A Multivariate Sign EWMA Control Chart

Changliang Zou and Fugee Tsung*

Hong Kong University of Science and Technology

Clear Water Bay, Kowloon, Hong Kong

Abstract

Nonparametric control charts are useful in statistical process control (SPC) when there is a lack of or limited knowledge about the underlying process distribution, especially when the process measurement is multivariate. This paper develops a new multivariate SPC methodology for monitoring location parameters. It is based on adapting a powerful multivariate sign test proposed by Randles (2000) to on-line sequential monitoring. The weighted version of the sign test is used to formulate the charting statistic by incorporating the exponentially weighted moving average control (EWMA) scheme, which results in a nonparametric counterpart of the classical multivariate EWMA (MEWMA). It is affine-invariant and has a strictly distribution-free property over a broad class of population models. That is, the in-control (IC) run length distribution can attain (or is always very close to) the nominal one when using the same control limit designed for a multivariate normal distribution. Moreover, when the process distribution comes from the elliptical direction class, the IC average run length can be calculated via a one-dimensional Markov chain model. This control chart possesses some other favorable features: its computation speed is fast with a similar computation effort to the MEWMA chart; it is easy to implement because only the multivariate median and the associated transformation matrix need to be specified (estimated) from the historical data before monitoring; it is also very efficient in detecting process shifts, particularly small or moderate shifts when the process distribution is heavy-tailed or skewed. Two real-data examples from manufacturing show that it performs quite well in applications.

Keywords: Affine-Invariant; Distribution-Free; Nonparametric Procedure; Multivariate Median; MEWMA; Robustness; Statistical Process Control

*Corresponding author. Email: season@ust.hk.

1 Introduction

In modern quality control, it is common to monitor several quality characteristics of a process simultaneously (Stoumbos et al. 2000). This is called multivariate statistical process control (MSPC) in the literature and it is the focus of this paper. One of the tasks of MSPC is to detect the change in a multivariate process location parameter $\boldsymbol{\theta}$ (mean, median or some percentile of the distribution) as quickly as possible. To be more specific, it is usually assumed that there are m_0 independent and identically distributed (i.i.d.) historical (reference) observations, $\boldsymbol{x}_{-m_0+1}, \ldots, \boldsymbol{x}_0 \in \mathbb{R}^p$, for some integer, $p \geq 1$, and the ith future observation, $\boldsymbol{x}_i$, is collected over time from the following multivariate change-point model

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

where τ is the unknown change point and $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$.

Methods for accomplishing the monitoring task are usually based on the following quadratic formulation of the test statistics:

$$(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_0)' \widehat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_0), \tag{2}$$

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. It is often called a Shewhart χ^2 chart when we use exact $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ instead of $\widehat{\boldsymbol{\mu}}_0$ and $\widehat{\boldsymbol{\Sigma}}_0$. In the literature, to accumulate information from past observations, many MSPC control charts are constructed in two steps. First, a sequence of multivariate vectors is constructed in the framework of a cumulative sum (CUSUM) or an exponentially weighted moving average (EWMA). Then, the charting statistic takes the quadratic form of the multivariate vectors in a similar way to (2) (cf., for instance, Healy 1987; Croisier 1988; Pignatiello and Runger 1990; Lowry et al. 1992; Runger and Prabhu 1996; and Zamba and Hawkins 2006). In particular, the multivariate EWMA (MEWMA) chart, proposed by Lowry et al. (1992), is powerful in detecting small or moderate sustained shifts in $\boldsymbol{\mu}$ with small or moderate weighting parameters. Its charting statistic is defined by

$$T_i^2 = \frac{2 - \lambda}{\lambda} \boldsymbol{z}_i' \boldsymbol{\Sigma}_0^{-1} \boldsymbol{z}_i, \tag{3}$$

where $\lambda \in (0, 1]$ is a weighting parameter and $\boldsymbol{z}_i$ is a vector operating in a recursive form,

$$\boldsymbol{z}_i = \lambda(\boldsymbol{x}_i - \boldsymbol{\mu}_0) + (1 - \lambda)\boldsymbol{z}_{i-1}. \tag{4}$$

The above mentioned MSPC research is mostly based on a fundamental assumption that the process data have multinormal distributions. However, it is well recognized that, in many applications, the underlying process distribution is unknown and not multinormal, so that statistical properties of commonly used charts, designed to perform best under the normal distribution, could potentially be (highly) affected. The problem of degraded statistical performance due to the non-normality is severe with small samples, particularly individual observation cases (c.f., Montgomery 2005) since the cental limit theorem is no longer (approximately) valid. 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 on this topic was presented by Chakraborti et al. (2001). The need for robust multivariate SPC has been noted in a number of articles, see, e.g., Woodall and Montgomery (1999) and references therein. Some effort has been devoted to this problem, such as the control schemes based on data-depth (see, e.g., Liu 1995; Liu et al. 2004) or support vector machines (Sun and Tsung 2003). However, Stoumbos and Sullivan (2002) argued that multivariate nonparametric control charts "are less powerful, more computationally intensive, and generally do not apply to skewed distributions". See Stoumbos and Jones (2000) for a nice analysis of the method presented by Liu (1995). Alternatively, Qiu and Hawkins (2001; 2003) suggested a computationally trivial nonparametric multivariate CUSUM procedure based on the antiranks of the measurement components. Both papers discuss only the case when the IC distribution is assumed to be known.

It needs to be emphasized that although the closed-forms of cumulative distribution functions (c.d.f.) or density functions may not be available, the statement that the IC distribution is assumed to be known is essentially equivalent to saying that m_0 is sufficiently large because we can always use various estimation approaches, such as empirical distribution functions or multivariate kernel density estimations, to obtain corresponding consistent estimators. Even the Shewhart χ^2 chart can then be regarded as "*distribution-free*". That is, the IC run length distribution for non-normal processes could be designed to achieve the nominal one by means of simulations through resampling from the IC distribution or from the m_0 IC historical samples directly. However, how large m_0 should be depends on the dimension p and it is always difficult to estimate a high-dimensional distribution because of the "curse of dimensionality" (Eaton 1983). Recently, Qiu (2008) proposed a distribution-free multivariate CUSUM procedure based on log-linear modeling, which presents us a new methodology for estimating the multivariate IC distribution from an IC reference dataset.

Stoumbos and Sullivan (2002) recommended that the MEWMA chart should be more appealing than multivariate nonparametric schemes because MEWMA charts can be quite *robust* in the sense that the IC run length distribution for a continuous non-normal process is quite close to the distribution for a multivariate normal process with the same control limit if the weighting parameter, λ, is small. With a large number of observations and a small smoothing parameter, a central limit theorem would ensure that the accumulation vector has approximately a multinormal distribution, which ensures robustness. Note that only $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ need to be estimated in MEWMA rather than the entire distribution, which relaxes the requirement of m_0. However, how small λ should be relies on the deviation of the actual measurement distribution from the multinormal distribution, which is difficult to measure in practice. Also, when λ is too small, the corresponding procedure would not be sensitive to relatively large shifts.

This paper develops a new multivariate SPC methodology for monitoring location parameters. This methodology adapts a powerful multivariate sign test proposed by Randles (2000) to on-line sequential monitoring by incorporating the EWMA scheme, which results in a nonparametric counterpart of the MEWMA chart. It is affine-invariant and it has an exact distribution-free property over a broad class of population models in the sense that the IC run length distribution can attain (or is always very close to) the nominal one when using the same control limit designed for a multinormal distribution. Moreover, when the process distribution comes from the elliptical direction class (including the multinormal distribution), the IC average run length (ARL) can be calculated via a one-dimensional Markov chain model. In addition, this control chart possesses some other favorable features: unlike some other nonparametric schemes such as data-depth-based charts, its computation speed is fast with a similar computation effort to the MEWMA chart. It is easy to implement because only the multivariate median and the transformation matrix need to be specified from the reference dataset before monitoring. It is also very efficient in detecting process shifts, especially for small or moderate shifts when the process distribution is heavy-tailed or skewed. The remainder of this paper is organized as follows: our proposed methodology is described in detail in Section 2. Its numerical performance is thoroughly investigated in Section 3. In Section 4, we demonstrate the method using two real-data examples from manufacturing industries. Several remarks draw the paper to its conclusion in Section 5. The IC ARL calibrations of MSEWMA via a Markov chain model are provided in the Appendix. Some other technical details, including proofs of some propositions, are provided in another appendix, which is available online as supplementary materials.

2 Methodology

Our proposed methodology is described in three parts. In Section 2.1, a brief introduction to Randles's (2000) multivariate sign test is presented. In Section 2.2, a new multivariate nonparametric EWMA control chart combined with multivariate sign test is derived. Its control limits, practical guidelines regarding its design and computational issues are addressed in Section 2.3.

