A Variable-Selection-Based Multivariate EWMA Chart for Process Monitoring and Diagnosis

WEI JIANG

Shanghai Jiaotong University, Shanghai 200052, China

KAIBO WANG

Tsinghua University, Beijing 100084, China

FUGEE TSUNG

The Hong Kong University of Science and Technology, Hong Kong

Fault detection and root cause identification are both important tasks in Multivariate Statistical Process Control (MSPC) for improving process and product quality. Most traditional control charts, including Hotelling's T^2 chart and the Multivariate Exponential Weighted Moving Average (MEWMA) chart, separate the two tasks into independent and successive procedures by signaling the existence of process faults followed by auxiliary methods to locate root causes. This paper proposes an integrated procedure, a Variable-Selection-based MEWMA (VS-MEWMA) chart, for multivariate process monitoring and fault diagnosis by utilizing dimensionality reduction techniques. The VS-MEWMA chart first locates potentially out-of-control variables via variable selection and then deploys such information in the monitoring statistics with the reduction in dimensionality providing increased sensitivity to out-of-control conditions. When a signal is given, the algorithm also identifies the suspected variables for further root cause diagnosis. Both numerical simulations and real examples are presented to illustrate the performance of the proposed chart, as well as design guidelines.

Key Words: Average Run Length; Fault Diagnosis; Forward Selection; MEWMA; Multivariate Statistical Process Control.

Introduction

WITH the rapid advances in sensing and metrology technology, many companies and organizations nowadays have easy access to data collected

Dr. Jiang is a Professor in Antai College of Economics and Management. He is a senior member of ASQ. His email address is jiangwei@sjtu.edu.cn.

Dr. Wang is an Associate Professor in the Department of Industrial Engineering. His email address is kbwang@tsinghua.edu.cn.

Dr. Tsung is a Professor in the Department of Industrial Engineering and Logistics Management. He is a Fellow of ASQ. His email address is season@ust.hk.

in production, business transactions, and service operations. These data streams often contain multiple or even dozens of variables and carry very useful knowledge and information, which can be extracted through modeling, characterization, monitoring, and forecasting. Statistical monitoring and surveillance of such data streams have been widely recognized as important and critical tools for detection of abnormal behavior and quality improvement. For example, in semiconductor and electronic manufacturing, high-dimensional multivariate data can be automatically acquired by sensors and other equipment; statistical process control techniques have been used routinely for on-line process control to improve process capability through variation reduction (Spanos (1992), Mason and Young (2002)). In service indus-

tries, many companies are applying data mining and surveillance tools to understand customer profiles with high-dimensional attributes to detect fraudulent behavior in real time, which is referred to as activity monitoring (Jiang et al. (2007)). For example, for credit card or insurance fraud detection, a collection of thousands of variables and transactions is closely monitored by a typical credit card company on a daily basis. In public health management, surveillance methods have been developed for timely detection and prevention of various types of adverse health events so that mitigation strategies can be initiated promptly. An example of public surveillance is an increased birth rate of babies with congenital malformations, which was especially apparent during the thalidomide tragedy in the early 1960s (Sonesson and Bock (2003)).

The development of Multivariate Statistical Process Control (MSPC) techniques for process monitoring and surveillance is continuously challenged by two basic requirements: the capability to detect process faults quickly and the ability to locate shifted variables accurately. The first requirement stresses the sensitivity of an MSPC scheme, while the second concerns the diagnostic capability of the scheme. However, lower sensitivity in fault detection would provide more out-of-control observations for diagnosis, offering improved diagnostic results. These challenges become crucial when the dimensionality of data streams is higher, which often prevents the adoption of well-known MSPC tools in practice due to the "curse of dimensionality" (Jiang and Tsui (2008), Wang and Tsung (2008), Zhu and Jiang (2009)).

However, most MSPC research has focused separately on the two objectives of detection and diagnosis. One solution for the "curse of dimensionality" lies in dimensionality-reduction algorithms that can be used before control charts. If the key information about process status can be summarized by a few variables, monitoring these variables and searching for process faults among a reduced set of variables is expected to be more effective than monitoring the whole set. Conventional dimensionality reduction techniques, such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA), have been adopted for the purpose of process monitoring and control. PCA- or ICA-based MSPC schemes first identify a small set of variables that are linear combinations of the original variables and then monitor these derivative statistics against pos-

sible process shifts (e.g., Mastrangelo et al. (1996), Runger and Alt (1996), Ding et al. (2006)).

A potential disadvantage of these methods is the use of in-control data to identify the variables to be monitored, possibly excluding variables with important information about shifts to out-of-control conditions. Moreover, the derived statistics may not yield any meaningful interpretations of the physical process and alarms. Once triggered from the derived statistics, the alarms must be checked against diagnosis procedures to identify the root causes of the alarms. For this purpose, root-cause identification algorithms have been studied in MSPC research. For example, Dunia et al. (1996) investigated the effectiveness of PCA for the identification of faulty sensors. Hawkins (1991), Hawkins (1993), and Hawkins and Maboudou-Tchao (2008) proposed a regression-adjusted method that regresses downstream variables with respect to upstream variables. The residuals are then checked to identify possible shifts. Mason et al. (1995) proposed another way to check the contribution of each variable based on the MYT-decomposition method proposed in Mason et al. (1995) and Mason and Young (1999). All possible combinations of variables are screened, and the most significant effects are determined based on statistical tests. However, there is still a lack of effective diagnosis methods for certain control charts that accumulate recent process information. For example, when an out-of-control alarm is triggered by an MEWMA or multivariate cumulative sum (MCUSUM) (Pignatiello and Runger (1990)) chart, it is difficult to reasonably associate the alarm to any specific variables.

Recently, Wang and Jiang (2009) and Zou and Qiu (2009) independently proposed using variable selection (VS) methods to screen for suspicious variables and monitor the selected variables in the reduced space. Wang and Jiang (2009) developed a Shewhart-type, variable-selection-based multivariate (VS-MSPC) chart that uses a forward variable selection (FVS) method to select suspicious variables. The LASSO-based EWMA (LEWMA) chart developed in Zou and Qiu (2009) used the exponentially weighted moving average (EWMA) statistic to accumulate recent observations and applied the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm for VS.

The basic principle underlying these methods is that, for most processes, the possibility of all variables in a high-dimensionality process shifting simul-

taneously is low. Instead, process changes are usually manifested in only one or a small fraction of the observed variables. Without assuming any *a priori* knowledge of the relationships between the hidden causes and the monitored variables, these control charts can adaptively screen suspicious variables for monitoring so that the dimensionality of monitoring is reduced. Additionally, whenever an alarm is triggered, the VS procedure can suggest potential out-of-control variables automatically.

The integration of detection and diagnosis makes the VS-MSPC charts appealing in reducing dimensionality and improving control chart performance. However, unlike the LEWMA chart, the VS-MSPC chart is a Shewhart-type chart that uses the information from only the current process observation. In this article, we extend the VS-MSPC chart by using accumulated recent information for VS and process monitoring. A multivariate EWMA procedure is developed to accumulate recent observations, which is expected to make mean estimation and VS more accurate and, hence, benefit the control chart performance. Unlike the method used in Zou and Qiu (2009)'s method, the proposed scheme locates potential out-of-control variables via a FVS algorithm, which is popular in industrial quality control practice, and then deploys such information in the monitoring statistics. The VS algorithm also suggests a reduced subset of suspicious variables for further root cause diagnosis.

The remainder of this paper is organized as follows. As a stepping-stone, we first give a brief review of the VS-MSPC chart to motivate our new variable-selection-based MEWMA (VS-MEWMA) chart. Then we introduce the VS-MEWMA chart, which uses accumulated information for both VS and process monitoring. The statistical performance of the proposed scheme is studied, followed by a discussion of properties and design issues of the proposed control chart. A real example from a footwear manufacturing process is presented to illustrate the application of our method for fault detection and isolation. Finally, the paper concludes with suggestions for future research.

