially when n is large, because only the most w recent Z_i^ν of every permutation need to be computed. A detailed analysis of the algorithm complexity is discussed in Appendix A.1.

The proposed procedure of finding $H_n(\alpha, \mathcal{F}_{m_0+n})$ is based on the permutation principle, which is different from the usual bootstrap method. Similar to other permutation tests, the charting procedure does not need any distributional assumption to have the exact IC run-length distribution. This property makes the proposed chart significantly different from existing methods. In practice, a complete enumeration of all permutations is infeasible. Random permutations of size $b = 10q/\alpha$ should be sufficient for reliable approximations, which, in turn, roughly requires $b/(1-\alpha)^w$ permutation trials when $n \ge w$. As long as m_0 and n are not too small and b is sufficiently large, Eq. (7) holds well using the estimated limits $H_n(\alpha, \mathcal{F}_{m_0+n})$, $n = 1, \ldots,$ from the algorithm.

Satisfactory performance is possible because of the *data-dependent* nature of the charting procedure. In other words, after observing $\mathbf{X}_n$, along with Z_n we need to determine the corresponding limit $H_n(\alpha, \mathcal{F}_{m_0+n})$ on line. This is fundamentally different from the approach of dynamic control limits originally proposed by Margavio et al. (1995) and Lai (1995), which still uses a fixed sequence of limits for a given F_0. However, these data-dependent limits come at a cost of heavy computational load required in the permutation procedure. However, this procedure becomes still feasible as the high-performance computing advances. For example, for a chart with $q = 5$, $b = 10000$, and $w = 58$, at $n = 100$, it takes 5 seconds to calculate the charting statistic and the corresponding limit on a computational node with 32 threads. For time-critical applications, more computational resources can be invested to speed up the charting operation.

4. Simulation Studies

In this section, we present some simulation results to demonstrate the performance of DFMGoF. Unfortunately, fair comparisons between DFMGoF and alternative charts are difficult because DFMGoF is designed to detect general distributional changes and is self-starting, only requiring a small reference sample size. To this end, we consider two self-starting charts that can monitor the mean vector and the covariance matrix simultaneously, the self-starting EWMAC chart (SSEWMAC; Maboudou-Tchao and Hawkins (2011)) and the chart based on the change-point model together with the GLR test (ChangePt; Zamba and Hawkins (2009)). Both charts are designed to detect either mean shifts or covariance matrix shifts and are parametric, assuming that process variables follow the multivariate normal distribution. In addition, we also consider the chart based on real-time contrasts (RTC; Deng et al. (2012)), which is nonparametric but requires sufficient reference samples to attain the specified IC performance. Most recently, Holland and Hawkins (2014) proposed a nonparametric multivariate change-point model for MSPC. It is robust and works well for the elliptical distribution family. However, the chart is designed to

detect mean shifts only and hence is not included for the comparisons here.

To test the robustness of these charts, we consider the following distributions in our numerical studies: (i) multivariate normal; (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 $\text{Gam}_{p,\zeta}$. These distributions are commonly used in the literature to study the robustness of charting performance. For easy reference, random-number generation and useful moments of these distributions are included in Appendix A.2. In the simulation, we consider $p = 10$ and 30, representing low-dimensional and high-dimensional cases, respectively. We set $\text{ARL}_0 = 200$ and $m_0 = 100$. Clearly, such few reference samples are not able to provide any meaningful estimate of the distributional parameters when $p = 10$ or 30.

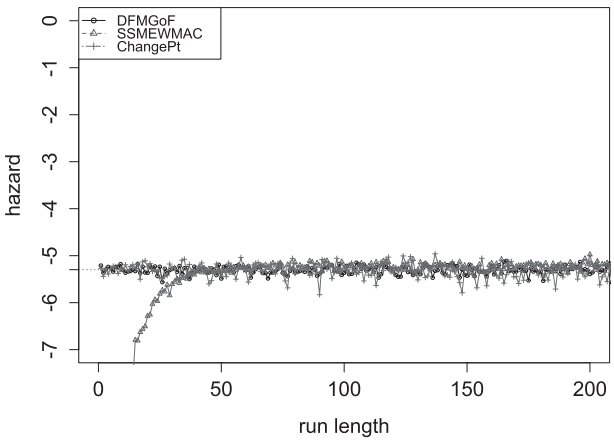
Without loss of generality, for each distribution, the mean vector $\boldsymbol{\mu}_0$ is set to be $\mathbf{0}$, the covariance matrix $\boldsymbol{\Sigma}_0$ is chosen to be diagonal as $\sigma^2\mathbf{I}$. For DFMGoF, in every replicate, we simply choose the $\lceil p/2 \rceil$ most correlated pairs of $\mathbf{X}$ based on its m_0 reference samples to form its corresponding $\mathcal{P}$. In addition, we set $w = 28$ when $\lambda = 0.1$ and $w = 58$ when $\lambda = 0.05$. For RTC, following the guidelines in Deng et al. (2012), the random forest algorithm is used to classify the IC samples and the on-line samples. A group of 10 samples from either class are used to train the classifier to get the charting statistic. In the subsequent simulation results, the quantities are obtained based on 10000 replicates without other notes. Additional simulation results, including cases with nondiagonal covariance matrices, are summarized in the supplementary material available upon request.

4.1. In-Control Performance Comparison

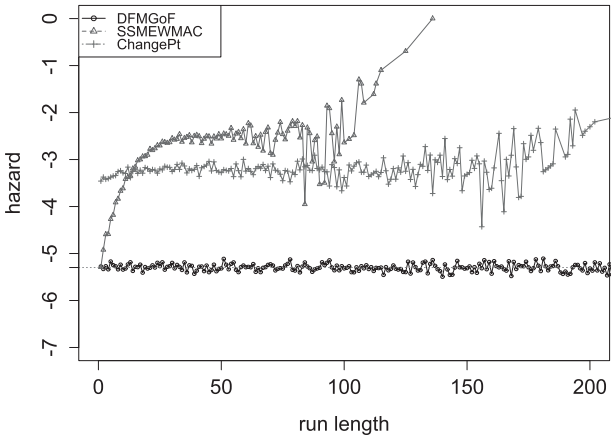
We first compare their IC performance in terms of ARL_0, standard deviation of the run-length (SDRL), and the false-alarm rate during the first 30 observations, i.e., $\text{FAR} = P(\text{RL} \leq 30)$. According to (Hawkins and Olwell (1998)), the IC run-length distribution of a chart is satisfactory if it is close to the geometric distribution. Correspondingly, if the run-length distribution is geometric, with $\alpha = 0.005$ and $\text{ARL}_0 = 200$, we expect SDRL= 200 and FAR= 0.140. Table 1 summarizes the IC performance of these charts for different distributions and dimensions. The control limits of SSEWMAC and ChangePt are set assuming that the data follow the multivariate normal distribution based on the methods in Maboudou-Tchao and Hawkins (2011) and Zamba and Hawkins (2009). On the other hand, the control limits of RTC are determined through resampling from the m_0 reference samples.

TABLE 1. IC Performance of the DFMGoF, RTC, ChangePt, and SSEWMAC Charts with Multivariate Normal, $t_{p,5}$, and $\text{Gam}_{p,5}$ Observations when $p = 10$ and $m_0 = 100$

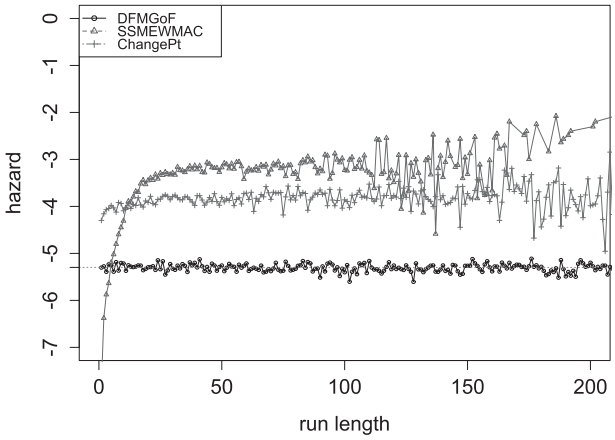
		$\lambda = 0.1$			$\lambda = 0.05$		
	Method	ARL_0	SDRL	FAR	ARL_0	SDRL	FAR
	Geometric	200	200	0.140	200	200	0.140
Norm	DFMGoF	199	199	0.142	192	192	0.147
	SEWMAC	202	189	0.077	198	171	0.032
	RTC	25.7	8.22	0.705	—	—	—
	ChangePt	199	191	0.128	—	—	—
$t_{p,5}$	DFMGoF	198	195	0.141	194	190	0.142
	SSEWMAC	24.2	14.3	0.721	0.8	16.8	0.531
	RTC	29	10.4	0.665	—	—	—
	ChangePt	24.3	23.5	0.711	—	—	—
$\text{Gam}_{p,3}$	DFMGoF	194	192	0.146	197	195	0.146
	SSEWMAC	42.2	29.4	0.415	53.8	33.6	0.239
	RTC	25.7	6.66	0.704	—	—	—
	ChangePt	56.0	51.1	0.379	—	—	—



(a) Multivariate normal



(b) Multivariate t_5



(c) Multivariate Gam_3

FIGURE 3. An Example of DFMGoF Chart. The black dots represent the charting statistic Z_n and the red dots represent the control limit $H_n(\alpha, \mathcal{F}_{m_0+1})$. The chart triggered an OC alarm at $n = 87$.