2.1 A Brief Review of Multivariate Sign Test

Recall model (1) and associated notation. In what follows, we elaborate on the individual observation model. The extension to the group case is presented at the end of Section 2.2. The monitoring problem (1) is closely related to nonparametric statistical tests of hypotheses for the one-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 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. from $F(\boldsymbol{x}-\boldsymbol{\theta})$, where $F(\cdot)$ represents a continuous p-dimensional distribution "located" at the vector $\boldsymbol{\theta}$. We want to test the null hypothesis, H_0, that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ against H_1 that $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Without loss of generality, we assume that $\boldsymbol{\theta}_0 = \boldsymbol{0}$. Otherwise, we substitute $\boldsymbol{x}_i - \boldsymbol{\theta}_0$ in place of $\boldsymbol{x}_i$. In creating tests for this problem, different levels of assumption have been proposed for the distribution of the $\boldsymbol{x}_i$'s. The classical parametric test, Hotelling's T^2, rejects

H_0 if $T^2 = n\bar{\boldsymbol{x}}'\mathbf{S}^{-1}\bar{\boldsymbol{x}}$ is large, where $\bar{\boldsymbol{x}}$ and $\mathbf{S}$ are the sample mean vector and sample covariance matrix, respectively. In the nonparametric setting, many efforts have been devoted to this problem in the literature, such as Randles (1989), Chakraborty et al. (1998), etc. A nice overview on this topic and related references can be found in Oja (1999) and Oja and Randles (2004). Specially, Randles (2000) develops a simple multivariate sign test based on the transformation-retransformation approach (Chakraborty et al. 1998) together with the directional transformation proposed by Tyler (1987). Tyler's transformation is to find a data-driven transformation, $\mathbf{V}_x$, that is the positive-definite symmetric $p \times p$ matrix with trace$(\mathbf{V}_x) = p$ and satisfies that, for any $\mathbf{A}_x'\mathbf{A}_x = \mathbf{V}_x^{-1}$,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{A}_x\boldsymbol{x}_i}{||\mathbf{A}_x\boldsymbol{x}_i||}\right)\left(\frac{\mathbf{A}_x\boldsymbol{x}_i}{||\mathbf{A}_x\boldsymbol{x}_i||}\right)' = \frac{1}{p}\mathbf{I}_p, \tag{5}$$

where $||\cdot||$ is the Euclidean norm and $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. Such $\mathbf{V}_x$ is unique as Tyler showed, if the sample is drawn from a continuous p-dimensional distribution and $n > p(p-1)$. After obtaining $\mathbf{V}_x$, Randles (2000) proposes to use

$$Q = n\bar{\boldsymbol{v}}'[\widehat{\mathrm{Cov}}(\boldsymbol{v})]^{-1}\bar{\boldsymbol{v}} = np\bar{\boldsymbol{v}}'\bar{\boldsymbol{v}}, \tag{6}$$

as a test statistic and H_0 is rejected for large values, where

$$\boldsymbol{v}_i = \frac{\mathbf{A}_x\boldsymbol{x}_i}{||\mathbf{A}_x\boldsymbol{x}_i||}, \quad \bar{\boldsymbol{v}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{v}_i,$$

and we use the fact that $[\widehat{\mathrm{Cov}}(\boldsymbol{v})] = n^{-1}\sum_{i=1}^{n}\boldsymbol{v}_i\boldsymbol{v}_i' = p^{-1}\mathbf{I}_p$. This test is affine-invariant and it uses only the direction of an observation from the origin and does not use its distance from the origin. Randles (2000) shows that the Q is distribution-free under H_0 for the class of distributions with elliptical directions in which random variables are generated via $\boldsymbol{x}_i = r_i\mathbf{D}\boldsymbol{u}_i$, where the $\boldsymbol{u}_i$'s are i.i.d. uniform on the unit p sphere, $\mathbf{D}$ is a $p \times p$ nonsingular matrix, and the r_i's are positive scalars. The elliptical directions family contains all the elliptically symmetric distributions, such as multinormal and multivariate t distributions and certain skew distributions. This test not only has a small-sample distribution-free property over a broad class of distributions, but it also performs very well in comparison with Hotelling's T^2 and other multivariate nonparametric tests on non-normal distributions. Therefore, we are interested in tackling the monitoring problem (1) using this sign test.

2.2 A Multivariate Sign EWMA Control Chart

Firstly, it is worth pointing out that although the monitoring problem (1) is closely related to the standard hypothesis tests in Section 2.1, they are completely different and distinguished by the fundamental difference between *on-line* and *off-line* decision issues (c.f., Woodall and Montgomery 1999). This difference will become clear after the derivation of the proposed charting statistic.

The proposed control scheme contains two steps. The first step is to extract information from the reference sample of size m_0 by obtaining, say, a multivariate median, $\boldsymbol{\theta}_0$, and a transformation matrix, $\mathbf{A}_0$. This step is similar to that of constructing the MEWMA chart in which $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are estimated from the historical data before monitoring. Various definitions of multivariate median have been proposed to robustly describe "multivariate center", such as the most well-known one, the multivariate L_1 median (Gower 1974; Brown 1983), defined by the minimizer of $\sum_{i=-m_0+1}^{0} ||\boldsymbol{x}_i - \boldsymbol{\theta}||$ using the associated notation of model (1). We recommend using Hettmansperger and Randles's (2002) affine equivariant multivariate median, called the AEM-median for short. This definition of the median is more ideal than the L_1 median because it serves the sign test purpose and the by-product of finding such median is just the desired transformation matrix. The AEM-median, $\boldsymbol{\theta}_0$, and the associated transformation matrix, $\mathbf{A}_0$, are defined by the solutions of the following equations:

$$E\left(\frac{\mathbf{A}(\boldsymbol{x}-\boldsymbol{\theta})}{||\mathbf{A}(\boldsymbol{x}-\boldsymbol{\theta})||}\right) = 0, \tag{7}$$

$$E\left(\frac{\mathbf{A}(\boldsymbol{x}-\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\theta})'\mathbf{A}'}{||\mathbf{A}(\boldsymbol{x}-\boldsymbol{\theta})||^2}\right) = \frac{1}{p}\mathbf{I}_p, \tag{8}$$

and the corresponding sample version, $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{A}}_0)$, is defined by the solution of the sample equations based on m_0 historical observations,

$$\frac{1}{m_0}\sum_{i=-m_0+1}^{0}\left(\frac{\mathbf{A}(\boldsymbol{x}_i-\boldsymbol{\theta})}{||\mathbf{A}(\boldsymbol{x}_i-\boldsymbol{\theta})||}\right) = 0, \tag{9}$$

$$\frac{1}{m_0}\sum_{i=-m_0+1}^{0}\left(\frac{\mathbf{A}(\boldsymbol{x}_i-\boldsymbol{\theta})(\boldsymbol{x}_i-\boldsymbol{\theta})'\mathbf{A}'}{||\mathbf{A}(\boldsymbol{x}_i-\boldsymbol{\theta})||^2}\right) = \frac{1}{p}\mathbf{I}_p, \tag{10}$$

in which $\mathbf{A}$ is a $p \times p$ upper triangular positive-definite matrix with a one in the upper left-hand element. Note that Eq.(10) is essentially equivalent to Eq.(5) when

$\boldsymbol{\theta}$ is given and Eq.(8) is simply the population (asymptotical) version. The existence and uniqueness of $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ are shown by Hettmansperger and Randles (2002) under the assumption that the population is directionally symmetric, which is a weaker assumption than elliptical directions (see Randles 2000 for detailed comparisons of various assumptions). The convergence rate of $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{A}}_0)$ is the same as other classical descriptive statistics, say $m_0^{-1/2}$. Hettmansperger and Randles (2002) also present an iterative procedure to solve (9) and (10) simultaneously which is quite efficient in obtaining $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{A}}_0)$ from a given sample. As a side note, in a multinormal distribution with mean $\boldsymbol{\mu}_0$ and variance-covariance matrix $\boldsymbol{\Sigma}_0$, it is easily seen that $\boldsymbol{\theta}_0 = \boldsymbol{\mu}_0$ and $\mathbf{A}_0'\mathbf{A}_0 = p^{-1}\text{trace}(\boldsymbol{\Sigma}_0)\boldsymbol{\Sigma}_0^{-1}$. In what follows, we use $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ rather than $(\widehat{\boldsymbol{\theta}}_0, \widehat{\mathbf{A}}_0)$ unless indicated otherwise, as a SPC Phase II convention.

In light of (5) and (6), after $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ is specified or estimated, for on-line collected observations $\boldsymbol{x}_i, i = 1, 2, \ldots$, it is straightforward to standardize and transform them to obtain the unit vector $\boldsymbol{v}_i$ through

$$\boldsymbol{v}_i = \frac{\mathbf{A}_0(\boldsymbol{x}_i - \boldsymbol{\theta}_0)}{||\mathbf{A}_0(\boldsymbol{x}_i - \boldsymbol{\theta}_0)||}. \tag{11}$$

With this choice, the unit vectors of the transformed data have a variance-covariance structure like that of a random variable that is uniform on the unit p-sphere. Then, we define an EWMA sequence similar to (4)

$$\boldsymbol{w}_i = (1 - \lambda)\boldsymbol{w}_{i-1} + \lambda \boldsymbol{v}_i, \tag{12}$$

where the initial vector, $\boldsymbol{v}_0$, is usually taken to be $E(\boldsymbol{v}_i)$ and thus should be $\mathbf{0}$ due to our definition in (7). Finally, the proposed control chart triggers a signal if

$$Q_i = \frac{2 - \lambda}{\lambda} p \boldsymbol{w}_i' \boldsymbol{w}_i > L, \tag{13}$$

where $L > 0$ is a control limit chosen to achieve a specific IC ARL. Note that according to the multivariate sign test (6), we use $\text{Cov}(\boldsymbol{w}_i) \approx \lambda \text{Cov}(\boldsymbol{v}_i)/(2-\lambda) = p^{-1}\lambda \mathbf{I}_p/(2-\lambda)$, which yields (13). The weighted average sum (12) plays a similar role to that of $\bar{\boldsymbol{v}}$ but reflects the relevance of the data: the more recent observations are more informative for detecting the change and thus getting the larger weights. Another difference between (6) and (13) is that the former involves a current data-driven transformation, $\mathbf{A}_x$, but the latter uses a population (or deemed as historically data driven) transformation matrix, $\mathbf{A}_0$. This is analogous to the difference between Hotelling's

T^2 statistic and the quadratic statistic (3). Hereafter, this chart is referred to as the multivariate sign EWMA (MSEWMA) control chart. In what follows, we show some useful properties of the MSEWMA chart and the proofs are given in the supplemental file.

Proposition 1 *The MSEWMA chart is affine-invariant.*

This proposition says that if the data points are rotated or if they are reflected around a $p-1$ dimensional hyperplane or if the scales of measurement are altered, the value of the charting statistic stays the same. This property is intuitively appealing, and it also ensures that the performance of MSEWMA is the same for any variance-covariance.

Proposition 2 *The MSEWMA chart is strictly distribution free in the sense that its in-control run length distribution is the same for the class of distributions with elliptical directions.*

This proposition is particularly useful in determining the control limit, L, because, for any continuous process distribution with elliptical directions, it is the same as achieving the desired IC run-length distribution.