Review of the VS-Based Multivariate Control Chart

In this section, we give a brief review of the VS-MSPC chart proposed by Wang and Jiang (2009) to motivate the VS-MEWMA chart. In essence, the VS-MSPC chart is derived from a generalized likelihood

ratio (GLR) statistic for a hypothesis test. Let the p-dimensional measurement vector at time t be represented by $\mathbf{y}_t$, which is centered (without loss of generality) by subtracting the in-control mean vector, so that $\mathbf{y}_t \sim N_p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}_t = \mathbf{0}_p$ for $t = 1, 2, \ldots, \tau$ and $\boldsymbol{\mu}_t \neq \mathbf{0}_p$ for $= \tau + 1, \ldots$, where τ is unknown. Using the sequential observations $\{\mathbf{y}_t\}$, our task is to trigger an alarm when a mean change is detected and suggest which of the monitored variables have shifted.

We first consider testing the hypotheses H_0: $\boldsymbol{\mu}_t = \mathbf{0}_p$ and H_1: $\boldsymbol{\mu}_t \neq \mathbf{0}_p$. Using only the current observation $\mathbf{y}_t$, negative two times the logarithm of the generalized likelihood ratio is $\Lambda(\mathbf{y}_t) = \mathbf{y}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{y}_t - \min_{\boldsymbol{\mu}_t}((\mathbf{y}_t - \boldsymbol{\mu}_t)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t))$, which is the Hotelling's T^2 statistic (since the second term is zero), and the corresponding control chart signals when an upper control limit is exceeded.

Let

$$L_0 = \min_{\boldsymbol{\mu}_t}((\mathbf{y}_t - \boldsymbol{\mu}_t)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t)). \qquad (1)$$

As the minimum of Equation (1) is achieved when $\boldsymbol{\mu}_t = \mathbf{y}_t$, Hotelling's T^2 statistic can be viewed as using $\mathbf{y}_t$ as an estimate of the true process mean (Jiang and Tsui (2008)). However, in general, $\mathbf{y}_t$ is not the best estimator for $\boldsymbol{\mu}_t$ at step t. When more observations, $\mathbf{y}_t, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots$, are available, the average or weighted average of these observations could be a better estimate than $\mathbf{y}_t$ alone. Another opportunity for improvement is based on the observation that many special causes do not shift all of the variables being monitored. Instead, a common phenomenon observed in practice is that a subset of variables, which is dominated by a common latent physical mechanism or component, deviates from its normal condition due to abnormal changes in a common mechanism or component (Mastrangelo et al. (1996), Choi et al. (2006)). Therefore, we can infer that components of the vector $\boldsymbol{\mu}_t$ that are estimated to be close to zero are in fact zero.

In order to identify the components that are significant in the process mean estimate, Wang and Jiang (2009) proposed to modify the optimization in Equation (1) by limiting the total number of nonzero elements in $\boldsymbol{\mu}_t$, which corresponds to the number of suspicious out-of-control variables; that is,

$$\begin{cases} \min_{\boldsymbol{\mu}_t}((\mathbf{y}_t - \boldsymbol{\mu}_t)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t)) \\ \text{s.t.} \quad \sum_j I(|\mu_{t(j)}| \neq 0) \leq s, \end{cases} \qquad (2)$$

where s is a prescribed integer upper bound on the number of nonzero coefficients; $\mu_{t(j)}$ is the jth element of vector $\boldsymbol{\mu}_t$ and $I(|\mu_{t(j)}| \neq 0)$ is an indicator function that takes 1 when $|\mu_{t(j)}| \neq 0$ and 0 otherwise. Let $\boldsymbol{\Sigma}^{-1} = \mathbf{R}^{\mathrm{T}}\mathbf{R}$ be the Cholesky decomposition of $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{z}_t = \mathbf{R}\mathbf{y}_t$. Equation (2) can be equivalently expressed as the following constrained least square estimation problem:

$$\begin{cases} \min_{\boldsymbol{\mu}_t}((\mathbf{z}_t - \mathbf{R}\boldsymbol{\mu}_t)^{\mathrm{T}}(\mathbf{z}_t - \mathbf{R}\boldsymbol{\mu}_t)) \\ \text{s.t.} \quad \sum_j I(|\mu_j| \neq 0) \leq s, \end{cases} \quad (3)$$

where $\mathbf{R}$ can be regarded as a design matrix and $\mathbf{z}_t$ as a response variable. This constrained optimization causes estimated coefficients that are close to zero in the regression model to be set to *exactly* zero.

Let $\boldsymbol{\mu}_t^*$ denote the solution to Equation (3) (which depends on the parameter s). Generically, we assume that only those variables with nonzero coefficients are potential out-of-control variables and will be monitored for possible process shifts. Although this assumption may not hold all the time in practice, especially when s is chosen unwisely small, it holds potential for productive dimensionality reduction and thereby improved detection sensitivity by identifying a set of variables with the greatest information about a potential shift.

Wang and Jiang (2009) adopted the FVS algorithm to find $\boldsymbol{\mu}_t^*$, the solution for Equation (3). A VS-MSPC chart is developed by substituting the estimated $\boldsymbol{\mu}_t^*$ into Equation (1) as follows:

$$\Lambda(\mathbf{y}_t) = 2\mathbf{y}_t\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_t^* - \boldsymbol{\mu}_t^{*\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_t^*,$$

which signals when the upper control limit is exceeded. Appendix A further shows that monitoring this statistic is equivalent to monitoring

$$\Lambda_1(\mathbf{y}_t) = \mathbf{y}_t^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_t^*, \quad (4)$$

which is the inner product of the current observation $\mathbf{y}_t$ and the constrained mean estimate $\boldsymbol{\mu}_t^*$ with respect to the variance-covariance matrix $\boldsymbol{\Sigma}$. In fact, if $s = p$, with all variables selected in Equation (3), then $\Lambda(\mathbf{y}_t)$ and $\Lambda_1(\mathbf{y}_t)$ reduce to Hotelling's T^2 statistic. If $s < p$, intuitively, $\Lambda_1(\mathbf{y}_t)$ measures a kind of "match" or "agreement" between the observation and the constrained mean estimate.

A VS-MEWMA Chart

The previous section illustrates the principle of the VS-MSPC chart. Although the FVS procedure is

used to narrow down the set of variables with suspected mean shifts, the observations are random and subject to noise so that an estimate of $\boldsymbol{\mu}_t$ based on only one observation can be unreliable. We now modify the objective function in Equation (3) and propose a new way to estimate the process mean. On one hand, using the information from several recent observations may benefit the estimation accuracy. On the other hand, observations too ancient may not accurately represent the current status of a process since the process may have changed. Therefore, a reasonable treatment is to put more emphasis on recent data and give less weight to older data, which leads to an EWMA-based VS method.

VS via Exponential Smoothing

As shown in Table 1, we propose to discount observations in an exponentially decaying way by a factor of $\omega_i = \lambda(1-\lambda)^{t-i} / \sum_{j=1}^{t} \lambda(1-\lambda)^{t-j}$, where $0 < \lambda \leq 1$ is the smoothing parameter. Based on the weighting scheme, we modify Equation (3) as follows:

$$\begin{cases} \min_{\boldsymbol{\mu}_t} \left(\sum_{i=0}^{t} \omega_i(\mathbf{z}_i - \mathbf{R}\boldsymbol{\mu}_t)^{\mathrm{T}}(\mathbf{z}_i - \mathbf{R}\boldsymbol{\mu}_t) \right) \\ \text{s.t.} \quad \sum_j I(|\mu_{t(j)}| \neq 0) \leq s, \end{cases} \quad (5)$$

It should be noted that the added constraint or penalty in Equation (5) is one of the options one may consider for VS. Using an L_1-type penalty instead gives the LASSO algorithm of Zou and Qiu (2009).