Table 1 shows that DFMGoF has satisfactory IC performance. For every distribution, DFMGoF's ARL_0 can attain the designed value 200 closely, and its run-length distribution is close to the geometric distribution based on the comparisons of its ARL_0, SDRL, and FAR with the ideal ones. On the other hand, both ChangePt and SSEWMAC have satisfactory ARL_0s when $\mathbf{X}$ is multivariate normal. This is not surprising because their control limits are obtained under the normality assumption. However, when $\mathbf{X}$ follows the $t_{p,5}$ or $\text{Gam}_{p,3}$ distribution, their ARL_0s are far smaller than 200, indicating that excessive false alarms are expected for nonnormal distributions. Table 1 also reveals that RTC has unsatisfactory IC performance for all the three distributions. This is because, to set the control limit correctly, RTC requires knowing the exact distribution type and parameters. If either is unknown, a large number of reference samples are required to obtain the limit through resampling. Samples of size $m_0 = 100$ are clearly insufficient to get the accurate limit, and hence lead to unsatisfactory IC performance. In fact, many other nonparametric MSPC charts share the same problem: if the reference sample size is too small, the IC performance is not guaranteed without knowing the exact F_0.

To further validate our conclusions, Figure 4 plots the hazard rates of these IC run-length distributions. Ideally, if the run-length distribution is geometric, the hazard rate is constant along the streamline. In contrast, elevated hazard rates at the beginning often lead to excessive early false alarms. Figure 4 clearly demonstrates that regardless of the form of F_0, DFMGoF always has the geometrically distributed run-length distribution. While SSEWMAC and ChangePt have satisfactory run-length distributions when $\mathbf{X}$ is multivariate normal, their distribu-

tions are heavily distorted when $\mathbf{X}$ is different from the multivariate normal distribution. This demonstrates that DFMGoF is exactly distribution free and has a satisfactory IC run-length distribution, as indicated in Section 3. Its robust performance makes it especially useful when m_0 is small and F_0 is unknown.

4.2. Out-of-Control Performance Comparison

In this section, we compare the OC performance of the four competing charts. Here we only consider the steady-state ARL (SSARL), meaning that any series where a signal occurs before the true change point ξ is discarded. We fix $\xi = 25$ in all cases. Furthermore, to have a fair comparison, we adjust the control limits of all charts such that their $\text{ARL}_0 = 200$ for each of the three distributions considered. It should be noted that this adjustment could only be used for simulation comparisons, but not applicable in practice because the true IC distribution is usually unknown.

Similar to other MSPC studies, it is impossible to enumerate all the change patterns to allow a full-scale study of the charts' performances. Following similar studies in the literature (Zou and Tsung (2011), Zou et al. (2012), Maboudou-Tchao and Hawkins (2011), Zamba and Hawkins (2009)), here we consider three scenarios as examples: (1) mean shifts in the first $\lfloor p/5 \rfloor$ dimensions of size δ, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \delta\mathbf{e}$ with

$$\mathbf{e} = (\underbrace{1, \ldots, 1}_{1,\ldots,\lfloor p/5 \rfloor}, 0, \ldots, 0)^{\mathrm{T}};$$

(2) variance shifts in all the p dimensions of size δ, i.e., $\boldsymbol{\Sigma}_1 = \delta\boldsymbol{\Sigma}_0$; (3) correlation shifts in the first $\lfloor p/5 \rfloor$ dimensions of size ρ, i.e., $\sigma_{i,i+1} = \sigma_{i+1,i} = \rho$, for $i = 1, 3, \ldots, \lfloor p/5 - 1 \rfloor$.

4.2.1. Comparisons in Detecting Mean Shifts

We compare their performance in detecting mean shifts of size $\delta = 0.25, 0.5, 1, 2, 4$ respectively. Table 2 illustrates that DFMGoF outperforms the other three charts when the shift size is small ($\delta = 0.25, 0.5$) and has comparable detection performance as RTC and SSEWMAC when the shift size is moderate to large. On the other hand, ChangePt does not perform satisfactorily compared with the other three charts, especially when $p = 30$. From the comparison, DFMGoF is particularly good at detecting small shifts. Similar observations have been made in the univariate GoF chart (Zou and Tsung (2010)).

We also note that, in our simulation settings, the charts generally have smaller OC ARLs when $p = 30$ than when $p = 10$ given the same δ. This is because the OC performance is largely determined by the Mahalanobis distance of the shifted mean vector from the IC one, i.e., $\Delta = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$ (see Maboudou-Tchao and Hawkins (2011) for a related discussion). Given the $\boldsymbol{\Sigma}_0$ and the change pattern in our study, we have $\Delta_{p=30} > \Delta_{p=10}$ given the same δ. This can partially explain the better performance when $p = 30$. Moreover, this also assures us that, even if we only consider a specific change pattern, the simulation results are representative as long as another change pattern $\boldsymbol{\mu}_1$ has the same Mahalanobis distance from $\boldsymbol{\mu}_0$ as the current pattern. As a result, it might not be necessary to compare exhaustive change patterns, especially for the elliptical distribution family.

4.2.2. Comparisons in Detecting Variance Shifts

We first compare their performance in detecting increases in variance of size $\delta = 1.25, 1.5, 2, 4$. This is often of more interest than decreases in variance because the former generally leads to a larger number of nonconforming parts and indicates presence of some assignable causes. (See Montgomery (1991) for a detailed discussion.) As Table 3 summarizes, when the distribution is multivariate normal, DFMGoF does not have advantages in the detection speed. SSEWMAC, which is designed under the multivariate normal assumption, has consistently better results. However, when $\mathbf{X}$ follows the $t_{p.5}$ distribution, SSEWMAC performs worse than DFMGoF. When $\mathbf{X}$ follows the $\text{Gam}_{p,3}$ distribution, the comparison is intriguing. As noted in Appendix A.2, the increases in variance of the shifted dimensions also lead to the increases in the mean of them. As a result, all charts have very good performance in detecting such changes. Table 3 also shows that RTC has superior performance across different types of distributions, especially in detecting small to moderate shifts. However, this is mainly because that RTC is ARL-biased in monitoring variance shifts. In other words, when the variance decreases, RTC has a larger ARL than ARL_0.

To demonstrate this point, we compare their performance in detecting decreases in variance, as shown in Figure 5 for the 10-dimensional multivariate normal distribution. Simulations for other distributions and other dimensions reveal similar results. Figure 5 clearly shows that the charts perform quite differently in detecting decreases in variance. In particular, RTC is not able to detect the shifts efficiently.

TABLE 2. OC ARL Comparison in Detecting Mean Shifts when $m_0 = 100$ and $\lambda = 0.05$. Numbers in parentheses are SDRL values

	p	δ	DFMGoF	RTC	ChangePt	SSEWMAC
Norm	10	0.25	93.9 (144)	132 (144)	186 (187)	122(139)
		0.5	36.1 (36.2)	51.8 (55.0)	134 (138)	70.3 (86.9)
		1.0	12.4 (5.62)	12.1 (6.93)	36.4 (20.3)	11.6 (5.88)
		2.0	5.83 (1.93)	6.23 (1.62)	12.4 (4.40)	4.04 (1.75)
		4.0	3.93 (1.12)	4.72 (1.10)	4.81 (1.59)	1.59 (0.59)
	30	0.25	83.1 (114)	100 (112)	186 (183)	87.8 (95.7)
		0.5	20.6 (11.4)	24.7 (22.0)	148 (151)	22.9 (15.6)
		1.0	8.40 (3.01)	7.62 (3.00)	52.9 (32.0)	9.09 (3.03)
		2.0	4.08 (1.17)	4.82 (1.22)	23.9 (9.82)	4.60 (1.27)
		4.0	2.74 (0.69)	3.90 (0.98)	12.2 (5.53)	2.37 (0.62)
$t_{p,5}$	10	0.25	140 (161)	151.6 (191)	188 (183)	162 (166)
		0.5	58.5 (79.5)	68.3 (80.9)	155 (158)	88.9 (109)
		1.0	16.4 (8.50)	15.1 (11.6)	59.9 (46.0)	21.4 (15.4)
		2.0	7.37 (2.64)	6.74 (1.80)	19.9 (8.23)	8.10 (2.94)
		4.0	4.42 (1.26)	5.21 (1.35)	8.21 (3.16)	4.11 (1.24)
	30	0.25	127 (155)	136 (164)	184 (181)	140 (133)
		0.5	37.7 (45.7)	45.6 (50.9)	154 (166)	55.3 (63.6)
		1.0	11.5 (4.52)	9.38 (3.93)	64.4 (46.3)	14.2 (5.36)
		2.0	5.46 (1.64)	5.77 (1.52)	31.5 (15.0)	6.43 (1.87)
		4.0	3.31 (0.84)	4.58 (1.17)	15.8 (8.53)	3.60 (0.97)
$\text{Gam}_{p,3}$	10	0.25	88.3 (115)	147 (169)	188 (184)	137 (144)
		0.5	26.8 (15.3)	59.3 (76.9)	152 (163)	67.2 (87.8)
		1.0	13.2 (4.99)	10.5 (5.32)	52.9 (38.9)	17.4 (9.30)
		2.0	7.27 (2.29)	6.43 (1.64)	17.2 (6.91)	7.39 (2.68)
		4.0	4.57 (1.31)	5.05 (1.25)	6.87 (2.56)	3.40 (1.20)
	30	0.25	43.4 (37.2)	126 (140)	183 (177)	109 (110)
		0.5	17.2 (6.88)	28.4 (26.6)	159 (154)	35.7 (36.9)
		1.0	9.20 (2.89)	7.43 (1.82)	62.5 (43.1)	11.9 (4.63)
		2.0	5.16 (1.46)	5.25 (1.27)	27.8 (11.5)	5.72 (1.67)
		4.0	3.30 (0.48)	4.13 (1.00)	14.7 (6.50)	2.85 (0.91)

In addition, DFMGoF performs better than SSEWMAC, which is contrary to the cases when variance increases. As a result, it is important to recognize the shifts that are of most importance. Without clear preference, SSEWMAC and DFMGoF give more balanced protection against unknown variance shifts.