Proposition 3 *The Q_i process is a Markov chain if the underlying distribution is from the class of distributions with elliptical directions.*

By this result and some similar arguments in Runger and Prabhu (1996), the MSEWMA shares a similar key property with its parametric counterpart, MEWMA. That is, the IC ARL of MSEWMA for distributions with elliptical directions can be calculated via the Markov chain model. The details are presented in the Appendix. Although the two-dimensional Markov chain model developed by Runger and Prabhu (1996) can be extended to the MSEWMA chart to evaluate its OC ARL, this is not of great interest here because different distributions and OC models have different representations of the transition probability matrices and it seems quite difficult to present a unified framework. Hence, we choose to use simulation to evaluate the OC ARL performance in the next section.

Finally, by Theorem 1 of Randles (2000) and the arguments in Tyler (1987) we can obtain the following asymptotic results without much difficulty.

Proposition 4 *Under the IC model, $Q_i \xrightarrow{d} \chi_p^2$ as $\lambda \to 0$, $i \to \infty$ and $\lambda i \to \infty$.*

To end this subsection, we note that when a group of g observations, say $\{\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{ig}\}$, are taken sequentially from the process at each time point, the MSEWMA chart can be readily defined in a similar way to (13) by using

$$\boldsymbol{v}_i = \frac{1}{g}\sum_{j=1}^{g} \frac{\mathbf{A}_0(\boldsymbol{x}_{ij} - \boldsymbol{\theta}_0)}{||\mathbf{A}_0(\boldsymbol{x}_{ij} - \boldsymbol{\theta}_0)||}.$$

2.3 Design and Implementation of the Proposed Scheme

On computation: For on-line detection, the computation burden of the MSEWMA chart is similar to that of MEWMA since both of them only require computing a working EWMA sequence and a quadratic form. Of course, unlike the MEWMA chart, in Phase I analysis, estimating $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ involves iterative routines and thus it is a bit more complicated than estimating $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. However, by using some efficient algorithms provided by Tyler (1987) and Hettmansperger and Randles (2002), convergence of $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ from the historical data with any practical p and m_0 is guaranteed and the convergence is usually quite fast. The computation task is trivial by virtue of the massive computing and data storage capabilities of modern computers. For instance, given $m_0 = 20,000$, for $p = 10, 100$ and 1000, usually about 0.1 second, 1.2 minute and 130 minutes are required to complete the iterative procedure using a Pentium-M 2.1MHz CPU, respectively. The detailed algorithm is provided in the supplemental file.

On the control limits and robustness: Based on Proposition 2, the control limits for distributions with elliptical directions are the same. Hence, we use the standard multivariate normal distribution to find the control limit (see the supplemental file for details). Table 1 provides the control limits of the MSEWMA chart for various commonly used combinations of λ, p and IC ARL's, obtained using a Markov chain with $m = 200$ transition states. We have conducted simulations to verify the accuracy of the Markov chain approximation, and the results are very satisfactory as long as $m > 50$. The Fortran code for implementing the proposed scheme, including the procedures for finding $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ and the control limits, are available from the authors upon request. The simulation results shown in Section 3 demonstrate that the IC run-length performance of MSEWMA is quite robust under various process distributions including very skewed distributions. Therefore, the control limits tabulated in Table 1 can be used for any continuous distribution.

Table 1: The control limits of the MSEWMA chart for IC ARL=200, 370 and 500 under p-variate distributions with elliptical directions.

IC ARL	λ	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 7$	$p = 10$
	0.4	6.009	7.920	9.668	11.321	14.448	18.841
	0.2	7.831	9.830	11.674	13.414	16.708	21.329
200	0.1	8.043	10.052	11.896	13.636	16.911	21.532
	0.05	7.225	9.177	10.963	12.646	15.819	20.288
	0.025	5.895	7.691	9.345	10.906	13.864	18.066
	0.4	6.276	8.294	10.125	11.847	15.083	19.628
	0.2	8.567	10.687	12.626	14.448	17.876	22.649
370	0.1	9.183	11.303	13.249	15.077	18.511	23.310
	0.05	8.605	10.700	12.607	14.404	17.774	22.472
	0.025	7.399	9.392	11.205	12.918	16.124	20.644
	0.4	6.390	8.459	10.329	12.083	15.388	19.983
	0.2	8.904	11.074	13.058	14.924	18.409	23.284
500	0.1	9.716	11.887	13.877	15.750	19.247	24.147
	0.05	9.265	11.417	13.375	15.216	18.663	23.462
	0.025	8.126	10.198	12.081	13.852	17.165	21.812

It should be pointed out that when m_0 is not large, there would be considerable uncertainty in the parameter estimation, which in turn would distort the IC run length distribution of the MSEWMA control chart. Even if the control limit of the chart were adjusted properly to obtain the desired IC run length behavior, its OC run length would still be severely compromised (cf., Jones 2002). This is essentially analogous to the estimated parameters problem in the context of parametric control charts (see Jensen et al. 2006 for an overview). We use simulated examples to show that the performances of MSEWMA and MEWMA are similarly affected when m_0 is not large. To deal with the situation when a sufficiently large reference dataset is unavailable, self-starting methods that handle sequential monitoring by simultaneously updating parameter estimates and checking for OC conditions have been developed accordingly (see, e.g., Quesenberry 1995). The further studies, including thorough investigations of the effect of m_0 on the MSEWMA chart and the development of corresponding self-starting charts, are beyond the scope of this paper but should be subjects of future research.

On choosing the smoothing weight, λ:

Unlike the MEWMA chart in which the choice of λ should be chosen to balance the robustness to non-normality and the detection ability to various shift magnitudes (c.f., Stoumbos and Sullivan 2002), the MSEWMA chart is robust under IC with any weight, $\lambda \in (0, 0.2]$, except for very skewed distributions and high dimensional cases. In general, a smaller λ leads to a quicker detection of smaller shifts (c.f., e.g., Lucas and Saccucci 1990; Prabhu and Runger 1997). This statement is still valid with MSEWMA. Based on our simulation results, we suggest choosing $\lambda \in [0.05, 0.2]$, which is a reasonable choice in practice, and using $\lambda \in [0.05, 0.1]$ when *a priori* indicates the underlying distribution is very skewed.

3 Numerical Performance Assessment

We present some simulation results in this section regarding the numerical performance of the proposed MSEWMA chart and compare it with some other procedures in the literature. The MEWMA chart (defined by Eqs.(3)-(4)) and the CUSUM chart of the first antirank (denoted as ARCUSUM) proposed by Qiu and Hawkins (2001) are considered. The ARCUSUM is briefly reviewed as follows. Define A_i as the first antirank of $\boldsymbol{x}_i$ and $\boldsymbol{\xi}_i = (\xi_{i,1}, \ldots, \xi_{i,p})^T$ with a single nonzero component 1 located in the jth position if $A_i = j$. The charting statistic of ARCUSUM, y_i, is defined by

$$y_i = (\boldsymbol{S}_i^{(1)} - \boldsymbol{S}_i^{(2)})^T \mathrm{diag}\{1/S_{i,1}^{(2)}, \ldots, 1/S_{i,p}^{(2)}\}(\boldsymbol{S}_i^{(1)} - \boldsymbol{S}_i^{(2)}),$$

where $\boldsymbol{S}_i^{(1)} = \boldsymbol{S}_i^{(2)} = \mathbf{0}$ if $C_i \le k$ (a reference constant); otherwise,

$$\begin{aligned}
\boldsymbol{S}_i^{(1)} &= (\boldsymbol{S}_{i-1}^{(1)} + \boldsymbol{\xi}_i)(C_i - k)/C_i, \\
\boldsymbol{S}_i^{(2)} &= (\boldsymbol{S}_{i-1}^{(2)} + \boldsymbol{g})(C_i - k)/C_i.
\end{aligned}$$

Here $\boldsymbol{S}_0^{(1)} = \boldsymbol{S}_0^{(2)} = \mathbf{0}$, $\boldsymbol{g} = (g_1, \ldots, g_p)^T = E_{H_0}(\boldsymbol{\xi}_i)$ and

$$\begin{aligned}
C_i =& [(\boldsymbol{S}_i^{(1)} - \boldsymbol{S}_i^{(2)}) + (\boldsymbol{\xi}_i - \boldsymbol{g})]^T \mathrm{diag}\{(S_{i-1,1}^{(2)} + g_1)^{-1}, \ldots, (S_{i-1,p}^{(2)} + g_p)^{-1}\} \\
& \times [(\boldsymbol{S}_i^{(1)} - \boldsymbol{S}_i^{(2)}) + (\boldsymbol{\xi}_i - \boldsymbol{g})].
\end{aligned}$$

We start by assuming that m_0 is sufficiently large, in this case thirty thousand. In all the underlying distributions considered, we first generate m_0 i.i.d. samples and then estimate $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $(\boldsymbol{\theta}_0, \mathbf{A}_0)$. For the ARCUSUM chart, the Phase II observations are firstly standardized through $\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{x}_i - \boldsymbol{\mu}_0)$ following Qiu and Hawkins's (2001) suggestion. Control limits of the MEWMA and ARCUSUM charts are determined by simulations to attain the nominal IC ARL under the standard multinormal distribution, while the control limits given in Table 1 are used for MSEWMA. Since the zero-state and steady-state ARL (SSARL) comparison results are similar, only the SSARLs are provided. To evaluate the SSARL behavior of each chart, any series in which a signal occurs before the $(\tau + 1)$-th observation is discarded (c.f., Hawkins and Olwell 1998). Because a similar conclusion holds for other cases, here we only present the results when IC ARL=200 and $\tau = 50$ for illustration. All the ARL results in this section are obtained from 100,000 replications.