Equation (5) can be expressed in a recursive form. Let

$$\mathbf{w}_t = (1-\lambda)\mathbf{w}_{t-1} + \lambda\mathbf{y}_t, \quad (6)$$

with $\mathbf{w}_0 = \mathbf{0}_p$. Appendix B shows that, when $t \to \infty$, the solution of Equation (5) is equivalent to the solu-

TABLE 1. Weight Allocation of Recent Data for Variable Selection and Shift Estimation

Time	Response	Covariate	Weight
t	$\mathbf{z}_t$	$\mathbf{R}$	ω_t
$t-1$	$\mathbf{z}_{t-1}$	$\mathbf{R}$	ω_{t-1}
$t-2$	$\mathbf{z}_{t-2}$	$\mathbf{R}$	ω_{t-2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$

tion of the following objective function:

$$\begin{cases} \min_{\boldsymbol{\mu}_t}((\mathbf{w}_t - \boldsymbol{\mu}_t)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{w}_t - \boldsymbol{\mu}_t)) \\ \text{s.t.} \quad \sum_j I(|\mu_{t(j)}| \neq 0) \leq s, \end{cases} \quad (7)$$

That is, we can simply accumulate observations in an EWMA and obtain $\boldsymbol{\mu}_t^*$ by minimizing Equation (7); the solution could be viewed as the best match with the running, weighted average of the observations.

In solving Equation (7), without considering statistical significance, increasing the number of nonzero coefficients decreases the objective function, so that the constraint imposed by the choice of s is always binding. Therefore, we fix the total number of nonzero variables and search for the optimal value of $\boldsymbol{\mu}_t$ using an algorithm that resembles the usual FVS procedure. We first select one variable from all p candidates to minimize Equation (7), and then by fixing the first chosen variable we search for the second variable, together with preceding selection(s), to minimize Equation (7). The selection procedure is repeated until s variables are selected. Detailed steps are outlined in Appendix B. By limiting the number of selection steps, it is guaranteed that s nonzero coefficients are attained in $\boldsymbol{\mu}_t^*$.

Solving Equation (7) is expected to give a more robust estimate of process mean than solving Equation (3), since the former takes recent information into consideration and reduces the randomness. Charting statistics developed based on $\boldsymbol{\mu}_t^*$ are also expected to be more powerful in detecting shifts.

A VS-MEWMA Chart

The solution, $\boldsymbol{\mu}_t^*$, obtained from the above procedure can be viewed as a more robust and reliable estimate of the process mean. Once it is obtained, we can plug it into Equation (4) and evaluate its "agreement" with process observations. However, in Equation (4), $\mathbf{y}_t$ only contains information from the latest observation. In order to further improve sensitivity of the VS-MSPC method to small changes, more observations could be accumulated and incorporated into the monitoring statistics of $\mathbf{y}_t$ in Equation (4). Therefore, we use the EWMA statistic $\mathbf{w}_t$ to replace $\mathbf{y}_t$ in Equation (4), and obtain the following VS-MEWMA statistic:

$$M_t = \boldsymbol{\mu}_t^{*\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{w}_t. \quad (8)$$

The corresponding chart signals when $M_t > c$, an upper control limit chosen for desired performance.

As the above statistic M_t integrates VS into a directional MEWMA chart, we name this chart a VS-MEWMA chart. It should be noted that the VS-MEWMA chart inherits the diagnostic capability of the VS-MSPC chart. Whenever an out-of-control alarm is triggered, it is natural to use the variables having nonzero coefficients as the basis for further identification of root causes.

Zou and Tsung (2008) introduced a directional MEWMA chart for monitoring multistage processes. In their directional MEWMA chart statistic, prior knowledge of the *shift direction* and exponentially smoothed observations are combined to improve the performance for fault detection. However, the control chart defined in Equation (8) does not require any prior information about process shifts or their directions. Our experiments have also shown that incorporating the estimated shift $\boldsymbol{\mu}_t^*$ rather than the estimated shift *direction* further improves shift detection.

Comparison with the LEWMA Chart

The VS-MEWMA chart is similar to the LEWMA chart proposed by Zou and Qiu (2009). The LEWMA chart also starts from the EWMA statistic, $\mathbf{w}_t$. At step t, it calculates q LASSO estimators $\hat{\boldsymbol{\mu}}_{t,\tilde{\gamma}_{m_k^{\text{last}}}}$, $k = 1, 2, \ldots, q$, that minimize Equation (7) with the L_0 penalty replaced by a L_1-type adaptive LASSO penalty. The LEWMA chart signals when

$$Q_t = \max_{k=1,\ldots,q} \frac{W_{t,\tilde{\gamma}_{m_k^{\text{last}}}} - E\left(W_{t,\tilde{\gamma}_{m_k^{\text{last}}}}\right)}{\sqrt{\mathrm{Var}\left(W_{t,\tilde{\gamma}_{m_k^{\text{last}}}}\right)}} > c', \quad (9)$$

where, asymptotically,

$$W_{t,\gamma} = \frac{2 - \lambda}{\lambda} \frac{\left(\mathbf{y}_t^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}_\gamma\right)^2}{\hat{\boldsymbol{\mu}}_\gamma^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}_\gamma}.$$

The parameter q is critical to the algorithm; it specifies the number of alternatives to test at each step. Details about the LEWMA chart can be found in Zou and Qiu (2009).

Both LEWMA and VS-MEWMA charts use the EWMA statistic for VS and a MEWMA-type charting statistic for detecting process shifts. However, apart from different penalty functions and corresponding algorithms considered in VS, there are two major differences between them:

(a) The LEWMA chart uses a data-driven method to search along the VS path for potential out-

of-control variables; the critical value, q, limits the *maximum* number of potential out-of-control variables. However, the VS-MEWMA chart uses a *fixed* number of potential out-of-control variables. (Its robustness against parameter misspecification will be presented later.) Therefore, the LEWMA chart is suitable for cases without any shift information or with little knowledge of shift patterns, while the VS-MEWMA chart is expected to be suitable for cases with better understanding of the number of potential out-of-control variables. If the VS-MEWMA is extended to include *maximally s* values at each step, the two charts may be similar. Since LEWMA involves a complicated algorithm and pursues an exhaustive search among q alternatives along the VS path, the form of the VS-MEWMA chart statistic may look simpler to practitioners and be less complex to implement.

(b) The LEWMA chart monitors the maximum of q standardized statistics $W_{t,\tilde{\gamma}_{m_k^{\text{last}}}}$ ($k = 1, 2, \ldots, q$), where the mean and standard deviation of $W_{t,\tilde{\gamma}_{m_k^{\text{last}}}}$ are calculated from numerically simulated data. However, the distribution of the standardized statistic is not identical for all $k = 1, 2, \ldots, q$, especially with respect to the right-side tail probability; this may result in a suboptimal allocation of the overall false alarm rate when simultaneously monitoring the q standardized statistics. The VS-MEWMA chart does not have this issue, since only one statistic is monitored. The performance of these two charts will be compared in later sections.

Performance Analysis of the VS-MEWMA Chart

With the VS-MEWMA chart, the smoothing parameter, λ, determines the weight to be assigned to recent observations for mean estimation. A small value of λ makes the estimation of $\boldsymbol{\mu}_t^*$ more stable, so that the selection of potential out-of-control components will not change as often as new observations arrive when the process is in control. On the other hand, a large value of λ causes $\boldsymbol{\mu}_t^*$ to be updated quickly by assigning more weight to recent observations, but the selection of variables becomes less stable when the process is in control. In the following, we shall study the effect of λ on the properties of the proposed control charts.