4.2.3. Comparisons in Detecting Correlation Shifts

We also compare their performance in detecting correlation shifts. Among various shift patterns, here we focus on a pattern that is commonly used in the literature (Zamba and Hawkins (2009)): the correlation between two neighbor variables creeps into the process with size ρ ranging from 0 to 1. In addition, only the correlations of the first $[p/5]$ dimensions are changed to make the detection even harder. Table 4 summarizes the OC ARLs of different charts under different shift sizes and different distributions. It shows that DFMGoF consistently performs best in almost all the scenarios. In contrast, while SSEWMAC performs satisfactorily for the multivariate normal distribution, its performance deteriorates significantly when the distribution is nonnormal. ChangePt performs slightly better and more robustly than SSEWMAC, though still not as well

TABLE 3. OC ARL Comparison in Detecting Variance Shifts when $m_0 = 100$ and $\lambda = 0.05$. Numbers in parentheses are SDRL values

	p	δ	DFMGoF	RTC	ChangePt	SSEWMAC
Norm	10	1.25	114.6(130)	58.9(60.9)	176(178)	77.2(95.6)
		1.5	52.3(63.9)	25.6 (22.7)	91.9 (87.4)	24.8 (32.0)
		2.0	15.9 (10.2)	10.3 (5.59)	25.1 (13.2)	7.61 (4.90)
		4.0	5.80 (2.51)	5.69 (1.78)	6.55 (2.98)	2.23 (1.11)
	30	1.25	84.4 (109)	31.4 (39.5)	178 (183)	71.1 (92.9)
		1.5	21.9 (20.6)	12.3 (7.81)	100 (88.6)	17.8 (20.6)
		2.0	8.61 (4.40)	6.92 (2.48)	30.8 (14.9)	6.85 (2.28)
		4.0	3.70 (1.37)	4.35 (1.39)	9.29 (3.28)	2.62 (0.49)
$t_{p,5}$	10	1.25	119 (142)	71.4 (83.5)	153 (194)	125 (136)
		1.5	67.9 (82.6)	37.4 (40.4)	86.5 (96.1)	80.9 (92.3)
		2.0	24.7 (26.8)	15.6 (12.3)	34.2 (27.0)	37.5 (50.2)
		4.0	7.97 (4.14)	6.98 (2.39)	9.17 (5.59)	7.75 (5.68)
	30	1.25	91.9 (112)	51.5 (58.5)	147 (154)	125 (136)
		1.5	41.5 (50.6)	24.8 (25.0)	82.2 (83.5)	80.8 (92.3)
		2.0	15.1 (11.2)	10.7 (7.22)	36.7 (24.2)	37.5 (50.2)
		4.0	5.54 (2.62)	5.77 (2.01)	11.3 (6.30)	7.75 (5.68)
$\text{Gam}_{p,3}$	10	1.25	24.8 (19.0)	23.8 (20.7)	38.2 (31.7)	16.4 (12.9)
		1.5	10.3 (4.57)	9.04 (3.81)	12.9 (7.11)	6.36 (3.77)
		2.0	5.48 (2.05)	5.68 (1.81)	4.79 (2.59)	2.57 (1.41)
		4.0	2.71 (0.76)	3.89 (1.20)	1.30 (0.50)	1.06 (0.24)
	30	1.25	14.7 (7.04)	12.6 (7.34)	41.2 (23.3)	11.4 (5.71)
		1.5	6.80 (2.56)	6.29 (1.97)	17.3 (7.43)	5.10 (2.04)
		2.0	3.74 (1.18)	4.35 (1.25)	7.12 (2.68)	2.11 (0.85)
		4.0	1.99 (0.52)	2.94 (0.86)	1.90 (0.65)	1.02 (0.18)

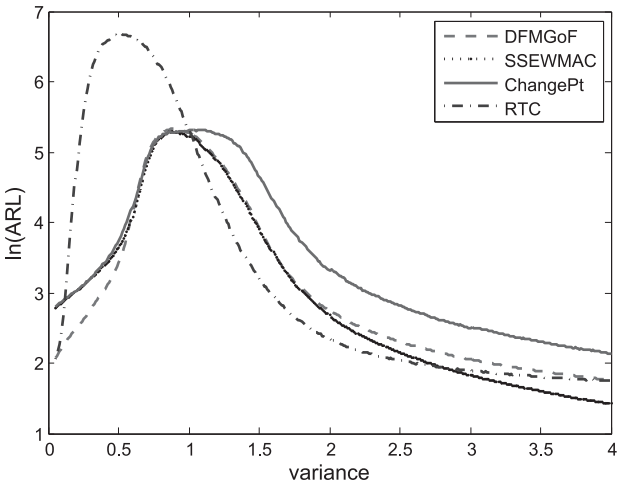


FIGURE 5. OC ARL Curses in Detecting Variance Shifts for the Multivariate Normal Distribution with $p = 10$, $m_0 = 100$, and $\lambda = 0.05$.

as DFMGoF. RTC almost has no detection power for correlation shifts.

We have also conducted many other simulations for different process shift patterns along with other target ARL_0s (370 and 500). Their results suggest that the general observations and conclusions made above still hold. These additional simulation results are available from the authors upon request.

5. Guidelines on Pair Selection

In this section, we investigate how the pair selection influences the charting performance. As mentioned earlier, in MSPC, there are innumerable shift patterns. Even for mean shifts only, it is difficult to find a single method that is dominantly better than the rest in detecting shifts in all directions in the p-dimensional space. We can demonstrate that the most appropriate pair selection in fact goes long with

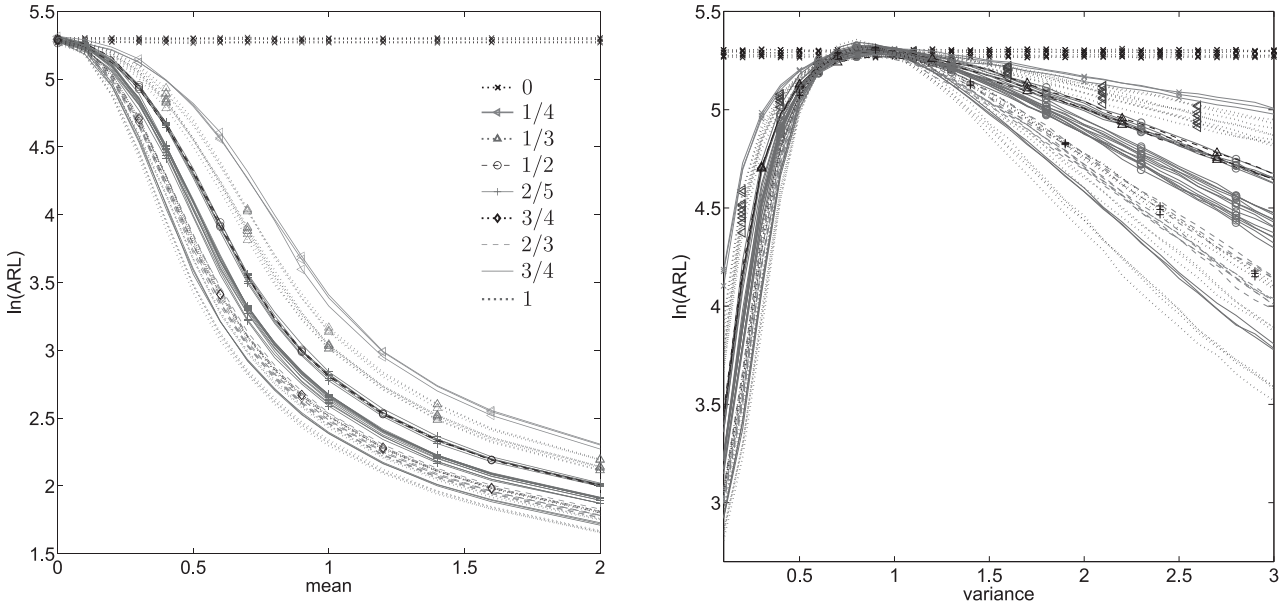
TABLE 4. OC ARL Comparison in Detecting Correlation Shifts when $m_0 = 100$ and $\lambda = 0.05$. Numbers in parentheses are SDRL values