Following the robustness analyses in Stoumbos and Sullivan (2002), 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 $\mathrm{Gam}_{p,\zeta}$. Details on the multivariate t and gamma distributions can be found in the Appendix to Stoumbos and Sullivan (2002). In addition, the following two distributions are involved in the comparison: (iv) measurement components are i.i.d. from chi-square distributions with ζ degrees of freedom, denoted as $\chi^2_{p,\zeta}$; (v) measurement components are i.i.d. from the Cauchy distribution, denoted as Cau_p. As discussed by Stoumbos and Sullivan (2002), since the multivariate normal and t distributions are elliptically symmetrical, the MEWMA's OC performance depends on a shift in the process mean vector only through a noncentrality parameter. This is still true for the MSEWMA chart because of its affine invariance. However, with the other distributions, such as multivariate gamma, the performance is not invariant to the covariance matrix of the "implicit" multivariate normal observation. 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 MSEWMA chart, and thus we only choose certain representative models for illustration. Specifically, for the (i)-(iii) distribution cases, the covariance matrix $\boldsymbol{\Sigma}_0 = (\sigma_{ij})$ is chosen to be $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.5^{|i-j|}$, for $i, j = 1, 2, \ldots, p$. In the interest of brevity, a shift of size δ in only the first component is used, say, $\boldsymbol{x}_i + \delta \boldsymbol{e}_1$ with $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$, unless stated otherwise.

Table 2: ARL and SDRL values with multinormal distributions.

	δ	MSEWMA				MEWMA				ARCUSUM			
		$\lambda = 0.2$		$\lambda = 0.05$		$\lambda = 0.2$		$\lambda = 0.05$		$k = 1.0$		$k = 0.5$	
		ARL	SDRL	ARL	SDRL	ARL	SDRL	ARL	SDRL	ARL	SDRL	ARL	SDRL
	0.00	199	194	200	187	200	194	199	183	201	197	200	204
	0.25	96.8	93.4	63.9	52.9	93.4	90.4	59.4	48.2	151	150	144	143
	0.50	35.4	29.9	26.5	15.9	31.5	26.8	23.4	14.1	98.6	96.5	88.5	86.6
	0.75	17.6	12.1	16.5	7.90	14.6	10.4	14.0	6.93	68.1	66.4	59.9	56.9
$p = 3$	1.00	11.3	6.15	12.5	5.11	8.81	5.21	9.90	4.40	51.4	49.2	44.0	40.4
	1.50	7.06	2.56	9.12	3.08	4.80	2.15	6.38	2.49	34.6	31.9	27.8	22.8
	2.00	5.72	1.61	7.85	2.41	3.34	1.27	4.72	1.72	26.4	23.5	21.0	15.5
	3.00	4.91	1.15	6.96	2.03	2.17	0.71	3.23	1.08	17.7	14.3	14.7	9.85
control limit		9.830		9.177		11.865		9.376		4.202		6.000	
	0.00	200	192	201	178	200	194	199	174	200	227	199	246
	0.25	126	123	78.1	66.5	128	125	77.8	65.7	209	209	222	225
	0.50	53.0	48.3	32.2	20.1	52.4	47.4	30.1	19.7	185	185	184	184
	0.75	24.5	18.7	19.3	9.39	23.2	18.3	18.3	8.79	161	161	154	154
$p = 10$	1.00	14.3	8.75	13.9	5.84	12.7	8.10	12.5	5.51	140	138	127	126
	1.50	7.72	3.18	9.45	3.28	6.38	3.01	7.88	3.00	99.9	97.5	83.1	77.8
	2.00	5.65	1.78	7.55	2.33	4.25	1.66	5.83	2.06	67.4	63.5	54.7	47.8
	3.00	4.24	1.00	6.00	1.68	2.65	0.85	3.90	1.25	31.6	26.9	27.3	20.3
control limit		21.329		20.288		24.059		20.701		14.097		13.675	

We first consider the multinormal distribution. A low-dimensional case with $p = 3$ and a higher-dimensional case $p = 10$ are involved. The simulation results for the MEWMA and MSEWMA charts with $\lambda = 0.05$ and $\lambda = 0.2$ and the ARCUSUM with reference values $k = 0.5$ and 1.0 are presented in Table 2. Besides the ARLs, the corresponding standard deviations of the run lengths (SDRL) are also included in this table to give a broader picture of the run-length distribution. From this table, we observe that the MEWMA chart has superior efficiency as we would expect, since the parametric hypothesis is the correct one in this case. The MSEWMA chart also offers quite satisfactory performance and the difference between MSEWMA and MEWMA is not significant, even when p is large. It should be pointed out that the

Table 3: ARL values with a multivariate t distribution of $t_{p,5}$.

	δ	MSEWMA			MEWMA			ARCUSUM	
		$\lambda = 0.2$	$\lambda = 0.05$	$\lambda = 0.01$	$\lambda = 0.2$	$\lambda = 0.05$	$\lambda = 0.01$	$k = 1.0$	$k = 0.5$
	0.00	201	200	199	91.6	177	204	200	201
	0.25	100	66.5	60.2	72.7	78.2	68.7	155	147
	0.50	38.7	28.1	31.9	43.3	32.9	35.3	102	93.6
	0.75	19.7	17.7	22.4	23.9	19.3	23.5	72.2	63.1
$p = 3$	1.00	12.7	13.4	17.9	14.1	13.4	17.6	54.9	47.8
	1.50	7.88	9.79	13.8	7.10	8.34	11.8	36.8	31.3
	2.00	6.31	8.36	12.0	4.62	6.10	8.81	27.5	23.4
	3.00	5.21	7.27	10.6	2.82	4.05	6.02	18.3	15.6
control limit		9.830	9.177	5.333	11.865	9.376	5.304	4.202	6.000
	0.00	200	200	199	47.0	133	197	200	199
	0.25	130	82.4	71.6	43.1	82.7	83.5	212	222
	0.50	56.9	34.6	39.1	35.6	42.9	46.2	188	189
	0.75	26.8	20.8	27.3	26.6	25.2	31.2	166	160
$p = 10$	1.00	15.9	15.0	21.1	18.6	17.5	23.7	146	134
	1.50	8.54	10.3	15.4	9.84	10.7	15.9	104	87.8
	2.00	6.24	8.19	12.7	6.23	7.79	12.1	71.5	57.9
	3.00	4.60	6.45	10.1	3.59	5.08	8.23	33.7	28.4
control limit		21.329	20.288	13.966	24.059	20.701	13.968	14.097	13.675

superiority of MEWMA becomes more significant when δ is quite large, say $\delta \geq 3$. The analogous phenomenon for univariate nonparametric charts has been mentioned in the literature, e.g., by Hackel and Ledolter (1991) and Zhou et al. (2009). The MSEWMA, which is essentially based on signs rather than distances, shares a similar drawback as those rank-based charts for univariate processes. That is, even though the shift is quite large, the ranks or signs of the observations may not be able to grow larger. In addition, both the MEWMA and MSEWMA significantly outperform the ARCUSUM chart. Note that in some cases, such as $p = 10$ and $\delta = 0.25$, the ARCUSUM is even not ARL-unbiased.

Next, the multivariate t distribution is considered. As this distribution belongs to

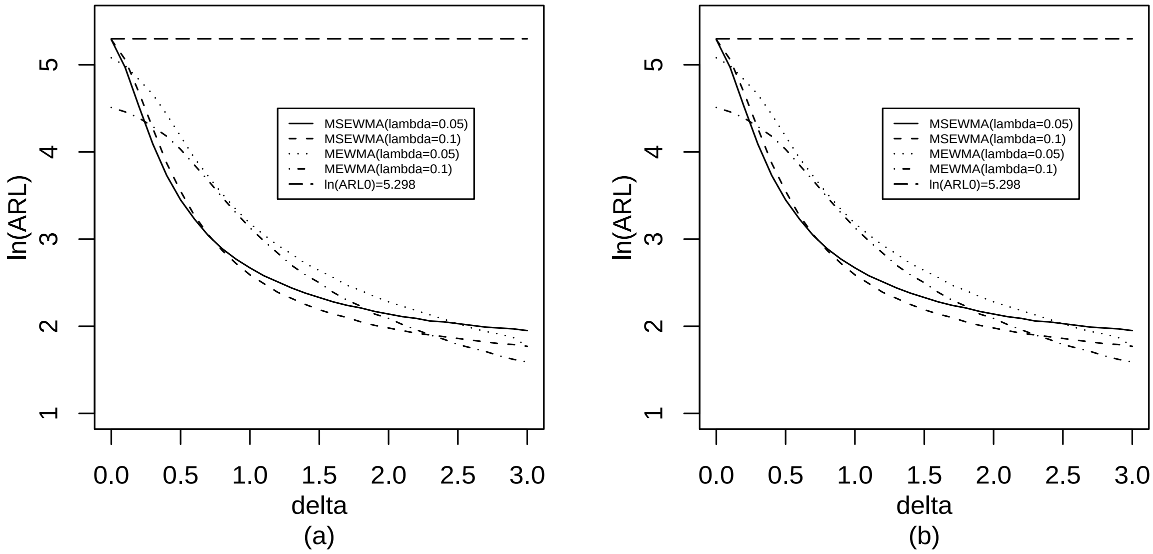

Figure 1: OC ARL comparison of the MSEWMA and MEWMA charts using $\lambda = 0.05$ and 0.1 with a shift in the first component under: (a) $t_{5,3}$; (b) $t_{5,10}$.

the class of distributions with elliptical directions, we do not focus on the robustness of IC ARL performance of the MSEWMA chart but on its OC ARL comparison with the other two procedures. Table 3 shows the ARL comparison with a shift in the first component of multivariate t measurements with five degrees of freedom when $p = 3$ or $p = 10$. For the two EWMA charts, besides 0.2 and 0.05, the ARLs with $\lambda = 0.01$ are also reported. Obviously, the MSEWMA and ARCUSUM charts can achieve the nominal IC ARL but the MEWMA has considerable bias in IC ARL except for $\lambda = 0.01$. The MSEWMA chart is more efficient in detecting the small and moderate shifts than is the MEWMA chart with the same value of λ in the sense that even when the IC ARL is much larger than that of the MEWMA, the OC ARLs decrease much faster than with the MEWMA. When λ is small, say $\lambda = 0.01$ in this example, the MEWMA chart is robust to non-normality under the IC situation; however, its ability to detect moderate and large shifts is largely compromised. In particular, when $p = 10$, the MSEWMA with $\lambda = 0.05$ performs uniformly better than MEWMA does, and the difference is quite remarkable. Again, although it is robust to multivariate t distribution for IC performance, the ARCUSUM chart is not as sensitive to the process shift as the other two charts are.