We first demonstrate the performance of the VS algorithm for correctly identifying shifted variables. Then we discuss the fault diagnostic capability of the control chart when an out-of-control signal is triggered. Finally, we compare the average run length (ARL) performance of the proposed method with conventional MEWMA charts and Hotelling's T^2 chart and discuss the robustness of the proposed methods against the selection of the parameter s.

In the simulation studies, the dimension p takes values of 10 and 50 to represent medium and high dimensions of a process, respectively. For simplicity, we used $\mathbf{0}_p$ and $\mathbf{I}_{p \times p}$ as the in-control mean vector and covariance matrix in the simulation. Although transformation of correlated variables to independent components has been extensively studied (see, e.g., Huwang et al. (2007), Gonzalez and Sanchez (2008), Hawkins and Maboudou-Tchao (2008), Wang and Jiang (2009), Chenouri et al. (2009)), the proposed VS-MEWMA is capable of monitoring processes with correlated variables without transformation. Examples with general variance-covariance matrices will be studied in later sections. We further assume that when a shift in one or more means occurs, it *equally* affects the level of all of the shifted variables. The number of shifted variables is denoted by p_0. Without any loss of generality, we apply the shift to the first p_0 variables. Therefore, for $p_0 = 2$, for example, $\mathbf{y}_t \sim N_p([\delta, \delta, 0, \ldots, 0]^{\mathrm{T}}, \mathbf{I})$ when the process is out-of-control. In this analysis, unless otherwise stated, the in-control ARL (denoted by ARL_0) is 200 and the simulation results are based on 10,000 replications for each value shown.

Performance of the VS Algorithm

From the EWMA statistic in Equation (6), we expect that the smoothing parameter λ would significantly affect the way that the VS works. Therefore, we first study the stability of the VS-MEWMA chart in capturing suspicious out-of-control variables using different values of λ. As a demonstration, a simulated process with 50 variables is run for 100 steps with no shift in the process mean vector. Further, we choose $s = p_0 = 2$. When using the algorithm in Equation (6) to identify suspicious variables, Figure 1 shows the indices of the two selected (nonzero) variables at each step when λ is set to different alternative values.

Figure 1(b) shows that, when $\lambda = 0.2$, the same variables persist as suspicious in successive steps (when there has been no shift) due to the inertia of
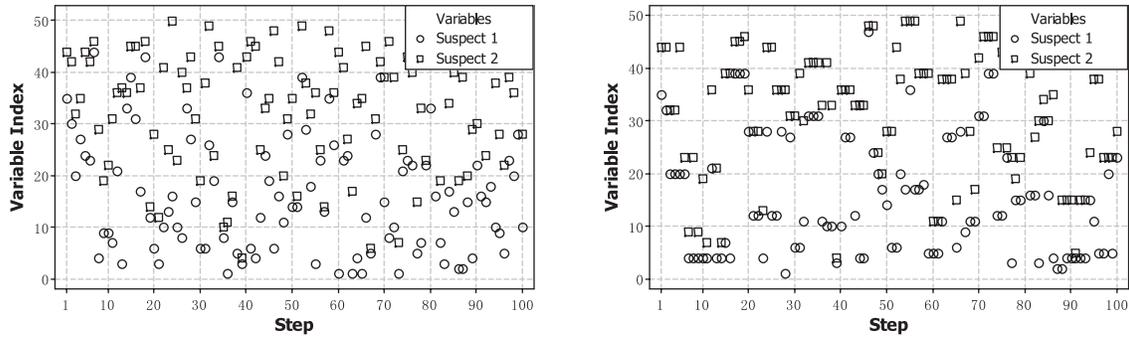
FIGURE 1. Stability of VS-MEWMA in Selecting Suspicious Variables (when there is no shift). (left) $\lambda = 1.0$; (right) $\lambda = 0.2$.

the EWMA statistics. This persistence effect would become stronger as λ becomes smaller.

In order to demonstrate the capability of the VS algorithm in identifying which variables have shifted, we add a shift of magnitude 1.0 to the first two variables, starting from the first step, and draw similar plots in Figure 2. When $\lambda = 1.0$, the selected suspicious variables are often different from the true shifted variables, although the shifted variables are slightly more frequently identified. With a small shift size and a comparatively large value for λ, the algorithm is prone to incorrectly identify the shifted variables (Wang and Jiang (2009)). When λ decreases, the probability that the shifted variables are identified is improved. For example, when $\lambda = 0.2$, the shifted variables are correctly identified most of the time.

Although using a small value of λ could improve the probability of correctly identifying the shifted variables, that would (slightly) increase the ARL with a large shift, compared with using a larger value of λ that would be optimal for the large shift. Figure 3 shows the same set of graphs when a mean shift of magnitude 1.0 is introduced at the 41st step. When $\lambda = 0.2$, the VS algorithm first correctly identifies the first shifted variable at step 47 and both shifted variables at step 48. The delay in identifying the correct variables is due to the inertia of the EWMA statistic. Before the change point, the algorithm may pick up the wrong variables. When a shift happens at step 41, the algorithm needs several steps to switch the selection to the correct set since the most recent information receives rather small weights. The inertia would become smaller as λ is increased and less weight is assigned to recent observations, but the accuracy of the identification of the shifted variables would suffer, as the case with $\lambda = 1.0$ shows.

In summary, the smoothing parameter for VS, λ, is critical for maintaining the sensitivity as well as the consistency of VS. A large value of λ may appear to be desirable for a quick response to large mean shifts but would result in a degraded accuracy
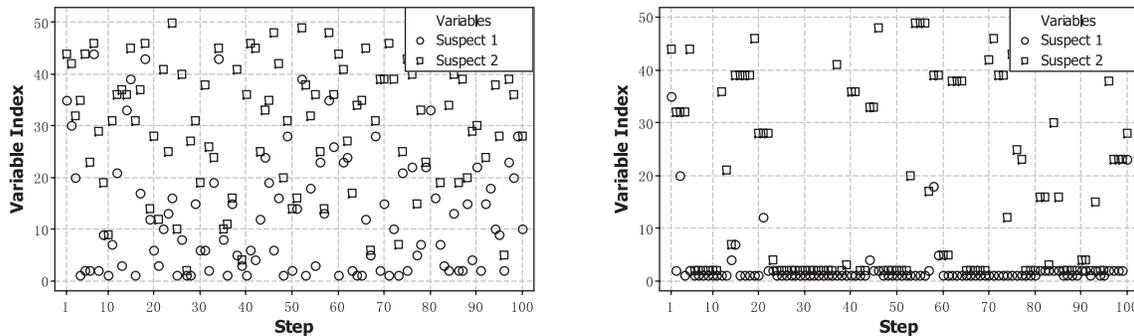

FIGURE 2. Stability of VS-MEWMA in Selecting Suspicious Variables (process shift of 1.0 in the 1st step). (left) $\lambda = 1.0$; (right) $\lambda = 0.2$.
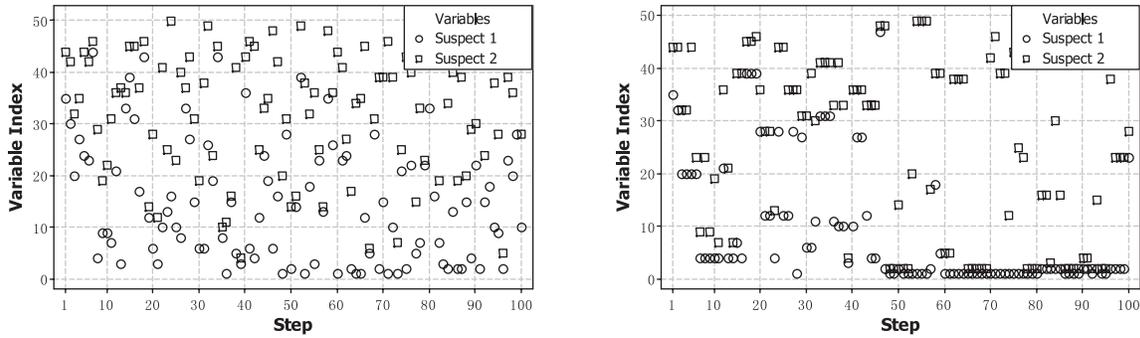
FIGURE 3. Stability of VS-MEWMA in Selecting Suspicious Variables (with a sustained shift of 1.0 in the first two variables beginning with step 41). (left) $\lambda = 1.0$; (right) $\lambda = 0.2$.

in identification of the shifted variables. In general, a trade-off is necessary in selecting appropriate λ values to balance quick response and accuracy. This also agrees with the property of the conventional EWMA statistics in process monitoring and control (Lucas and Saccucci (1990)).