	p	δ	DFMGoF	RTC	ChangePt	SSEWMAC
Norm	10	0.3	174 (188)	203 (215)	189 (176)	163 (171)
		0.5	123 (138)	193 (220)	161 (158)	120 (125)
		0.7	67.4 (77.9)	178 (195)	82.2 (64.0)	62.9 (56.3)
		0.9	28.9 (13.6)	162 (174)	32.4 (14.6)	32.1 (14.0)
	30	0.3	148 (158)	195 (207)	189 (185)	163 (147)
		0.5	84.9 (96.8)	191 (209)	165 (168)	129 (128)
		0.7	35.8 (24.9)	174 (187)	82.7 (55.9)	67.7 (44.3)
		0.9	19.8 (7.51)	163 (178)	37.8 (14.2)	40.7 (12.8)
$t_{p,5}$	10	0.3	168 (180)	192 (225)	189 (186)	184 (172)
		0.5	133 (150)	197 (224)	149 (49.3)	183 (175)
		0.7	77.0 (95.0)	188 (220)	83.2 (74.5)	177 (174)
		0.9	31.2 (16.1)	171 (200)	36.7 (16.8)	138 (134)
	30	0.3	158 (168)	204 (243)	184 (185)	179 (149)
		0.5	107 (137)	189 (227)	139 (143)	174 (145)
		0.7	48.7 (54.5)	181 (218)	75.9 (56.8)	172 (144)
		0.9	22.8 (9.38)	188 (233)	41.4 (15.4)	132 (105)
$\text{Gam}_{p,3}$	10	0.3	194 (197)	203 (231)	196 (185)	181 (173)
		0.5	187 (183)	200 (212)	186 (186)	169 (166)
		0.7	140 (142)	193 (212)	144 (147)	159 (162)
		0.9	45.1 (38.5)	174 (201)	49.9 (30.3)	84.2 (82.5)
	30	0.3	176 (179)	197 (211)	192 (190)	172 (150)
		0.5	159 (159)	197 (208)	185 (178)	171 (147)
		0.7	93.9 (106)	187 (205)	149 (152)	147 (132)
		0.9	26.2 (12.6)	169 (180)	51.7 (31.2)	68.5 (44.1)

the shift pattern of most interest. Of course, in practice, we might not be able to have sufficient domain knowledge in prioritizing the importance of different shift patterns. However, the bottom line is that, through the standard pair selection, our chart can detect all the marginal mean or variance shifts as well as certain correlation structure shifts. We illustrate this point using some additional numerical results as follows.

We consider a small-scale problem for illustration purposes, which detects changes in the 4-dimensional multivariate normal distribution. We enumerate all possible pair combinations, which correspond to all the nonempty subsets of $\mathcal{P}_0 = \{(x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4), (x_3, x_4)\}$. Altogether, there are $2^6 - 1 = 63$ nonempty subsets (pair selections). We set the IC mean vector $\boldsymbol{\mu}_0 = \mathbf{0}$, the covariance matrix $\boldsymbol{\Sigma}_0 = (\sigma_{ij,0})$, with $\sigma_{ii,0} = 1$ and $\sigma_{ij,0} = 0.3^{|i-j|}$ for $i \neq j, i, j = 1, 2, \ldots, p$. The OC scenarios considered include (1) mean shift in x_1 of size δ, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \delta\mathbf{e}$ with $\mathbf{e} = (1, 0, 0, 0)^T$; (2) variance shift in x_1 of size δ, i.e., $\sigma_{11,1} = \delta\sigma_{11,0}$; (3) shift in the correlation between x_1 and x_2 of size ρ, i.e., $\sigma_{12,1} = \sigma_{21,1} = \rho$, where ρ changes from 0.3 to other values. We set $m_0 = 100, \lambda = 0.05$ and $\text{ARL}_0 = 200$.

We first consider the charting performance of all 63 $\mathcal{P}$s for mean and variance shifts in x_1. As Figure 6 shows, if a $\mathcal{P}$ does not have any pair of x_1, such like $\mathcal{P} = \{(x_2, x_3)\}$, we denote it by a black dot curve with cross marks. Otherwise we denote $\mathcal{P}$s by eight different line styles and marks with respect to their different proportions of pairs of x_1, i.e., 1/4, 1/3, 1/2, 2/5, 3/5, 2/3, 3/4, 1. We can see that the charts with no pairs of x_1 cannot detect the shift in x_1 at all, while the other charts can detect the shift to some degree. As a result, we can conclude that, as long

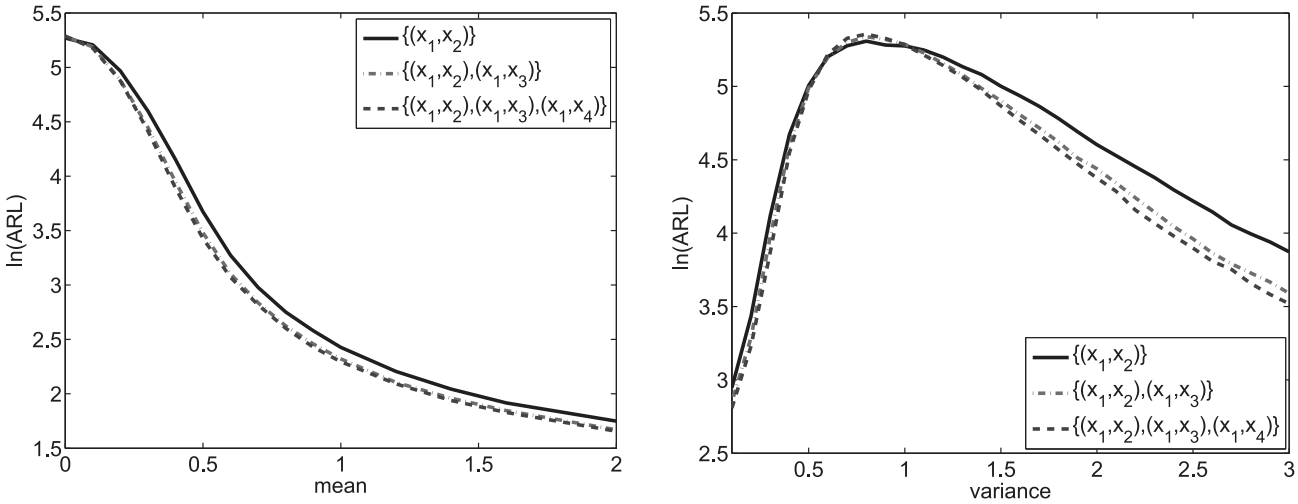


(a) Mean shift (b) Variance shift

FIGURE 6. OC ARL Curves of the 63 Possible $\mathcal{P}$ for x_1 Shifts with the Proportion of Pairs of x_1 as 0, 1/4, 1/3, 1/2, 2/5, 3/5, 2/3, 3/4, 1.

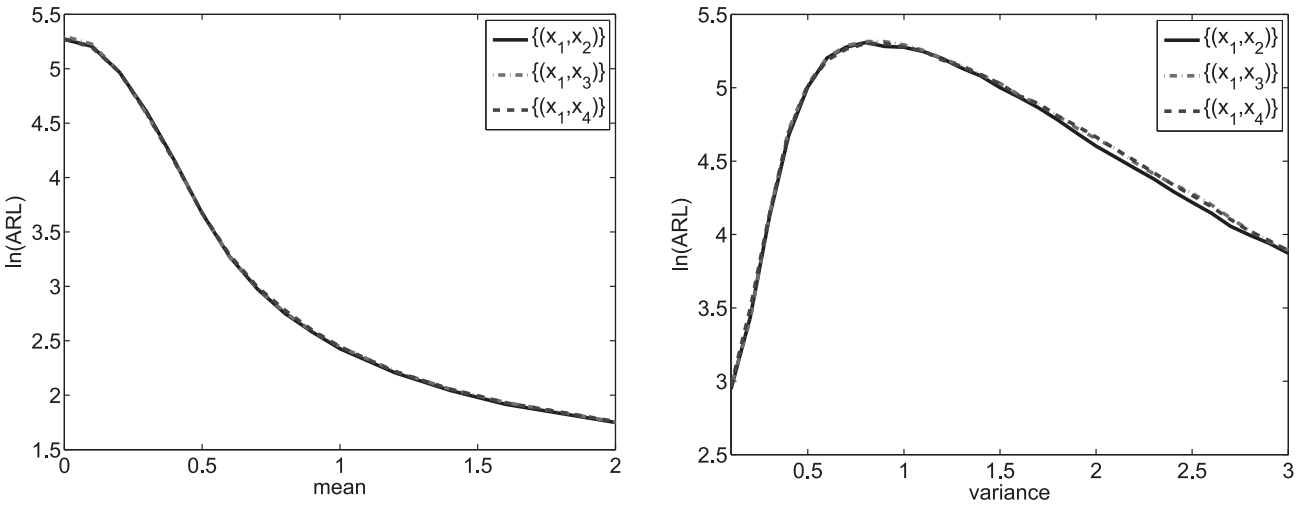
as the chart includes at least one pair of the shifted dimension, the chart has detection power for its mean or variance shift.

Furthermore, as Figure 6 shows, the proportion of pairs of x_1 in $\mathcal{P}$ has influence on the charting performance. Generally, the higher the proportion, the better the detection power is. This is because the lower proportion of pairs of x_1 means more irrelevant pairs involved in the chart. These irrelevant pairs will introduce extra noise to the chart and consequently deteriorate the detection power.



(a) Mean shift (b) Variance shift

FIGURE 7. OC ARL Curves of $\mathcal{P} = \{(x_1, x_2)\}$, $\mathcal{P} = \{(x_1, x_2),(x_1, x_3)\}$, and $\mathcal{P} = \{(x_1, x_2),(x_1, x_3),(x_1, x_4)\}$ for x_1 Shifts.