Certainly, the superiority of MSEWMA over MEWMA depends on the degrees

Table 4: IC ARL values with multivariate gamma distributions of $\mathrm{Gam}_{p,\zeta}$.

	ζ	MSEWMA					MEWMA					ARCUSUM
		λ					λ					k
		0.4	0.2	0.1	0.05	0.025	0.4	0.2	0.1	0.05	0.025	0.5
$p=3$	1	152	179	187	190	191	34.1	62.2	105	167	206	200(196)
	2	176	191	196	197	196	42.6	76.8	129	175	204	202(201)
	3	184	194	197	198	197	49.7	88.3	140	183	204	201(200)
	4	188	194	199	199	197	56.2	96.7	148	185	199	203(202)
	5	192	197	198	202	200	62.7	105	158	189	198	199(197)
	10	195	200	200	201	198	85.7	135	175	194	199	202(200)
	15	198	199	199	200	202	105	155	183	201	203	200(200)
	30	199	200	199	199	199	130	168	188	195	199	199(198)
$p=5$	1	133	165	183	189	192	26.7	50.0	91.5	147	194	200(208)
	2	158	181	192	198	198	35.6	66.4	109	168	202	199(210)
	3	168	187	196	199	198	42.3	77.6	129	174	197	200(207)
	4	174	190	197	198	199	47.9	85.9	137	177	196	200(208)
	5	179	193	198	199	198	53.6	94.1	144	179	196	201(210)
	10	188	196	198	200	197	76.9	127	165	192	200	200(206)
	15	191	199	201	197	198	92.2	138	176	193	199	199(208)
	30	196	202	200	199	201	122	160	187	196	200	201(210)
$p=10$	1	116	158	185	191	196	20.7	40.3	76.1	130	178	198(241)
	2	136	170	191	194	197	28.6	56.3	101	153	188	200(245)
	3	154	178	193	197	199	34.6	67.6	112	164	190	200(242)
	4	164	184	195	198	200	41.7	77.4	128	173	195	199(245)
	5	167	186	196	200	200	47.2	87.2	137	177	196	201(246)
	10	181	197	199	199	201	70.5	114	163	188	196	200(246)
	15	186	199	201	200	200	85.8	133	171	190	198	199(245)
	30	191	200	199	200	199	117	164	192	193	199	199(241)

of freedom, ζ; that is the deviation from multinormality. This can be clearly seen in Figure 1, which shows the ARL curves (in the log scale) of the MEWMA and MSEWMA charts in the left and right panels for $\zeta = 3$ and $\zeta = 10$, respectively, when $p = 5$. The MSEWMA chart can maintain the desired IC ARL with any λ and it outperforms the corresponding MEWMA chart except for very large shifts when $\zeta = 3$. This advantage reduces when $\zeta = 10$, however. We can expect that the performance comparison between them will become more similar to that of Table 1 where the multinormal distribution is considered as ζ becomes larger.

Now, we turn to Table 4, which gives IC ARL values with multivariate gamma observations. As before, various cases with combinations of dimensionality, λ and degrees of freedom ζ are considered. From this table, we can see that the MSEWMA is quite satisfactorily robust to the skewed distribution as long as λ is not too large (i.e., $\lambda > 0.2$). When $\lambda \leq 0.1$, the MSEWMA's IC ARL is always quite close to the nominal one even for the extremely non-normal and high-dimensional distribution of $\mathrm{Gam}_{10,1}$. In comparison, the MEWMA usually has a large bias in the IC ARL and the degradation becomes more pronounced as the dimensionality increases. For $\zeta \leq 5$, only when λ is 0.025 will the MEWMA chart maintain a desired IC ARL. The ARCUSUM chart is still robust in this case from the viewpoint of ARL. We also observe that the ARCUSUM may have an increase in its SDRL with increasing dimension, which is partly due to excessive false alarms in short runs.

Figures 2 and 3 respectively summarize the ARL curves of the MSEWMA and MEWMA with a shift of size δ in the first component and equal shifts of size δ in the first two components, with multivariate gamma distributions. In both figures, we set $\zeta = 2$ and the results of $p = 5$ and $p = 10$ are shown in left and right panels, respectively. We do not consider the ARCUSUM because it has been shown to be not as efficient as the other two procedures in the preceding examples and some additional simulations (not reported here). Note that for a fair comparison, the MSEWMA with $\lambda = 0.1$ or 0.05 and the MEWMA with $\lambda = 0.025$ or 0.01 are considered. Figures 2 and 3 present similar results: (i) with similar IC ARL, the MSEWMA is much better than MEWMA in detecting small and moderate shifts when $p = 5$ while MEWMA has a certain advantage for the large shifts as expected; (ii) when p becomes larger, the improvement in MSEWMA over MEWMA is tremendous.

Figure 4 shows ARL comparisons of the MSEWMA and MEWMA charts in mon-

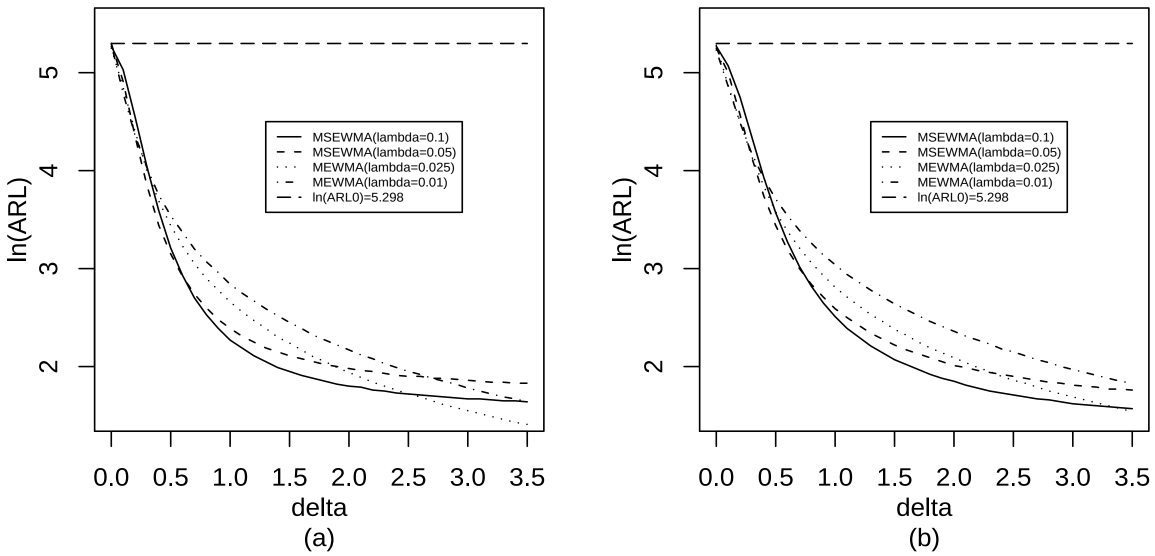

Figure 2: OC ARL comparison of the MSEWMA and MEWMA charts with a shift of size δ in the first component of multivariate gamma observations under (a) $\mathrm{Gam}_{5,2}$ and (b) $\mathrm{Gam}_{10,2}$.

itoring a shift in the first component with $\chi^2_{p,1}$ and Cau_p observations. We present only the results for $p = 5$ in this example and a similar conclusion holds for other cases. Clearly, with an appropriate value of λ, say 0.1 or 0.05, the MSEWMA chart not only attains the desired IC ARL, but it also outperforms the MEWMA chart in detecting small and moderate shifts by a quite large margin. This demonstrates the fact that the MSEWMA chart is more sensitive to process shifts from non-normal observations, especially for extremely skewed or heavy-tailed distributions, compared with the conventional parametric MEWMA chart. Interestingly, the MEWMA chart completely fails with the Cau_p observations. It has a much larger IC ARL than the nominal one and it hardly changes as δ increases. This is not surprising since this observation is consistent with findings on the advantage of rank-based tests or estimations over the associated parametric methods in the contexts of robust statistics, e.g., see Hettmansperger and McKean (1998). We should emphasize that the physical measurement for which the mean is not finite is rarely seen in practical applications and this Cauchy numerical example is just used for illustration of the robustness of MSEWMA.

We conducted some other simulations with various correlation structures, p and

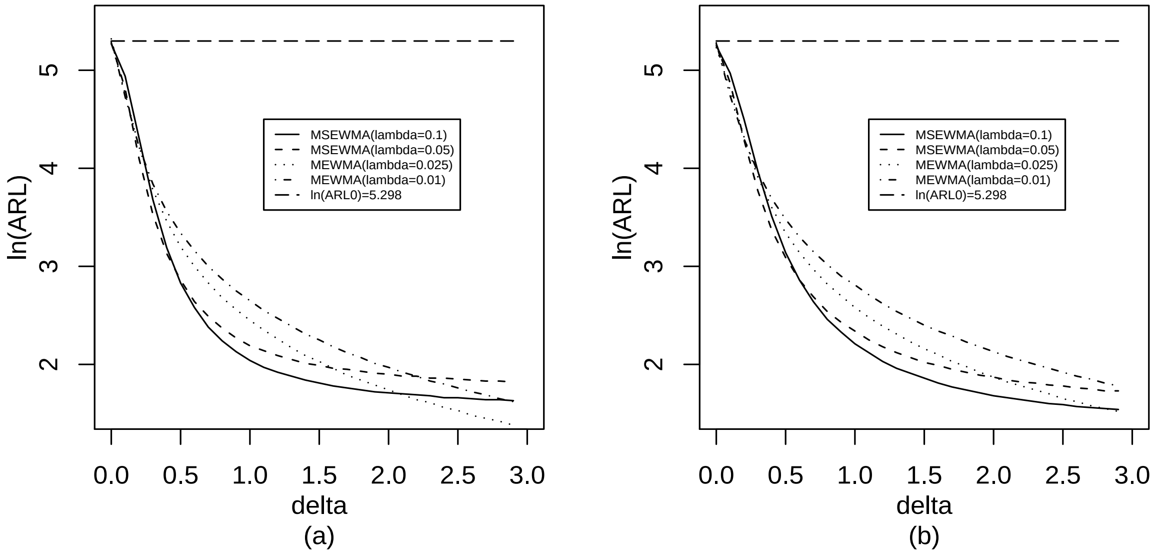

Figure 3: OC ARL comparison of the MSEWMA and MEWMA charts with equal shifts of size δ in the first two components with multivariate gamma observations under (a) $\mathrm{Gam}_{5,2}$ and (b) $\mathrm{Gam}_{10,2}$.