Performance of Out-of-Control Variable Identification

The above study demonstrates the accuracy of the VS algorithm in identifying the correct variables for monitoring purposes (pre-signal). We now investigate the capability of the algorithm in identifying the correct variables after a signal. Let p_0 be the actual number of variables shifted in a process and τ be the change point (the first shifted observation). In the following, we simulated a process with $p = 10$, $p_0 = 2$, $\tau = 100$, and 200 replicates. Simulated runs with a signal before the change point were ignored. We calculated the probability of correct identification after a valid signal, denoted by P_c, by counting the proportion of times that the shifted variables were correctly identified using $s = 2$. The two shifted variables are treated as two opportunities for detection. If only one of the two shifted variables is correctly identified, the probability is treated as 50%. The results are reported in Table 2.

First, a general pattern can be observed; i.e., when the shift magnitude increases, P_c increases accordingly. For most of the cases, P_c is higher than 80% when $\delta \geq 1.5$. Although this result might be expected, it is not trivial. Since large shifts are identified more quickly, there are fewer out-of-control observations available for diagnostic analysis. Thus, it is reasonable but not obvious that the identification probability would improve. Second, for each shift magnitude, there seems to exist an optimal λ, λ^*,

that achieves the highest P_c. In addition, λ^* is generally smaller for small shifts and larger for large shifts. This shows that for small shifts, a small λ is helpful in accumulating more recent information, hence capturing shift information more accurately. For a large shift, since the corresponding run length is usually short, a large value of λ makes the VS procedure more sensitive and accurate in capturing shifts in a short period of time.

ARL Performance Study and Comparison with Existing Methods

In this study, we simulate 10- and 50-dimensional processes with $p_0 = 2$, that is,

$$\mathbf{y}_t \sim N_p([\delta, \delta, 0, \ldots, 0]^{\mathrm{T}}, \mathbf{I})$$

when the process is out-of-control. Since the VS-MEWMA chart is inevitably influenced by the values used for the smoothing parameters, for certain choices of λ the inertia in Equations (5) and (8) could be important. Therefore, the steady-state ARL with

TABLE 2. Probability of Correct Identification After a Valid Signal ($p = 10$, $p_0 = 2$, $s = 2$, $\mathrm{ARL}_0 = 200$)

		P_c (%)			
δ	$\lambda =$	0.05	0.1	0.2	0.4
0.2		47.0	37.5	32.5	26.5
0.6		67.3	66.8	68.3	54.3
1.0		78.0	82.5	82.0	74.0
1.5		79.8	87.8	84.5	84.0
2.0		86.0	86.0	91.8	87.3
3.0		85.8	88.0	90.8	93.3
4.0		88.5	92.0	92.3	94.0
5.0		89.3	91.3	94.5	97.5

a burn-in length of 100 steps (shifts were added at the 101st step) was used for evaluating control chart performance. Each ARL was calculated using 10,000 replicates. The in-control ARL was set to 200 for all charts; the out-of-control ARL's and corresponding standard deviations of run lengths (SDRLs) were as shown in Table 3 and Table 4.

As Hotelling's T^2 chart is popular for general-purpose multivariate process monitoring and the MEWMA chart is known for its sensitivity in detecting small-shift, we also studied the ARL performance of Hotelling's T^2 and MEWMA charts and compared them with the VS-MEWMA chart (Hotelling's T^2 chart is shown as a MEWMA chart with $\lambda = 1.0$ in the tables). The VS-MSPC chart proposed by Wang and Jiang (2009) can be seen as a special case of the VS-MEWMA chart with $\lambda = 1.0$; its performance is also shown in the tables.

As we are studying a process with all variables having unit variance and independence with each other, it is natural to monitor these variables using multiple univariate EWMA charts (denoted by Multi-EWMA chart). Each variable is monitored by an individual EWMA chart with all charts using the same values of the smoothing parameter λ and control limits. The Multi-EWMA chart signals if any individual chart goes out-of-control. The in-control ARLs of all the individual charts are the same, with the overall in-control ARL of the Multi-EWMA chart adjusted to be 200. The idea of using multiple univariate charts has been explored by some researchers. Pignatiello and Runger (1990) compared MCUSUM with multiple univariate CUSUM charts and pointed out that performance of the multiple univariate scheme depends on the manner in which the process mean shifts. This happens since each CUSUM chart is optimally designed for a particular shift only. Han et al. (2007) also proposed a different multichart scheme with multiple CUSUM or EWMA charts; each chart is designed for a specific shift size; the whole multichart scheme is designed for detecting a range of shifts. In the following study, we also report the performance of Multi-EWMA charts when applied to the simulated processes.

In Table 3 and Table 4, the smallest ARL value of each row is shown in bold, showing that the λ value that achieves the lowest ARL value increases with δ. This is a typical feature of EWMA-type control charts, that a small value for λ is helpful in accumulating recent information and therefore increases the sensitivity to small shifts. In general, for a specified shift size and in-control ARL, there is an optimal choice of λ that minimizes the out-of-control ARL (Lucas and Saccucci (1990)).

We now compare the VS-MEWMA chart with the conventional MEWMA chart. When $p = 10$ and $s = p_0 = 2$, Table 3 shows that, for a given δ value that is not too small, the VS-MEWMA chart *with its optimal value for λ* typically achieves a smaller out-of-control ARL than the MEWMA chart *with its (different) optimal value for λ*, as would be expected. The apparent exceptions for small values of δ (0.2, 0.4, and 0.6) may be due to two reasons. First, the choice of values for λ to display in the table is limited. For example, with $\delta = 0.4$ and $\lambda = 0.05$, the ARL for both charts is 32.3. When λ is further decreased, the performance of both charts improves (results not shown here), and the VS-MEWMA chart may become slightly better than MEWMA. Second, when shift magnitude is too small, the probability of correct identification of shifted variables is low, which inevitably hurts the performance of the VS-MEWMA.

When $p = 50$, a similar pattern is exhibited except that the multiple univariate EWMA chart is slightly better than MEWMA for shifts with $\delta > 4.0$; but the VS-MEWMA chart is still superior to the multiple univariate EWMA chart in most cases.

If the VS-MEWMA chart uses $\lambda = 1.0$, which means only the most recent observation is used for VS and monitoring, the chart reduces to the VS-MSPC chart proposed by Wang and Jiang (2009). The MEWMA chart with $\lambda = 1.0$ reduces to Hotelling's T^2 chart, and the multiple univariate EWMA scheme reduces to a multiple Shewhart chart. Table 3 and Table 4 show that the VS-MEWMA chart with $\lambda = 1.0$ performs worse than the VS-MEWMA chart with $\lambda < 1.0$ for small shifts in most cases. This shows that the VS procedure using more than the most recent observation plays an important role in improving the performance in detecting small process changes. The tables also show that the VS-MEWMA chart with $\lambda = 1.0$ performs slightly worse than the MEWMA chart with $\lambda = 1.0$ for small shifts. This may be due to the lack of identification accuracy with small shifts and large λ for the VS-MEWMA chart. In summary, among the three charts, the VS-MEWMA chart is superior for detecting a wide range of moderate and large shifts and the MEWMA chart is the best for detecting small shifts.