(a) Mean shift

(b) Variance shift

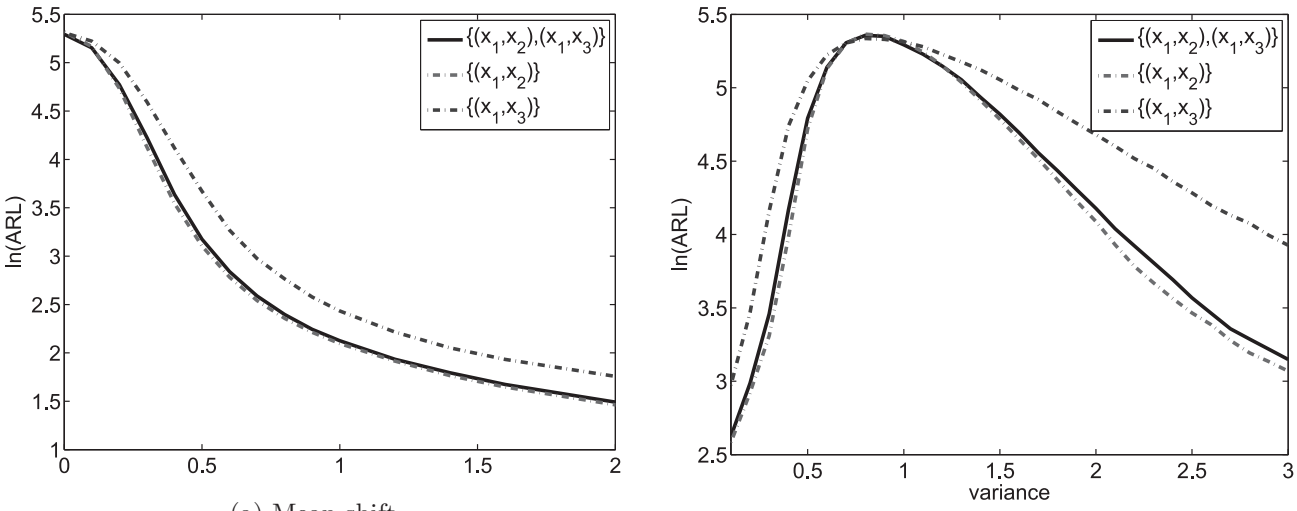
FIGURE 8. OC ARL Curves of $\mathcal{P} = \{(x_1, x_2)\}$, $\mathcal{P} = \{(x_1, x_3)\}$ and $\mathcal{P} = \{(x_1, x_2)\}$ for x_1 Shifts.

The chart detection power for x_1 shifts is not only influenced by the proportion of pairs of x_1, but also by the number of pairs of x_1 involved in $\mathcal{P}$. We demonstrate this point in Figure 7. We compare the performance of $\mathcal{P} = \{(x_1, x_2)\}, \mathcal{P} = \{(x_1, x_2), (x_1, x_3)\}$ and $\mathcal{P} = \{(x_1, x_2), (x_1, x_3), (x_1, x_4)\}$ separately. The proportions of pairs of x_1 in these three charts are all 100%. However, we can see the chart including the most pairs of x_1, i.e., $\mathcal{P} = \{(x_1, x_2), (x_1, x_3), (x_1, x_4)\}$ performs best, followed by $\mathcal{P} = \{(x_1, x_2), (x_1, x_3)\}$, with $\mathcal{P} = \{(x_1, x_2)\}$ the last choice. As a result, we can conclude that, given the same proportion, the higher the number of pairs of the shifted dimension in $\mathcal{P}$, the better the charting performance.

Generally, for $\mathcal{P}$s including different pairs of x_1, the correlation of the included pair of x_1 has little influence on the detection power. We demonstrate this point by comparing the following charts with a unique pair of x_1, i.e., $\mathcal{P} = \{(x_1, x_2)\}, \mathcal{P} = \{(x_1, x_3)\}$ and $\mathcal{P} = \{(x_1, x_4)\}$ separately, with their corresponding correlation coefficients as 0.3, 0.09, and 0.0027. From Figure 8, we can see that they have similar detection power and their curves are not distinguishable. Hence, we can conclude that, for $\mathcal{P}$s including multiple pairs of x_1, these pairs detect the shift in x_1 with equal efficiency. Say, for example, for $\mathcal{P} = \{(x_1, x_2), (x_1, x_3)\}$, these two pairs will contribute equally to the OC signal for x_1 shifts. This brings a lot of convenience for us when selecting the pairs.

For shifts in more than one dimension, the choice of pairs is more complicated. We consider mean or variance shifts in both x_1 and x_2 for $\mathcal{P} = \{(x_1, x_2), (x_1, x_3)\}$, $\mathcal{P} = \{(x_1, x_2)\}$ and $\mathcal{P} = \{(x_1, x_3)\}$. From Figure 9, we can see that generally $\mathcal{P} = \{(x_1, x_2)\}$ performs best, with a minor advantage to $\mathcal{P} = \{(x_1, x_2), (x_1, x_3)\}$. This is because, though the latter has more pairs of x_1, its proportion of pairs of x_2 is lower. The noise introduced by pair (x_1, x_3) is bigger than the detection power contributed by it. From this point, we can see that introducing more pairs is not always better.

Now we consider the charting performance of different $\mathcal{P}$ for correlation shifts, and deliver our conclusions from the following simulation results. We consider the correlation between x_1 and x_2 changing from -0.9 to 0.9 while the other correlations remain unchanged. As Figure 10 shows, we denote $\mathcal{P}$s that do not have the pair (x_1, x_2) by black dot curves with cross marks. Otherwise we denote $\mathcal{P}$s by six different line styles and marks with respect to their different proportions of the pair (x_1, x_2), i.e., 1/6, 1/5, 1/4, 1/3, 1/2, 1. We can see that, as long as the pair (x_1, x_2) is included in $\mathcal{P}$, the chart can detect its change to some degree. Otherwise, the chart almost has no detection power. This demonstrates that it is only the pair (x_1, x_2) that contributes to the OC signal. Furthermore, from Figure 10, we can see that the higher the proportion, the better the performance. This means that the detection power is interfered with the other unrelated pairs in $\mathcal{P}$. The



(a) Mean shift

(b) Variance shift

FIGURE 9. OC ARL Curves of $\mathcal{P}=\{(x_1, x_2),(x_1, x_3)\}$, $\mathcal{P}=\{(x_1, x_2)\}$, and $\mathcal{P}=\{(x_1, x_3)\}$ for Shifts in Both x_1 and x_2.

more unrelated pairs included, the more noise will be added to the chart and consequently the detection power will be worse.

From the discussion above, we can see that including all the $p(p-1)/2$ pairs in $\mathcal{P}$ brings about the most omnipotent detection performance. In this way, the chart can detect any marginal mean or variance shift, as well as any shift in the correlation structure, though for detecting a certain shift pattern, this charting performance is not the best compared with other pair selection mechanisms designed particularly for this shift pattern. For example, the detection power of $\mathcal{P} = \{(x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4), (x_3, x_4)\}$ for x_1 shifts only ranks in the middle of the total 63 charts, as Figure 11 shows. Here we also summarize some guidelines on the optimal pair selection targeting different shift patterns.

- For mean and variance shifts
  - — If no prior information is known, the bottom line is to include all the p dimensions in $\mathcal{P}$, which corresponds to $\lceil p/2 \rceil$ pairs. In this way, the chart can detect any marginal mean or variance shift. Furthermore, how to pair these p dimensions has no influence on the detection power, which brings convenience for practitioners.
  - — If shifts in a certain dimension are prioritized, including all pairs of this dimension and pruning away all irrelevant pairs bring about the best detection power.
  - — If shifts in more than one dimension are prioritized, only including these dimensions in $\mathcal{P}$ by combining them as pairs and trying the best to prune away pairs of the other irrelevant dimensions bring about the best detection power.

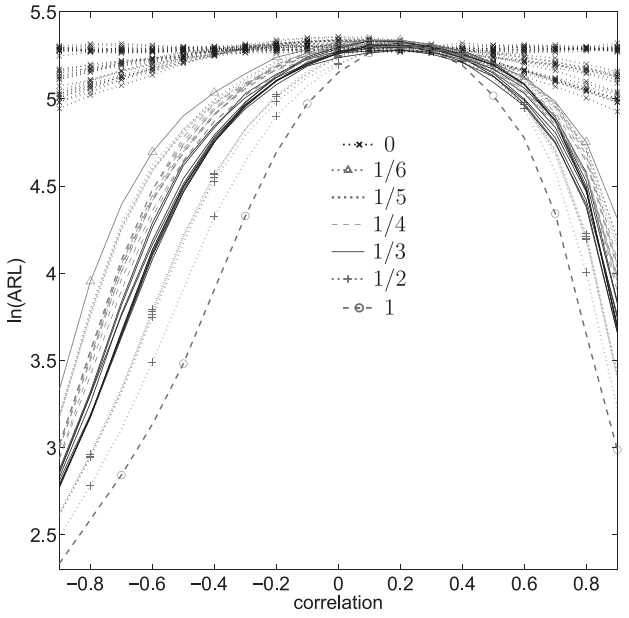
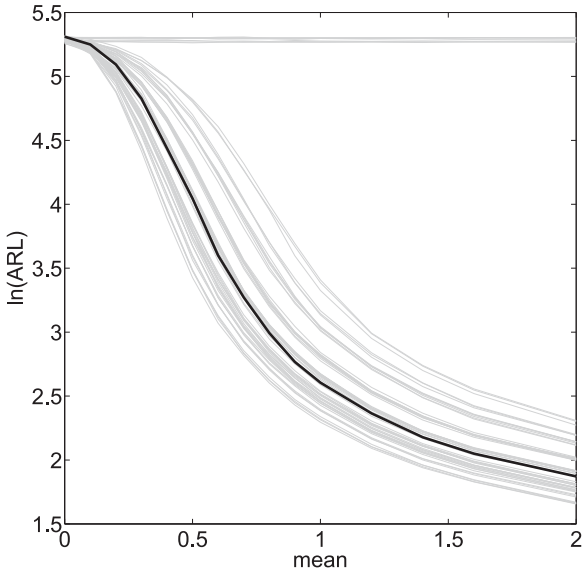
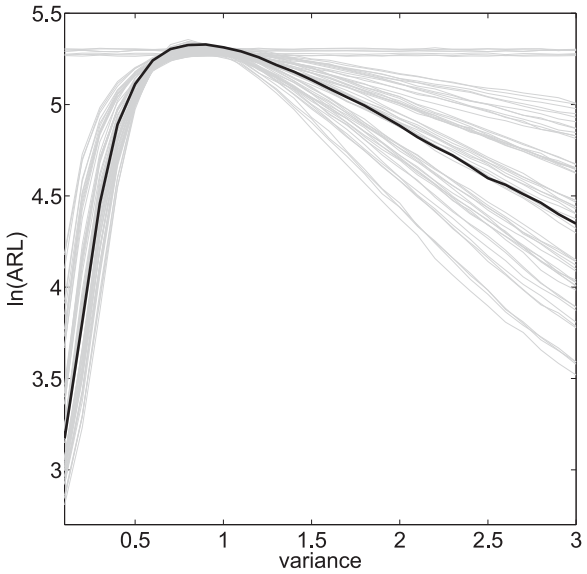