IC ARL, to check whether the above conclusions would change in other cases. These simulation results, not reported here but available from the authors, show that the MSEWMA chart works well for other correlation structures as well in terms of its OC ARL, and its good performance still holds for other choices of p and IC ARL.

In all the foregoing examples, it is assumed that the IC parameters are known or, equivalently, that they are estimated from a sufficiently large reference dataset. Finally, we study the performance of MSEWMA when this assumption is violated. To this end, we use the multinormal and multivariate gamma distributions with two degrees of freedom. Only the case $p = 5$ is considered and the nominal IC ARL is fixed as 200. Table 5 shows the IC ARLs and SDRLs of MSEWMA and MEWMA when the IC parameters $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ for MSEWMA and $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ for MEWMA are computed from an IC dataset with various historical sample sizes, m_0. In each replication, a sample of size m_0 is firstly generated and the IC parameters are estimated from this sample. Then, an independent sequence of multivariate observations is generated and both charts are used to obtain the corresponding run lengths. From this table, it can be seen that (i) when the sample size of the IC dataset is relatively small, the actual IC ARLs and SDRLs of the two charts are both quite far away from the

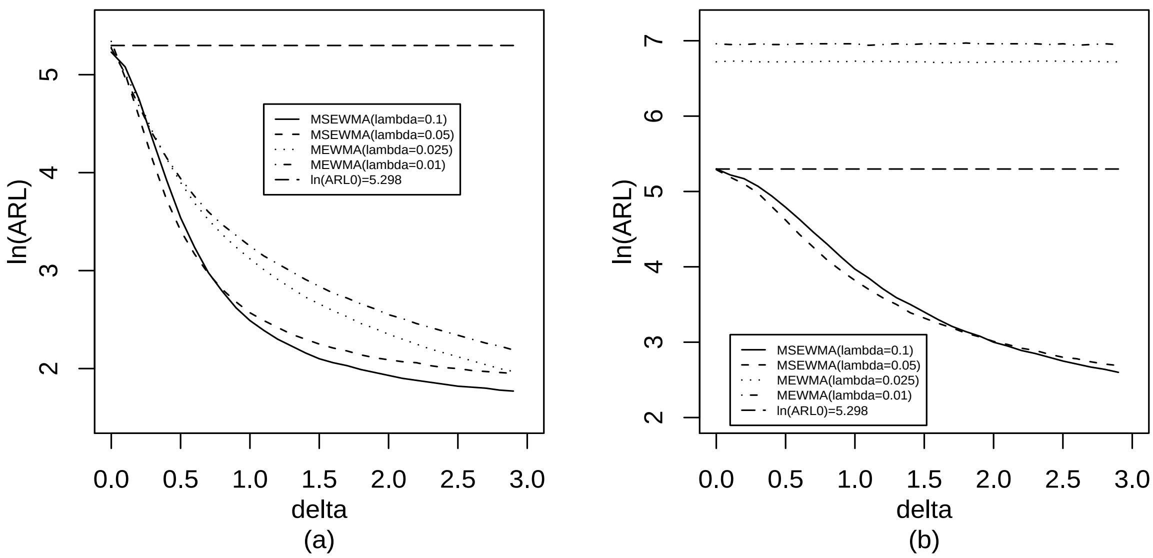

Figure 4: OC ARL comparison of the MSEWMA and MEWMA charts with a shift in the first component under: (a) five-dimensional chi-square distribution, $\chi^2_{5,1}$; (b) five-dimensional Cauchy distribution with 1 degree of freedom, Cau_p.

nominal level of 200, (ii) when the sample size of the IC dataset increases, such biases decrease, and (iii) the biases in IC ARL of the two charts are similar, although it appears that the chart with the smaller λ has a little larger bias in IC ARL, which is consistent with the findings in the studies of the univariate EWMA chart with estimated parameters (Jones et al. 2001). In this paper, we make no attempt to further analyze this problem, but we think that the designs with estimated parameters for both the MEWMA and MSEWMA charts certainly warrant future research.

4 Real Data Applications

In this section, we demonstrate the proposed methodology by applying it to two datasets: one is from an aluminium electrolytic capacitor manufacturing process; the other is the aluminum smelter example used by Qiu and Hawkins (2001). We elaborate on the first one and use it to illustrate the implementation of MSEWMA step by step. The application on the latter dataset will be briefly presented since that dataset is discussed in several papers (c.f., Zamba and Hawkins 2006; Hawkins and

Table 5: IC ARL and SDRL values with various Phase I sample sizes, m_0. Numbers in parentheses are SDRL values.

	$N_p(\mathbf{0}, \mathbf{\Sigma})$				$\mathrm{Gam}_{p,2}$			
	MSEWMA		MEWMA		MSEWMA		MEWMA	
m_0	λ				λ			
	0.1	0.05	0.025	0.01	0.1	0.05	0.025	0.01
300	155(148)	145(130)	133(111)	134(94.1)	157(146)	152(136)	130(114)	134(96.6)
400	165(157)	155(138)	144(120)	144(100)	163(154)	160(143)	140(123)	141(100)
500	168(163)	163(146)	152(124)	153(110)	170(160)	165(147)	149(131)	152(113)
750	180(170)	174(154)	165(139)	163(117)	179(168)	175(159)	163(144)	165(121)
1000	184(173)	180(163)	173(146)	171(127)	181(172)	178(162)	172(154)	172(128)
1500	189(177)	185(166)	180(153)	178(131)	184(173)	186(167)	176(154)	182(138)
2000	193(186)	190(170)	183(157)	184(137)	187(175)	188(168)	183(161)	188(141)
4000	197(188)	194(175)	190(163)	194(147)	191(180)	193(174)	190(167)	192(146)

Maboudou-Tchao 2007).

The aim of an aluminium electrolytic capacitor (AEC) process is to transform the raw materials (anode aluminum foil, cathode aluminum foil, guiding pin, electrolyte sheet, plastic cover, aluminum shell and plastic tube) into AECs. The whole process includes a sequence of operations, such as clenching, rolling, soaking, assembly, cleaning, aging and classifying. Before packing, a careful monitoring (or inspection) step is required by sampling from a production batch. The three most important characteristics in the specification of an AEC, the capacitance, loss tangent (or equivalently dissipation factor) and leakage current level (labeled as x_1, x_2 and x_3), are automatically measured by an electronic device at some given measuring voltage, frequency and temperature.

The dataset comprises 200 vectors (included in the supplemental file). Figure 5 (a)-(c) show the time series plots of the raw data. We use the first 170 vectors as the historical sample to calibrate the necessary parameters and the others for test. A calibration sample of this size may be smaller than ideal to determine fully the in-control distribution (c.f., Table 5) but it suffices to illustrate the use of the method in a real-world setting. The normal Q-Q plots for the three measurements based on those

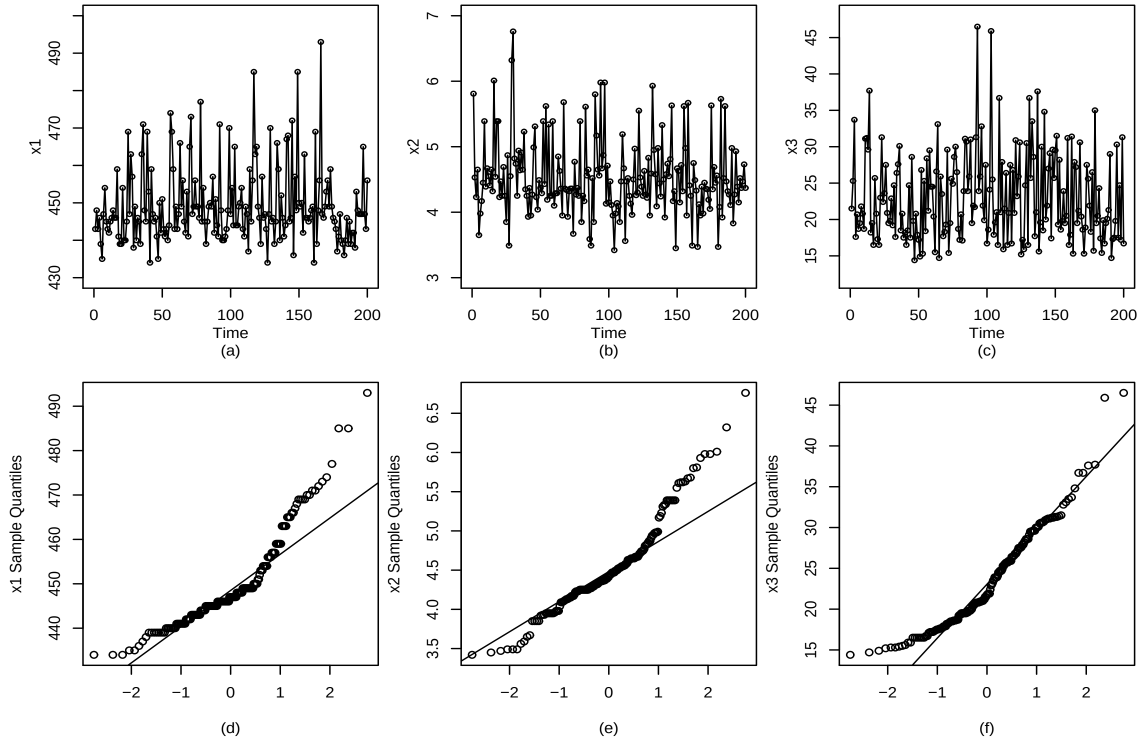

Figure 5: (a)-(c): The time series plots of the aluminium electrolytic capacitors data; (d)-(f): the normal Q-Q plots for x_1, x_2 and x_3 respectively.