TABLE 3. Steady-State ARL (SDRL) for $p = 10$ and $ARL_0 = 200$

δ	$\lambda = 0.05$	0.1	0.2	0.4	0.8	1.0
			VS-MEWMA ($s = p_0 = 2$)			
0.2	**86.2 (76.0)**	106 (101)	132 (130)	160 (159)	184 (183)	188 (191)
0.4	**32.3 (20.7)**	37.8 (29.6)	54.9 (49.5)	89.7 (87.4)	142 (143)	158 (160)
0.6	**18.5 (9.47)**	18.7 (11.1)	23.7 (18.4)	42.8 (40.3)	94.0 (94.2)	118 (118)
0.8	12.8 (5.75)	**11.8 (5.87)**	12.9 (8.35)	20.6 (17.6)	56.7 (55.2)	80.6 (80.6)
1.0	9.85 (4.07)	8.64 (3.74)	**8.46 (4.57)**	11.8 (9.18)	33.1 (31.9)	52.0 (51.9)
1.5	6.30 (2.29)	5.20 (1.86)	**4.54 (1.82)**	4.63 (2.50)	9.44 (8.43)	16.6 (16.1)
2.0	4.67 (1.57)	3.82 (1.23)	3.17 (1.07)	**2.85 (1.17)**	3.72 (2.72)	5.85 (5.33)
2.5	3.79 (1.21)	3.04 (0.93)	2.49 (0.77)	**2.10 (0.74)**	2.11 (1.20)	2.61 (2.08)
3.0	3.19 (0.99)	2.57 (0.76)	2.09 (0.60)	1.72 (0.57)	**1.48 (0.67)**	1.57 (0.94)
3.5	2.76 (0.84)	2.23 (0.64)	1.83 (0.53)	1.46 (0.51)	1.18 (0.41)	**1.17 (0.45)**
4.0	2.48 (0.75)	2.01 (0.56)	1.63 (0.50)	1.25 (0.44)	1.05 (0.23)	**1.05 (0.22)**
4.5	2.25 (0.67)	1.84 (0.51)	1.47 (0.50)	1.11 (0.32)	1.01 (0.11)	**1.01 (0.09)**
5.0	2.07 (0.60)	1.71 (0.49)	1.32 (0.47)	1.04 (0.20)	1.00 (0.04)	**1.00 (0.04)**
			MEWMA			
0.2	**83.0 (73.5)**	102 (97.1)	127 (126)	156 (154)	178 (177)	188 (184)
0.4	**32.3 (21.4)**	37.4 (29.6)	52.9 (48.2)	83.5 (79.9)	135 (134)	155 (152)
0.6	**18.4 (9.78)**	18.8 (11.8)	23.6 (18.7)	41.3 (38.5)	90.1 (89.8)	115 (114)
0.8	12.9 (6.13)	**12.1 (6.29)**	13.2 (8.66)	21.2 (18.4)	55.2 (53.6)	79.3 (78.6)
1.0	9.98 (4.29)	8.85 (4.01)	**8.84 (4.85)**	12.1 (9.45)	33.1 (32.2)	52.0 (51.2)
1.5	6.36 (2.37)	5.37 (2.00)	**4.69 (1.93)**	4.87 (2.70)	9.92 (8.90)	17.4 (17.0)
2.0	4.77 (1.68)	3.93 (1.34)	3.29 (1.18)	**2.95 (1.26)**	4.07 (3.07)	6.49 (5.94)
2.5	3.80 (1.30)	3.15 (0.99)	2.58 (0.82)	**2.20 (0.79)**	2.29 (1.36)	2.93 (2.39)
3.0	3.2 (1.06)	2.63 (0.80)	2.15 (0.65)	1.80 (0.60)	**1.58 (0.76)**	1.73 (1.14)
3.5	2.81 (0.89)	2.30 (0.69)	1.87 (0.55)	1.52 (0.53)	**1.24 (0.48)**	1.25 (0.56)
4.0	2.51 (0.79)	2.05 (0.59)	1.68 (0.51)	1.33 (0.47)	1.08 (0.28)	**1.07 (0.28)**
4.5	2.28 (0.70)	1.87 (0.53)	1.53 (0.50)	1.16 (0.36)	1.02 (0.14)	**1.02 (0.13)**
5.0	2.09 (0.63)	1.73 (0.50)	1.38 (0.49)	1.07 (0.25)	1.00 (0.05)	**1.00 (0.05)**
			Multiple Univariate EWMA (Multi-EWMA)			
0.2	**89.8 (79.1)**	112 (107)	138 (138)	165 (165)	184 (184)	191 (190)
0.4	**34.4 (22.2)**	40.3 (31.7)	59.4 (54.2)	95.9 (95.1)	148 (147)	164 (162)
0.6	**19.6 (10.0)**	20.0 (12.5)	26.1 (20.9)	47.5 (44.8)	102 (101)	126 (125)
0.8	13.5 (6.08)	**12.8 (6.50)**	14.2 (9.53)	23.7 (21.0)	63.8 (62.0)	90.5 (89.6)
1.0	10.5 (4.32)	9.31 (4.16)	**9.26 (5.07)**	13.5 (10.7)	37.7 (36.9)	60.3 (59.5)
1.5	6.68 (2.47)	5.60 (2.06)	**4.89 (1.99)**	5.25 (2.97)	11.2 (10.1)	20.0 (19.5)
2.0	4.95 (1.70)	4.06 (1.36)	3.44 (1.21)	**3.12 (1.35)**	4.52 (3.43)	7.47 (6.95)
2.5	3.99 (1.29)	3.25 (1.02)	2.67 (0.84)	**2.29 (0.84)**	2.47 (1.51)	3.32 (2.74)
3.0	3.36 (1.06)	2.72 (0.82)	2.22 (0.67)	1.86 (0.63)	**1.68 (0.85)**	1.91 (1.34)
3.5	2.92 (0.90)	2.37 (0.70)	1.94 (0.56)	1.57 (0.54)	**1.30 (0.53)**	1.32 (0.65)
4.0	2.61 (0.78)	2.12 (0.61)	1.73 (0.50)	1.37 (0.49)	1.11 (0.33)	**1.10 (0.33)**
4.5	2.36 (0.70)	1.93 (0.54)	1.57 (0.51)	1.20 (0.40)	1.03 (0.17)	**1.02 (0.16)**
5.0	2.16 (0.64)	1.79 (0.49)	1.42 (0.49)	1.09 (0.28)	1.01 (0.08)	**1.01 (0.07)**