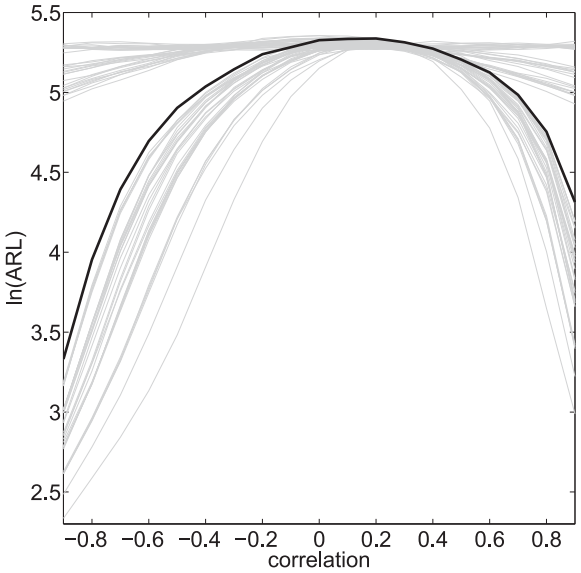
FIGURE 10. OC ARL Curves of the 63 Possible $\mathcal{P}$ for Correlation Shifts with the Proportion of Pairs (x_1, x_2) as 0, 1/6, 1/5, 1/4, 1/3, 1/2, 1.



(a) Mean shift



(b) Variance shift



(c) Correlation shift

FIGURE 11. OC ARL Curves of $\mathcal{P} = \{(x_1, x_2), (x_1, x_3), (x_2, x_3), (x_2, x_4), (x_3, x_4)\}$ for x_1 Shifts.

- For correlation shifts
  - — If we want to detect any change in the correlation structure, then involving all the $p(p-1)/2$ pairs is the most parsimonious.
  - — If some correlations are of more interest than others based on some prior information, then only involving the pairs representing these interesting correlations and pruning away any other irrelevant pair bring about the best charting performance.

6. A Real-Data Application

We use a real dataset from a semiconductor-manufacturing process to illustrate the application of DFMGoF. The dataset, which is publicly available in the UC Irvine Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/SECOM), contains a total of 1567 samples from a semiconductor manufacturing process. Each sample is a vector of 591 dimensions, consisting of 591 continuous measurements during the process in producing each batch. Among

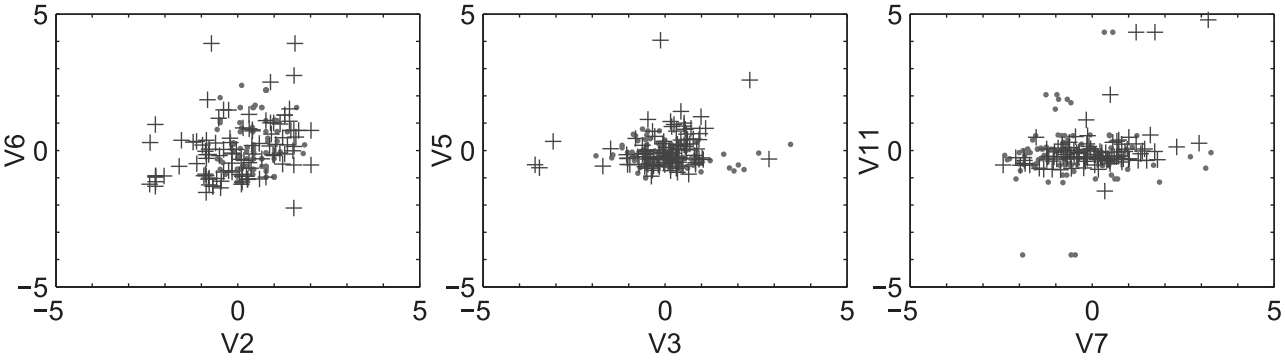


FIGURE 12. The Scatter Plots of Three Pairs of Variables. The red dots represent the IC samples and the blue cross represents the OC samples.

them, 1463 samples are classified as conforming ones (IC samples), while the remaining 104 samples are classified as nonconforming ones (OC samples). The goal of this section is to use this dataset to illustrate the on-line process quality control using DFMGoF.

As a preprocessing step, we remove 117 variables with constant values from the 591 variables in all the 1567 samples. In addition, we impute the missing data by replacing the missing values with the mean of the observed values from that variable because the fraction of missing values in the dataset is trivial. Among the remaining 474 variables, we find that 12 variables, namely $\{X_3, X_{15}, X_{38}, X_{99}, X_{126}, X_{146}, X_{148}, X_{264}, X_{348}, X_{350}, X_{374}, X_{383}\}$, have no significant difference in their mean values between the IC and OC samples. As a result, for these 12 variables, any MSPC chart designed merely to monitor the process mean vector might be ineffective. Subsequently, we focus on these 12 variables to demonstrate the advantage of DFMGoF. The variables are denoted as $\{V_1, \ldots, V_{12}\}$ for notation simplicity. Figure 12 compares some pairwise scatter plots of these 12 variables of 100 IC samples with those of 60 OC samples (these samples are randomly selected from the dataset because plotting all the IC and OC samples makes the points unidentifiable). It clearly reveals that the mean values of the variables do not differ much between IC and OC samples, but their variance and correlation structure change to some degree. Furthermore, the normal QQ plots (Figure 13) show that these variables do not have marginal normal distributions, indicating distribution-free charts might perform more robustly for this dataset.

To demonstrate the application of DFMGoF, we monitor the observations sequentially: we randomly draw $m_0 = 500$ IC observations without replace-

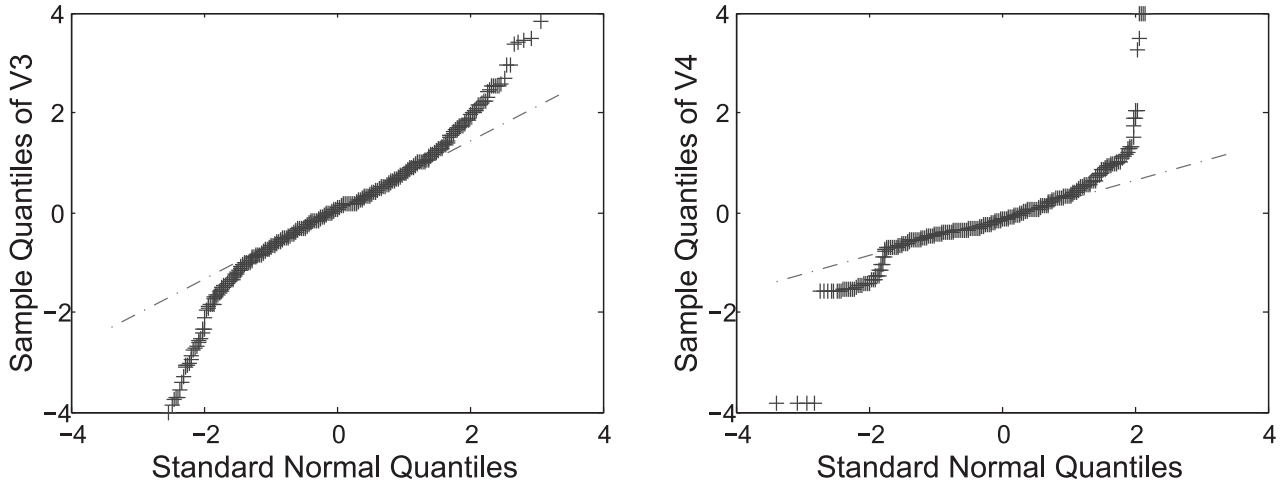


FIGURE 13. The Normal Q-Q Plots for V3 and V4.