170 vectors are shown in Figure 5 (d)-(f) which clearly indicate that the marginals are not normal, especially for x_1 and x_2. Both the Kolmogorov-Smirnov and Shapiro-Wilk goodness-of-fit tests for normality conclude that all the three variables in this dataset are not normally distributed (all the p-values are smaller than 1×10^{-5}). Mardia's (1970) multivariate normality test is also performed and the p-value is about 1.73×10^{-7}. All these tests together with Figure 5 (d)-(f) suggest that the multivariate normality assumption is not valid and thus we could expect that the MSEWMA chart would be more robust and powerful than normal-based approaches for this dataset. The estimated mean vectors, correlation matrices, and $(\boldsymbol{\theta}_0, \mathbf{A}_0)$ are presented in Table 6. In Table 6, it can be seen that the correlation matrices contain several large entries, which demonstrate that the variables have considerable interrelationships and consequently a multivariate control chart is likely to be more appropriate than a univariate control chart.

After computing all necessary estimates from the IC data, we are ready to construct the proposed chart for Phase II analysis. Its IC ARL is fixed at 200, and λ is

Table 6: The estimated mean vector, correlation matrix, multivariate median $\boldsymbol{\theta}_0$ and transformation matrix $\mathbf{A}_0$ for the AEC data

Sample mean vector			Sample multivariate median, $\boldsymbol{\theta}_0$		
449.82	4.535	23.41	448.27	4.457	22.48
Sample correlation matrix			Sample transformation matrix, $\mathbf{A}_0$		
1.000	-0.239	0.173	1.000	3.661	-0.179
-0.239	1.000	0.121	0.000	15.265	-0.367
0.173	0.121	1.000	0.000	0.000	1.130

chosen to be 0.1. The control limit is 10.052 given in Table 1. Figure 6 shows the resulting MSEWMA chart (solid curve connecting the dots) along with its control limit (the solid horizontal line). The corresponding MEWMA with $\lambda = 0.05$ (dashed curve connecting circles) and ARCUSUM with $k = 0.5$ (dotted curve connecting diamonds) are also presented in the figure, along with their control limits of 9.376 and 6.000 by dashed and dotted horizontal lines, respectively. Note that $\lambda = 0.05$ is used in MEWMA to make it robust to this non-normal data. From the plot, it can be seen that the MSEWMA chart passes control limit at around the 187th observation and it remains above the control limit for a while until the 195th observation. This excursion suggests that a marked step-change has occurred. The process may be adjusted after (or near) the end of this dataset and thus the MSEWMA charting statistics soon drop below the control limit. In comparison, the MEWMA chart does not give any signal until the 191st observation and the ARCUSUM statistics remain below the control limit throughout.

Finally, we apply the proposed MSEWMA chart to the aluminum smelter process dataset (kindly supplied to us by Peihua Qiu). The dataset contains five variables, the contents of SiO_2, Fe_2O_3, MgO, CaO, and Al_2O_3 and is comprised of 185 vectors. Similar to Qiu and Hawkins (2001), we use the first 95 vectors to estimate the parameters and the others for the test. Also, we first pre-whiten the dataset into residual vectors because the original measurements are substantially autocorrelated. Readers may refer to Section 4.3 of Qiu (2008) for the specifics of the pre-whiten process. Then the three charts with the same parameters as in the AEC example are applied to these residual vectors. Figure 7 shows the resulting MSEWMA, MEWMA and ARCUSUM charts along with their control limits of 13.636, 12.938 and 9.45.

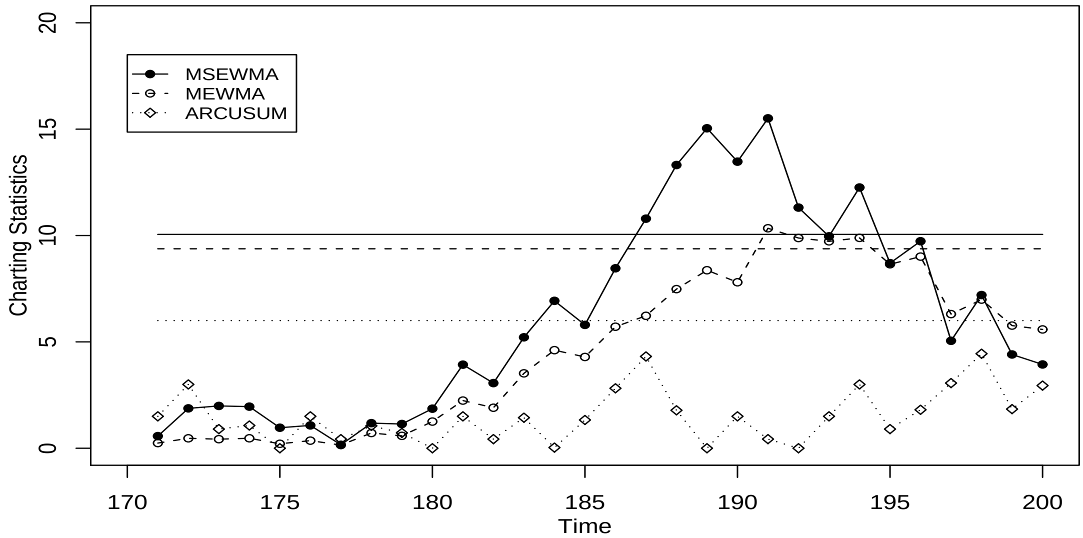

Figure 6: The MSEWMA, MEWMA and ARCUSUM control charts for monitoring the aluminium electrolytic capacitor process. The solid, dashed and dotted horizontal lines indicate their control limits, respectively.

The MSEWMA triggers an alarm at around the 133th observation (after the 38th test observation is collected) and remains above its control limit for ten observations. The MEWMA chart also gives a signal at the 146th observation. Once again, the ARCUSUM chart of the first antirank fails to signal, which is consistent with the analysis in Qiu and Hawkins (2001). Perhaps some other combined versions, such as the first-and-last-combined antiranks used in Qiu and Hawkins (2001) are more appropriate for this dataset. However, the choice of antiranks depends heavily on the shift directions, which are usually unknown before monitoring. Therefore, designing an effective ARCUSUM involves more undetermined parameters than MSEWMA, and the MSEWMA should be a reasonable alternative for non-multinormal processes by taking its convenience and robustness into account.

5 Concluding Remarks

In this paper, we propose a multivariate nonparametric control scheme. Instead of estimating the mean, $\boldsymbol{\mu}_0$, and covariance matrix, $\boldsymbol{\Sigma}_0$, from the IC historical dataset

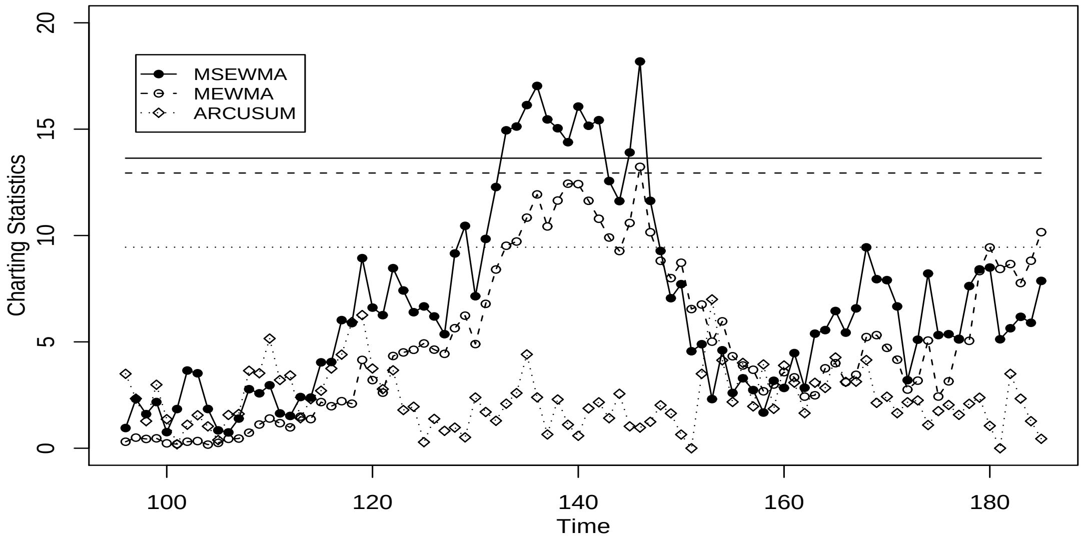

Figure 7: The MSEWMA, MEWMA and ARCUSUM control charts for monitoring the aluminum smelter process. The solid, dashed and dotted horizontal lines indicate their control limits, respectively.

as in the construction of the conventional MEWMA chart, we propose to obtain the multivariate affine-equivariant median, $\boldsymbol{\theta}_0$, first and an associated transformation matrix, $\mathbf{A}_0$. Then, the proposed chart is developed based on integrating Randle's (2002) powerful transformation-retransformation sign test with EWMA process monitoring. This nonparametric chart shares some appealing properties with its parametric counterpart MEWMA: (1) its computation speed is fast; (2) it is affine-invariant; (3) for the distributions with elliptical directions, the charting sequence is a Markov chain process and correspondingly its IC ARL can be easily calculated through a one-dimensional Markov chain model. In comparison with MEWMA, it is not only much more robust in IC performance, but it is also generally more sensitive to the small and moderate shifts in location parameters for skewed and heavy-tailed multivariate observations. In many cases, the improvement is quite remarkable. The drawback of the MSEWMA chart, which is common to almost all rank-based nonparametric charts, is that it is not as efficient as MEWMA for very large shifts because it only uses the direction of observations from the origin. Certainly, this disadvantage is mainly due to the *trade-off* between robustness and sensitivity.