TABLE 4. Steady-State ARL (SDRL) for $p = 50$ and $ARL_0 = 200$

δ	$\lambda = 0.05$	0.1	0.2	0.4	0.8	1.0
			VS-MEWMA $(s = p_0 = 2)$			
0.2	**125 (119)**	151 (147)	174 (173)	187 (188)	195 (196)	197 (195)
0.4	**46.8 (32.2)**	62.9 (53.8)	99.9 (95.1)	145 (144)	178 (179)	185 (183)
0.6	**24.7 (12.7)**	27.3 (17.8)	43.2 (37.1)	87.1 (85.0)	149 (152)	166 (166)
0.8	16.6 (7.17)	**16.1 (8.22)**	20.6 (14.7)	44.0 (41.0)	113 (115)	140 (140)
1.0	12.4 (4.76)	**11.2 (4.79)**	12.2 (7.03)	22.2 (19.1)	77.3 (76.5)	110 (108)
1.5	7.76 (2.54)	6.45 (2.17)	**5.80 (2.30)**	6.84 (4.18)	22.0 (21.1)	43.5 (43.3)
2.0	5.69 (1.73)	4.60 (1.38)	3.85 (1.24)	**3.68 (1.59)**	7.00 (5.77)	14.2 (13.9)
2.5	4.56 (1.31)	3.61 (1.01)	2.98 (0.87)	**2.57 (0.89)**	3.24 (2.14)	5.32 (4.81)
3.0	3.81 (1.05)	3.04 (0.81)	2.44 (0.67)	2.02 (0.62)	**1.96 (1.02)**	2.50 (1.91)
3.5	3.31 (0.90)	2.63 (0.68)	2.10 (0.54)	1.73 (0.53)	**1.44 (0.62)**	1.52 (0.89)
4.0	2.93 (0.77)	2.33 (0.60)	1.87 (0.47)	1.49 (0.51)	1.17 (0.40)	**1.16 (0.44)**
4.5	2.66 (0.69)	2.11 (0.52)	1.72 (0.47)	1.29 (0.45)	1.05 (0.22)	**1.04 (0.21)**
5.0	2.42 (0.64)	1.94 (0.46)	1.58 (0.50)	1.14 (0.34)	1.01 (0.11)	**1.01 (0.08)**
			MEWMA			
0.2	**120 (115)**	145 (142)	164 (163)	181 (181)	192 (191)	196 (195)
0.4	**53.3 (41.1)**	69.4 (61.8)	99.5 (97.4)	136 (138)	171 (169)	181 (181)
0.6	**29.2 (17.4)**	34.3 (25.6)	51.7 (47.5)	89.3 (87.9)	140 (138)	159 (160)
0.8	**19.7 (9.76)**	20.4 (12.2)	28.0 (22.7)	53.2 (50.3)	110 (108)	132 (134)
1.0	14.7 (6.56)	**14.2 (7.08)**	16.9 (11.6)	32.0 (29.2)	81.2 (81.1)	106 (108)
1.5	9.17 (3.47)	7.93 (3.10)	**7.63 (3.58)**	10.5 (7.64)	33.3 (32.5)	53.4 (52.8)
2.0	6.65 (2.30)	5.59 (1.90)	**4.91 (1.82)**	5.24 (2.81)	13.1 (12.1)	24.6 (24.4)
2.5	5.27 (1.74)	4.35 (1.34)	3.65 (1.20)	**3.42 (1.41)**	6.00 (4.88)	11.2 (10.6)
3.0	4.39 (1.39)	3.58 (1.05)	2.97 (0.90)	**2.58 (0.92)**	3.30 (2.21)	5.38 (4.87)
3.5	3.80 (1.16)	3.10 (0.89)	2.52 (0.73)	**2.12 (0.67)**	2.16 (1.15)	2.92 (2.34)
4.0	3.35 (1.01)	2.70 (0.75)	2.20 (0.61)	1.83 (0.56)	**1.62 (0.74)**	1.85 (1.28)
4.5	3.02 (0.88)	2.44 (0.67)	1.97 (0.50)	1.63 (0.52)	**1.31 (0.52)**	1.35 (0.69)
5.0	2.74 (0.79)	2.23 (0.60)	1.83 (0.47)	1.44 (0.50)	1.14 (0.36)	**1.13 (0.38)**
			Multiple Univariate EWMA (Multi-EWMA)			
0.2	**132 (123)**	158 (156)	178 (174)	190 (191)	195 (195)	198 (197)
0.4	**49.1 (34.5)**	67.4 (57.1)	106 (101)	151 (149)	182 (183)	189 (190)
0.6	**25.7 (13.6)**	29.1 (19.8)	46.9 (40.9)	94.7 (92.1)	155 (156)	172 (172)
0.8	17.3 (7.48)	**16.8 (8.84)**	22.1 (16.4)	47.3 (44.2)	120 (119)	146 (144)
1.0	12.9 (5.05)	**11.7 (5.10)**	13.1 (7.95)	24.4 (21.4)	82.0 (81.4)	115 (113)
1.5	8.07 (2.73)	6.76 (2.33)	**6.16 (2.52)**	7.41 (4.62)	24.6 (22.9)	47.0 (46.7)
2.0	5.93 (1.86)	4.80 (1.48)	4.07 (1.36)	**3.94 (1.77)**	8.09 (7.01)	16.3 (15.7)
2.5	4.70 (1.36)	3.79 (1.09)	3.10 (0.92)	**2.73 (1.00)**	3.68 (2.54)	6.13 (5.68)
3.0	3.94 (1.13)	3.16 (0.87)	2.56 (0.72)	**2.15 (0.69)**	2.19 (1.23)	2.98 (2.46)
3.5	3.41 (0.93)	2.72 (0.73)	2.19 (0.59)	1.81 (0.55)	**1.58 (0.74)**	1.74 (1.11)
4.0	3.03 (0.81)	2.42 (0.63)	1.94 (0.49)	1.57 (0.52)	**1.26 (0.48)**	1.26 (0.57)
4.5	2.72 (0.73)	2.18 (0.56)	1.77 (0.47)	1.37 (0.49)	1.09 (0.30)	**1.08 (0.29)**
5.0	2.50 (0.66)	2.00 (0.50)	1.64 (0.49)	1.21 (0.40)	1.02 (0.16)	**1.02 (0.14)**

TABLE 5. Zero-State ARL Comparison

δ	MEWMA			VS-MEWMA		
	$\lambda = 0.1$	0.2	0.4	0.1	0.2	0.4
			$p = 10$			
0.2	**105 (94.2)**	131 (124)	157 (153)	107 (95.6)	135 (129)	160 (157)
0.4	39.5 (28.1)	54.8 (48.7)	86.6 (82.3)	**39.5 (27.7)**	56.8 (49.8)	90.7 (87.9)
0.6	20.5 (11.0)	24.9 (18.6)	42.3 (39.4)	**19.7 (10.3)**	24.4 (18.4)	43.1 (40.4)
0.8	13.4 (5.49)	14.0 (8.34)	21.6 (18.3)	**12.6 (5.24)**	13.5 (8.04)	21.2 (18.1)
1.0	9.93 (3.36)	9.38 (4.66)	12.5 (9.40)	9.25 (3.23)	**8.92 (4.35)**	11.9 (8.83)
1.5	6.14 (1.49)	5.17 (1.70)	5.11 (2.58)	5.60 (1.43)	4.82 (1.61)	**4.81 (2.47)**
2.0	4.53 (0.90)	3.64 (0.94)	3.18 (1.18)	4.09 (0.87)	3.35 (0.89)	**2.95 (1.10)**
2.5	3.64 (0.65)	2.87 (0.65)	2.38 (0.69)	3.25 (0.60)	2.61 (0.62)	**2.21 (0.65)**
3.0	3.09 (0.47)	2.40 (0.52)	1.96 (0.49)	2.74 (0.52)	2.19 (0.43)	**1.81 (0.52)**
3.5	2.71 (0.48)	2.09 (0.32)	1.69 (0.49)	2.32 (0.47)	1.98 (0.29)	**1.50 (0.51)**
4.0	2.34 (0.47)	1.98 (0.22)	1.43 (0.50)	2.07 (0.26)	1.82 (0.39)	**1.24 (0.43)**
4.5	2.08 (0.26)	1.88 (0.33)	1.20 (0.40)	2.00 (0.12)	1.58 (0.49)	**1.08 (0.27)**
5.0	2.01 (0.09)	1.69 (0.46)	1.06 (0.24)	1.96 (0.20)	1.31 (0.46)	**1.02 (0.14)**
			$p = 50$			
0.2	**147 (133)**	164 (158)	180 (179)	152 (137)	173 (167)	186 (183)
0.4	73.4 (58.9)	102 (95.2)	136 (136)	**65.7 (52.0)**	101 (92.7)	146 (141)
0.6	37.8 (24.0)	54.0 (47.2)	88.9 (86.5)	**29.3 (17.3)**	44.0 (36.5)	86.8 (83.4)
0.8	23.3 (11.1)	29.7 (22.4)	54.8 (51.7)	**17.3 (7.51)**	21.4 (14.5)	43.8 (40.5)
1.0	16.6 (6.06)	18.2 (11.2)	32.6 (28.9)	**12.2 (4.27)**	12.9 (6.98)	22.7 (19.1)
1.5	9.90 (2.30)	8.65 (3.22)	11.0 (7.36)	6.96 (1.70)	**6.16 (2.10)**	7.04 (4.02)
2.0	7.27 (1.30)	5.79 (1.51)	5.65 (2.62)	4.94 (0.99)	4.09 (1.06)	**3.79 (1.51)**
2.5	5.81 (0.86)	4.47 (0.92)	3.82 (1.31)	3.86 (0.69)	3.13 (0.70)	**2.67 (0.82)**
3.0	4.89 (0.66)	3.70 (0.67)	2.94 (0.81)	3.22 (0.51)	2.56 (0.56)	**2.14 (0.52)**
3.5	4.24 (0.51)	3.20 (0.48)	2.43 (0.57)	2.82 (0.46)	2.17 (0.39)	**1.84 (0.45)**
4.0	3.81 (0.47)	2.87 (0.42)	2.14 (0.38)	2.42 (0.50)	2.01 (0.22)	**1.56 (0.50)**
4.5	3.36 (0.48)	2.56 (0.50)	2.00 (0.25)	2.11 (0.31)	1.92 (0.27)	**1.30 (0.46)**
5.0	3.07 (0.26)	2.23 (0.42)	1.91 (0.30)	2.01 (0.11)	1.75 (0.43)	**1.11 (0.31)**