TABLE 5. The ARLs of Different Charts for Monitoring the Semiconductor Production Process with $p = 12$, $m_0 = 500$, $\lambda = 0.05$, and $\tau = 60$. Numbers in parentheses are SDRL values

	DFMGoF	RTC	ChangePt	SSEWMAC
IC	199 (181)	70.1 (11.5)	52.5 (55.5)	21.0 (4.54)
OC	38.5 (30.3)	83.6 (85.0)	92.3 (20.8)	40.0 (20.1)

ment as reference samples from the 1463 IC samples, then we draw the subsequent observations sequentially as on-line testing samples. We set $\text{ARL}_0 = 200$, $\lambda = 0.05$, and $w = 58$. We first evaluate the IC performance of the chart by drawing the testing samples independently from the remaining 963 IC samples. In each replicate, the chart runs until an OC signal is generated and the corresponding run length is recorded. The procedure is repeated 2000 times and the estimated ARL_0s of DFMGoF and the other three charts are reported in Table 5. It shows that, except DFMGoF, the other three charts have excessive false alarms after short runs, indicating their unacceptable IC performance. This is because the nonnormality of the data makes the normal assumption of SSEWMAC and ChangePt invalid, and the reference sample size $m_0 = 500$ is insufficient for RTC to get a proper control limit. In contrast, DFMGoF demonstrates superior performance in this case.

Next we compare the OC performance of DFMGoF with the other charts. Similar to Section 4.2, we adjust the control limits of the other charts to make their ARL_0s close to 200 through resampling from the IC samples. We choose $\xi = 60$, meaning that the first 60 on-line samples are drawn from the remaining IC samples and the subsequent on-line samples are drawn from the OC samples. The procedure is repeated 2000 times. Then the OC ARLs are compared in Table 5, where DFMGoF has the best performance among these four charts with OC ARL= 38.5 and SDRL= 30.31. For illustration, Figure 14 shows one replicate of DFMGoF, where the blue solid dot represents the monitoring statistic Z_n and the red circle represents the corresponding control limit $H_n(\alpha, \mathcal{F}_{m_0+n})$ for every step. We can see that DFMGoF has a quick response to the process shift with a timely increase in $Z_n(w, \lambda)$ after the ξth sample. Finally, Z_n exceeds the control limit at $n = 67$, signalling an OC alarm with run length 7.

7. Concluding Remarks

Though nonparametric MSPC has been extensively studied in the literature, the challenges associated with designing distribution-free control schemes for monitoring both the mean vector and the covariance matrix simultaneously are yet to be well addressed. This paper presents a new MSPC method to fill in this gap. Specifically, we propose a new chart

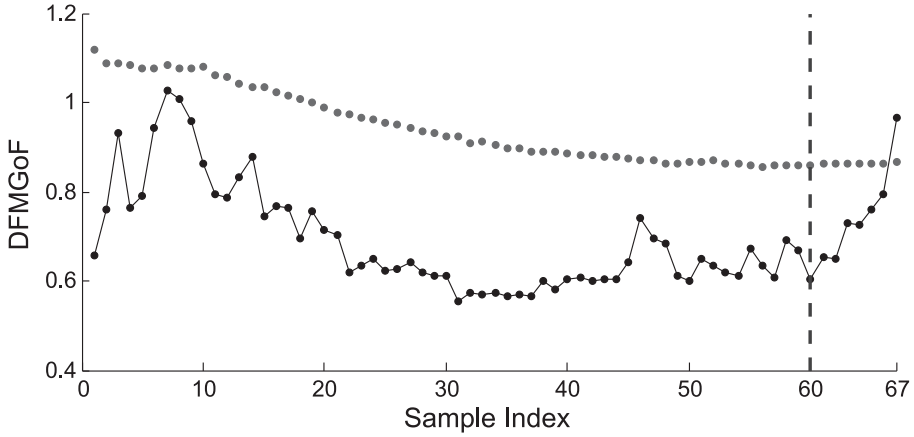


FIGURE 14. One replicate of DFMGoF for Monitoring the Semiconductor Production Process when Changes Occur at $\tau = 60$. The black dot represents the Z_n calculated at each step and the red dot represents the control limit $H_n(\alpha, \mathcal{F}_{m_0+1})$ at each step.

based on a multivariate GoF test to detect general distributional changes. We also propose to use *data-dependent* control limits to achieve the distribution-free property regardless of the type or dimension of the IC distribution. A computational algorithm based on the permutation principle is proposed to find the control limits on line along with the charting statistics. Numerical studies show that the proposed chart has exactly the distribution-free property and proves to be robust in detecting general process changes. Its application in a real dataset is also illustrated through a case study.

Along this research direction, there are also some valuable extensions. First, when the dimension is large, the current calculation procedure of the control limits requires a large amount of computation time. It is of interest to design a more efficient computation procedure taking advantage of modern computing techniques, e.g., distributed computing or importance sampling, to reduce the execution time for on-line applications. Second, extensions of the proposed chart to autocorrelated processes are also worth studying. In these cases, the charting statistics should also be able to detect changes in the process autocorrelation structure. Furthermore, when calculating the control limits, the permutation procedure should keep the autocorrelation structure as well. Then more advanced algorithms, such as the block bootstrap method, need to be considered.

Appendices

A.1. Computational Complexity of DFMGoF

Further computational issues deserve our consideration for fast implementation of DMFGoF. Computing the monitoring statistic Z_n requires at most $O(qn)$ computation time (i.e., comparing (X_{nj}, X_{nk}) with (X_{ij}, X_{ik}) for $i = -m_0 + 1, \ldots, n-1$ for every pair (j,k)). But it is computationally expensive to compute the control limit $H_n(\alpha, \mathcal{F}_{m_0+n})$ by permutation, as Algorithm 1 shows below. Every permutation $\mathcal{S}_n^v = \{\mathbf{X}_{v_{-m_0+1}}, \ldots, \mathbf{X}_{v_0}, \mathbf{X}_{v_1}, \ldots, \mathbf{X}_{v_n}\}$ needs $O(n)$ computation time. To calculate Z_n^v from $\mathcal{S}_n^v$, the new probabilities $\{\hat{P}_{0,jk}^{v_n,r}(\mathbf{X}_{v_i,jk}), -m_0 + 1 \le i \le n\}$ can be obtained directly from $\{\hat{P}_{0,jk}^{n,r}(\mathbf{X}_{i,jk}), -m_0 + 1 \le i \le n\}$ (by the same permutation sequence). However, $\{\hat{P}_{jk}^{v_n,r}(\mathbf{X}_{v_i,jk}), \max\{1, n-w+1\} \le i \le n\}$ are different from $\{\hat{P}_{jk}^{n,r}(\mathbf{X}_{i,jk}), \max\{1, n-w+1\} \le i \le n\}$, because the former are based on the ranks of $\mathbf{X}_{v_i}(\max\{1, n-w+1\} \le i \le n)$ in $\mathcal{S}_n^v$ rather than those in $\mathcal{S}_n$. Therefore they have to be recalculated with additional $O(qw^2)$ computation time. Z_t^v for $\max\{1, n-w+1\} \le t < n$ can be calcu-

ALGORITHM 1. Permutation Procedure to Find Control Limits

Define flag = 1
Draw a permutation $\mathcal{S}_n^v = \{\mathbf{X}_{v_{-m_0+1}}, \ldots, \mathbf{X}_{v_0}, \mathbf{X}_{v_1}, \ldots, \mathbf{X}_{v_n}\}$
Get $\{\hat{P}_{0,jk}^{v_n,r}(\mathbf{X}_{v_{-m_0+1}:v_n,jk})\}$ directly from $\{\hat{P}_{0,jk}^{n,r}(\mathbf{X}_{-m_0+1:n,jk})\}$
Compute $\{\hat{P}_{jk}^{v_n,r}(\mathbf{X}_{v_i,jk}), \max\{n-w+1, 1\} \le i \le n\}$
Calculate Z_n^v

for $t = n-1$ to $\max\{n-w+1, 1\}$ **do**
    Update $\{\hat{P}_{0,jk}^{v_t,r}(\mathbf{X}_{v_i,jk})\}$ from $\{\hat{P}_{0,jk}^{v_{t+1},r}(\mathbf{X}_{v_i,jk})\}$ for $\max\{t-w+1, 1\} \le i \le t$
    Update $\{\hat{P}_{jk}^{v_t,r}(\mathbf{X}_{v_i,jk})\}$ from $\{\hat{P}_{jk}^{v_{t+1},r}(\mathbf{X}_{v_i,jk})\}$ for $\max\{t-w+1, 1\} \le i \le t$
    Calculate Z_t^v

    if $Z_t^v > H_t(\alpha, \mathcal{F}_{m_0+t})$ **then**
        Discard current permutation, set flag = 0
    end if
end for
Calculate Z_n^v

if flag = 1 **then**
    return current permutation Z_n^v
end if

lated for every time point t in a recursive manner and only $O(qw)$ computation time is needed in each update. Then the total computational complexity is $O(qbw^2 + qn)$, linear in n, b, and p. Such computational complexity is implementable with the powerful computing resources nowadays. With the help of parallel computing, the computation time would be reduced further and therefore it is easy to apply DFM-GoF for high-dimensional process monitoring.

A.2. The Multivariate t and Gamma Distributions

The multivariate t distribution used in this paper is based on the proposal by Johnson and Kotz (1972). It is defined as follows. Let $\mathbf{X}$ follow the p-dimensional multivariate normal distribution $N_p(\mathbf{0}, \mathbf{\Sigma})$ with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. Let Z follow an independent chi-square distribution with ζ degrees of freedom. Then $\mathbf{T} = \boldsymbol{\mu} + \mathbf{X}/\sqrt{Z/\zeta}$ follows a p-dimensional t distribution with ζ degrees of freedom with noncentrality parameter $\boldsymbol{\mu}$. In addition, we have

$$\mathrm{E}(\mathbf{T}) = \boldsymbol{\mu}, \qquad \mathrm{Var}(\mathbf{T}) = \frac{\zeta}{\zeta - 2}\mathbf{\Sigma}.$$

For a more detailed discussion on the multivariate t distribution, please refer to Anderson (1984).