There are a number of issues not thoroughly addressed here that could be topics of future research. First, this paper focuses on Phase II monitoring only and presumes that all of the historical observations used for estimating the IC parameters are i.i.d. In practical applications, there is no such assurance. Hence, it requires much future research to extend our method to Phase I analysis, in which detection of outliers or change-points in a historical dataset would be of interest. Second, the performance of both MEWMA and MSEWMA is affected by the amount of data in the reference dataset. Thus, determination of required Phase I sample sizes to reduce the effects of estimated parameters and a general recommendation are needed. Third, the current version of the proposed chart is designed for detecting location shifts only. In real multivariate processes, changes affecting only the location vector are very rare. We believe that, after certain modifications, the proposed method should be able to handle cases in which monitoring both the location and covariance structure is of interest (cf., e.g., Huwang et al. 2007). Much future research is also needed to propose a self-starting version of the MSEWMA chart (cf., Hawkins and Maboudou-Tchao 2007). Finally, statistical monitoring and surveillance of high-dimensional data stream involving dozens or even hundreds of variables have been widely recognized as important and critical tools for detection of abnormal behavior and quality improvement. Note that in such cases, the estimated parameters problem would become more prominent because estimated covariance matrices would be even rank-deficient if $m_0 < p + 1$. It is of interest to study the performance of MSEWMA in such monitoring environments and to investigate how to improve its efficiency by using some variable selection techniques (Zou and Qiu 2009).

Acknowledgement

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. This work was completed when Zou was a Post-doctoral Fellow at the Hong Kong University of Science and Technology, whose hospitality is appreciated and acknowledged. This research was supported by the NNSF of China Grants 10771107 and RGC Competitive Earmarked Research Grants 620508 and 620707.

Appendix: IC ARL Calibrations of MSEWMA via a Markov Chain Model

Based on Propositions 2 and 3, without loss of generality, we assume that $\boldsymbol{x}_i$ are i.i.d. standard p-dimensional multinormal variables, which will facilitate the derivation of the transition probabilities as shown later. The Markov chain model described here can be regarded as an extension of Brook and Evans (1972) and Runger and Prabhu (1996) to MSEWMA, and hence we only briefly describe the approximation method, but highlight some necessary modifications and formulas. For more details on the Markov chain approximation for the conventional EWMA and MEWMA charts, readers may refer to Lucas and Saccucci (1990) and again Runger and Prabhu (1996).

A one-dimensional Markov chain is used to approximate the IC ARL. Define the $(m+1)$ by $(m+1)$ transition probability matrix, $\mathbf{P} = (p_{ij})$, where the element p_{ij} denotes the probability of a transition from state i to j, and $(m+1)$ is the number of transition states. Denote $g = 2[L\lambda/(p(2-\lambda))]^{\frac{1}{2}}/(2m+1)$. Now, we have for $i = 0, 1, 2, \cdots$, where m and j are not equal to 0, that

$$
\begin{aligned}
p_{ij} &= \Pr\left\{(j-0.5)g < ||\lambda\boldsymbol{v}_t + (1-\lambda)\boldsymbol{w}_{t-1}|| < (j+0.5)g \big| ||\boldsymbol{w}_{t-1}|| = ig\right\} \\
&= \Pr\left\{(j-0.5)g < ||\lambda\boldsymbol{v}_t + (1-\lambda)ig\boldsymbol{u}|| < (j+0.5)g\right\} \\
&= \Pr\left\{(j-0.5)g/\lambda < ||\boldsymbol{v}_t + (1-\lambda)ig\boldsymbol{e}_p/\lambda|| < (j+0.5)g/\lambda\right\},
\end{aligned}
$$

where we use the arguments from the proof of Proposition 3 that the distribution of $\boldsymbol{w}_{t-1}$ given $||\boldsymbol{w}_{t-1}|| = ig$ is uniformly distributed on $S(||\boldsymbol{w}_{t-1}||)$, say as $ig\boldsymbol{u}$. The last equality comes from the fact that $\boldsymbol{v}_i$ and $\boldsymbol{u}$ are independent. Let $\xi = [(1-\lambda)ig/\lambda]$. By noting that $||\boldsymbol{v}_t|| = 1$, simple calculation yields that

$$
p_{ij} = \Pr\left\{(j-0.5)^2g^2/\lambda^2 < 1 + \xi^2 + 2\xi\boldsymbol{e}_p'\boldsymbol{v}_t < (j+0.5)^2g^2/\lambda^2\right\}.
$$

Then, for $i = 0$ and $j = 1, \ldots, m$,

$$
p_{0j} = I_{\{1\in[(j-0.5)^2g^2/\lambda^2,(j+0.5)^2g^2/\lambda^2]\}},
$$

where $I_{\{\cdot\}}$ is the indicator function. For $i, j = 1, \ldots, m$, we have

$$
p_{ij} = G\left(\frac{1}{2}[(j+0.5)^2g^2/\lambda^2 - 1 - \xi^2]/\xi\right) - G\left(\frac{1}{2}[(j-0.5)^2g^2/\lambda^2 - 1 - \xi^2]/\xi\right),
$$

where $G(\cdot)$ denotes the c.d.f. of the random variable, $y_1/\sqrt{y_1^2+\ldots,y_p^2}$ and $y_i, i = 1,\ldots,p$ are i.i.d. from the standard normal distribution. It is easy to verify that $G(\cdot)$ has the following closed form,

$$G(x) = \begin{cases} 1 - \frac{1}{2}F_{p-1,1}\left(\frac{x^{-2}-1}{p-1}\right), & \text{for} \quad x \geq 0, \\ \frac{1}{2}F_{p-1,1}\left(\frac{x^{-2}-1}{p-1}\right), & \text{for} \quad x < 0, \end{cases}$$

where $F_{p-1,1}(\cdot)$ is the c.d.f. of the F distribution with $(p-1,1)$ degrees of freedom. The remaining case is that for $j = 0$,

$$p_{ij} = G\left(\frac{1}{2}[0.25g^2/\lambda^2 - 1 - \xi^2]/\xi\right).$$

Finally, the IC ARL can then be evaluated by

$$\text{ARL} = \boldsymbol{e}'_{m+1}(\mathbf{I}_{m+1} - \mathbf{P})^{-1}\mathbf{1},$$

where $\mathbf{1}$ is a vector of ones.

References:

Brook, D., and Evans, D. A. (1972), "An Approach to the Probability Distribution of CUSUM Run Length," *Biometrika*, 59, 539–549.

Brown, B. M. (1983), "Statistical Uses of the Spatial Median," *Journal of the Royal Statistical Society, Series B*, 45, 25–30.

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.

Chakraborty, B., Chaudhuri, P., and Oja, H. (1998), "Operating Transformation Retransformation on Spatial Median and Angle Test," *Statistica Sinica*, 8, 767–784.

Croisier, R. B. (1988), "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes," *Technometrics*, 30, 243–251.

Eaton, M. L. (1983), *Multivariate Statistics*, Wiley, New York.

Gower, J. S. (1974), "Algorithm AS 78. The Mediancentre," *Applied Statistics*, 23, 466–470.

Hackel, P., and Ledolter, J. (1991), "A Control Chart Based on Ranks," *Journal of Quality Technology*, 23, 117–124.

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

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

Healy, J. D. (1987), "A Note on Multivariate CUSUM Procedure," *Technometrics*, 29, 409–412.

Hettmansperger, T. P., and Randles, R. H. (2002), "A Practical Affine Equivariant Multivariate Median," *Biometrika*, 89, 851–860.

Hettmansperger, T. P., and McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, Arnold, London.

Huwang, L., Yeh, A. B., and Wu, C. (2007), "Monitoring Multivariate Process Variability for Individual Observations," *Journal of Quality Technology*, 39, 258–278.

Jensen, W. A., Jones, L. A., Champ, C. W., and Woodall, W. H. (2006), "Effects of Parameter Estimation on Control Chart Properties: A Literature Review," *Journal of Quality Technology*, 38, 349–364.

Jones, L. A. (2002), "The Statistical Design of EWMA Control Charts with Estimated Parameters," *Journal of Quality Technology*, 34, 277–288.

Jones, L. A., Champ, C. W., and Rigdon, S. E. (2001), "The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters," *Technometrics*, 43, 156–167.

Liu, R. (1995), "Control Charts for Multivariate Processes," *Journal of the American Statistical Association*, 90, 1380–1388.

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

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

Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Scheme Properties and Enhancements," *Technometrics*, 32, 1–29.

Mardia, K. V. (1970), "Measures of Multivariate Skewness and Kurtosis With Applications," *Biometrika*, 57, 519–530.

Montgomery, D. C. (2005), *Introduction to Statistical Quality Control, 6th ed.*, John Wiley & Sons, New York.

Oja, H. (1999), "Affine Invariant Multivariate Sign and Rank Tests and Corresponding Estimates: A Review," *Scandinavian Journal of Statistics*, 26, 319–343.

Oja, H., and Randles, R. (2004), "Multivariate Nonparametric Tests," *Statistical Science*, 19, 598–605.

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

Prabhu, S. S., and Runger, G. C. (1997), "Designing a Multivariate EWMA Control Chart," *Journal of Quality Technology*, 29, 8–15.

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

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

Qiu, P., and Hawkins, D. M. (2003), "A Nonparametric Multivariate CUSUM Procedure for Detecting Shifts in All Directions," *Journal of the Royal Statistical Society, Series D*, 52, 151–164.

Quesenberry, C. P. (1995). "On Properties of Q Charts for Variables". *Journal of Quality Technology*, 27, pp. 184–203.

Randles, R. H. (1989), "A Distribution-Free Multivariate Sign Test Based on Interdirections," *Journal of the American Statistical Association*, 84, 1045–1050.

Randles, R. H. (2000), "A Simpler, Affine Invariant, Multivariate, Distribution-Free Sign Test," *Journal of the American Statistical Association*, 95, 1263–1268.

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, 1701–1706.

Stoumbos, Z. G., and Jones, L. A. (2000), "On the Properties and Design of Individual Control Charts Based on Simplicial Depth," *Nonlinear Studies*, 7, 147–178.

Stoumbos, Z.G., Reynolds, M. R., 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.

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.

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.

Tyler, D. E. (1987), "A Distribution-Free M-Estimator of Multivariate Scatter," *The Annals of Statistics*, 15, 234–251.

Woodall, W. H., and Montgomery, D. C. (1999), "Research Issues and Ideas in Statistical Process Control," *Journal of Quality Technology*, 31, 376–386.

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

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

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