The multivariate gamma distribution considered here is first proposed by Krishnamoorthy and Parthasarathy (1951). It could be generated as follows. Let $\mathbf{X}$ be a matrix of dimension $\zeta \times p$. Each row of $\mathbf{X}$ follows a p-dimensional multivariate normal distribution $N_p(\mathbf{0}, \mathbf{\Sigma})$ independently. Then

$$\mathbf{G} \equiv \frac{1}{2}\mathrm{diag}(\mathbf{X}^T\mathbf{X}) + \boldsymbol{\mu}$$

is defined to follow the p-dimensional gamma distribution with ζ degrees of freedom, i.e., $\mathrm{Gam}_{p,\zeta}$, with noncentrality parameter $\boldsymbol{\mu}$. In addition, denoting $\mathbf{G} = (G_1, \ldots, G_p)$, then $G_i = \sum_{j=1}^{\zeta} X_{ji}^2/2 + \mu_i (1 \le i \le p)$ follows a gamma distribution with scale parameter $\theta_i = \sigma_{ii}$ and shape parameter $\beta = \zeta/2$. Therefore, the density of the marginal distribution can be expressed as

$$f(G_i = x) = \frac{(x - \mu_i)^{\zeta/2-1} \exp\left[-\frac{x-\mu_i}{\sigma_{ii}}\right]}{(\sigma_{ii})^{\zeta/2}\Gamma(\zeta/2)}, \quad x \ge \mu_i. \tag{A.1}$$

Its expectation and variance follow the properties of a gamma distribution as $\mathrm{E}(G_i) = \zeta\sigma_{ii}/2 + \mu_i$ and $\mathrm{var}(G_i) = \zeta\sigma_{ii}^2/2$. In addition, we can find the covariance between G_i and G_j by definition as

$$\begin{aligned}\mathrm{Cov}(G_i, G_k) &= \mathrm{Cov}\left(\sum_{j=1}^{\zeta} X_{ji}^2/2 + \mu_i, \sum_{j=1}^{\zeta} X_{jk}^2/2 + \mu_k\right) \\ &= \frac{\zeta}{4}\mathrm{Cov}(X_{ji}^2, X_{ik}^2) = \frac{\zeta}{2}\sigma_{ik}^2.\end{aligned}$$

Thus, the correlation between any two dimensions of $\mathrm{Gam}_{p,\zeta}$ is nonnegative. Furthermore, we could see that the mean vector of $\mathrm{Gam}_{p,\zeta}$ depends on both the noncentrality parameter $\boldsymbol{\mu}$ and the variance of the underlying multivariate normal distribution, i.e., $\mathrm{diag}(\mathbf{\Sigma})$. Therefore, the change in the variance of $\mathrm{Gam}_{p,\zeta}$ will lead to the change in its mean vector.

Acknowledgments

The authors thank the editor and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. Nan Chen is partially supported by Singapore AcRF Tier 1 funding #R-266-000-078-112, National Research Foundation Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE). He also expresses thanks for the partial support by the NNSF of China under Grant 71471096. Changliang Zou is supported by the NNSF of China under Grants 11431006, 11131002, 11371202 and the Foundation for the Author of National Excellent Doctoral Dissertation of China 201232.

References

Alt, F. B. (1985). "Multivariate Quality Control". *Encyclopedia of the Statistical Science*, pp. 111–122.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*. New York, NY: Wiley.

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

Chen, G.; Cheng, S. W.; and Xie, H. (2005). "A New Multivariate Control Chart for Monitoring Both Location and Dispersion". *Communications in Statistics—Simulation and Computation®* 34, pp. 203–217.

Chen, N.; Zi, X.; and Zou, C. (2015), "A Distribution-Free Multivariate Control Chart". *Technometrics*, in press, doi: 10.1080/00401706.2015.1049750.

Conover, W. J. (1999), *Practical Nonparametric Statistics*, 3rd edition. New York, NY: Wiley.

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

Deng, H.; Runger, G.; and Tuv, E. (2012) "System Monitoring with Real-Time Contrasts". *Journal of Quality Technology*, 44.

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

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

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

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

Hawkins, D. M. and Maboudou-Tchao, E. M. (2008). "Multivariate Exponentially Weighted Moving Covariance Matrix". *Technometrics* 50, pp. 155–166.

He, Q. P.; Hawkins, D. M.; Maboudou-Tchao, E. M.; and Wang, J. (2007). "Fault Detection Using the K-Nearest Neighbor Rule for Semiconductor Manufacturing Processes". *IEEE Transactions on Semiconductor Manufacturing* 20, pp. 345–354.

Healy, J. D. (1987), "A Note on Multivariate Cusum Procedures". *Technometrics* 29, pp. 409–412.

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

Hwang, W.; Runger, G.; and Tuv, E. (2007). "Multivariate Statistical Process Control with Artificial Contrasts". *IIE transactions* 39, pp. 659–669.

Johnson, N. and Kotz, S. (1972), *Distributions in Statistics: Continuous Multivariate Distributions*. In Wiley series "Probability and Mathematical Statistics: Applied Probability and Statistics". New York, NY: Wiley.

Jones, L. A.; Champ, C. W.; and Rigdon, S. E. (2001). "The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters". *Technometrics* 43.

Khoo, M. B. (2004). "A New Bivariate Control Chart to Monitor the Multivariate Process Mean and Variance Simultaneously". *Quality Engineering* 17, pp. 109–118.

Krishnamoorthy, A. and Parthasarathy, M. (1951). "A Multivariate Gamma-Type Distribution". *The Annals of Mathematical Statistics*, pp. 549–557.

Lai, T. L. (1995). "Sequential Changepoint Detection in Quality Control and Dynamical Systems". *Journal of the Royal Statistical Society, Series B (Methodological)* 57, pp. 613–658.

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

Liu, R. Y.; Singh, K.; and Teng, J. H. (2004). "DDMA-Charts: Nonparametric Multivariate Moving Average Control Charts Based on Data Depth". *Allgemeines Statistisches Archiv* 88, pp. 235–258.

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

Maboudou-Tchao, E. M. and Hawkins, D. M. (2011) "Self-Starting Multivariate Control Charts for Location and Scale". *Journal of Quality Technology* 43, pp. 113–126.

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

Montgomery, D. and Wadsworth, H. (1972). "Some Techniques for Multivariate Quality Control Applications". In *ASQC Technical Conference Transactions*, Washington, DC, pp. 427–435.

Montgomery, D. C. (1991). *Introduction to Statistical Quality Control*. New York, NY: Wiley.

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

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

Qiu, P. and Hawkins, D. (2001). "A Rank-Based Multivariate Cusum Procedure". *Technometrics* 43, pp. 120–132.

Reynolds, M. R. and Cho, G.-Y. (2006). "Multivariate Control Charts for Monitoring the Mean Vector and Covariance Matrix". *Journal of Quality Technology* 38, pp. 230–253.

Runger, G. C. and Prabhu, S. S. (1996). "A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart". *Journal of the American Statistical Association* 91, pp. 1701–1706.

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

Sukchotrat, T.; Kim, S. B.; and Tsung, F. (2009). "One-Class Classification-Based Control Charts for Multivariate Process Monitoring". *IIE Transactions* 42, pp. 107–120.

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

Woodall, W. H. (2000). "Controversies and Contradictions in Statistical Process Control". *Journal of Quality Technology* 32, pp. 341–350.

Yeh, A. B.; Huwang, L.; and Wu, C.-W. (2005). "A Multivariate EWMA Control Chart for Monitoring Process Variability with Individual Observations". *IIE Transactions* 37, pp. 1023–1035.

Yeh, A. B.; Huwang, L.; and Wu, Y.-F. (2004). "A Likelihood-Ratio-Based EWMA Control Chart for Monitoring Variability of Multivariate Normal Processes". *IIE Transactions* 36, pp. 865–879.

Yeh, A. B. and Lin, D. K. (2002) "A New Variables Control Chart for Simultaneously Monitoring Multivariate Process Mean and Variability". *International Journal of Reliability, Quality and Safety Engineering* 9, pp. 41–59.

Zamba, K. and Hawkins, D. M. (2009), "A Multivariate Change-Point Model for Change in Mean Vector and/or Covariance Structure". *Journal of Quality Technology* 41, pp. 285–303.

Zhang, J. (2002), "Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64, pp. 281–294.

Zhang, J.; Li, Z.; and Wang, Z. (2010). "A Multivariate Control Chart for Simultaneously Monitoring Process Mean and Variability". *Computational Statistics & Data Analysis* 54, pp. 2244–2252.

Zou, C. and Tsung, F. (2010). "Likelihood Ratio-Based Distribution-Free EWMA Control Charts". *Journal of Quality Technology* 42, pp. 174.

Zou, C. and Tsung, F. (2011). "A Multivariate Sign EWMA Control Chart". *Technometrics* 53, pp. 84–97.

Zou, C.; Wang, Z.; and Tsung, F. (2012). "A Spatial Rank-Based Multivariate EWMA Control Chart". *Naval Research Logistics (NRL)* 59, pp. 91–110.