Zero-State ARL Performance Study

For completeness, the zero-state ARL performance of the VS-MEWMA chart with certain parameter combinations is also shown in Table 5 for comparison with the MEWMA chart. The smallest value within each row is highlighted in bold. The VS-MEWMA chart usually outperforms the MEWMA chart when $\delta > 0.6$ or 0.8 if $p = 10$. The margin becomes larger when p increases to 50. This shows that the advantage of the VS-MEWMA chart is persistent for both steady-state and zero-state ARL.

The difference between the steady-state and zero-state ARLs is mainly caused by the initial value used in Equation (6). $\mathbf{w}_0$ usually takes the in-control mean value of the process. If the process is out-of-control (e.g., due to initial setup bias) at the first step, the performance of both VS-MEWMA and MEWMA charts will be affected. For convenience, we here used a constant control limit rather than time-varying limits in the above simulations; this could be another reason that makes the chart slow in detecting process shifts present in the early observations. In practice, time-varying control limits could be used, and the design of univariate EWMA charts for fast initial response can be adapted to the VS-MEWMA chart as well (Montgomery (2005)). It should be noted that, as reported in Stoumbos and Sullivan (2002), using

time-varying control limits based on the exact covariance matrix of $\mathbf{w}_0$ makes the chart less robust to distributions having high kurtosis.

Discussions

Dependent Variables and Comparison with LEWMA

The preceding studies assume that process variables are independent. However, in many industrial processes, the variables are correlated, and many variables may shift as a result of a common cause. For example, Apley and Lee (2010) proposed a model to characterize variable patterns in autobody assembly, where observed variables are linked with sources of variations via a general matrix $\mathbf{C}$. Depending on the specific structure of $\mathbf{C}$, shifts of one variable source may lead to changes of multiple observed variables. Similarly, in blind source separation (see, e.g., Apley and Lee (2003), Shan and Apley (2008)), one source of variation (or signal) may affect multiple observable variables. Hawkins (1991) discussed a process with five correlated variables and different types of shift patterns. Zou and Qiu (2009) analyzed this process using their proposed LEWMA chart. In the following, we use this example to discuss the performance of the VS-MEWMA chart when monitoring dependent variables.

The in-control covariance matrix of five dependent variables, denoted by $\mathbf{\Sigma}_0$, is

$$\mathbf{\Sigma}_0 = \begin{pmatrix} 1.0000 & 0.1388 & 0.3496 & 0.0829 & 0.2652 \\ 0.1388 & 1.0000 & 0.7324 & 0.9130 & 0.6932 \\ 0.3496 & 0.7324 & 1.0000 & 0.6824 & 0.8214 \\ 0.0829 & 0.9130 & 0.6824 & 1.0000 & 0.7640 \\ 0.2652 & 0.6932 & 0.8214 & 0.7640 & 1.0000 \end{pmatrix}$$

Similar to Hawkins (1991) and Zou and Qiu (2009), twenty different types of hypothetical out-of-control mean shift patterns, with $\boldsymbol{\mu}_1 = \boldsymbol{\delta}$, are added to the process starting with step 25. The values of the elements of $\boldsymbol{\delta}$ are given in Table 6.

In the following, we apply the VS-MEWMA chart to the simulated process and compare its performance with the MEWMA chart and the LEWMA chart. The corresponding ARLs are shown in Table 6. For a fair comparison, we aligned our simulations settings with those in Zou and Qiu (2009) in this study. The in-control ARL is set to 500; each ARL is calculated from 20,000 replicates. The smallest ARL value of each row is again highlighted in bold. The last column shows the *actual* number of shifted variables, p_0, in each row. Since the magnitudes of some shifts in

patterns 16–20 are very small, we count the number of variables having shift magnitude larger than 0.01 and show them as the *effective* p_0 in parentheses.

Table 6 shows that for 18 of the 20 shift patterns the smallest ARL values are achieved by the VS-MEWMA, while the MEWMA chart is the best in the other two cases. Specifically, for the first five shift patterns (each having only one shifted variable), the VS-MEWMA chart with $s = 1$ always gives the best performance. For the 10 patterns having only two shifted variables (patterns 6–15), the VS-MEWMA chart with $s = 2$ gives the best performance in nearly half of the cases. For shift patterns 16–20, the best value for s is also close to the effective value for p_0. Note that the LEWMA chart searches among subsets including $k = 1, 2, \ldots, q$ variables, whereas the VS-MEWMA chart searches only subsets that contain exactly s variables.

Another important observation from Table 6 is that the VS-MEWMA chart is at least as good as the MEWMA chart (except for one case) if $s > p_0$, e.g., for the first five shift patterns when the VS-MEWMA chart has $s > 1$ and for the shift patterns 6–15 when the VS-MEWMA chart has $s > 2$.

Overall, the VS-MEWMA chart has effective performance in monitoring mean shifts of this correlated process. As would be expected, it usually has the smallest out-of-control ARL if the parameter s is close to p_0. If s is misspecified, its performance may be adversely affected. In the following, we further study the robustness of the chart against misspecification of s.

This example shows that the VS-MEWMA chart works well with processes with general covariance matrices. Equation (7), with the decomposition $\mathbf{\Sigma}^{-1} = \mathbf{R}^{\mathrm{T}}\mathbf{R}$, shows that the VS-MEWMA chart involves a variable transformation based on the Cholesky decomposition and uses the transformed variable in VS and monitoring. If that transformation does not identify the changed variable(s) correctly (which may be very likely, since identification of variation patterns is a more complicated task than a simple transformation), a single failure pattern may result in shifts in multiple components. In such cases, the performance of the VS-MEWMA chart could deteriorate. Nonetheless, if multiple sources of variations can be captured in a unified model such as linear regression to link the source of variations with the observed variables, we expect that VS methods may still be applicable to identify the true source of

WEI JIANG, KAIBO WANG, AND FUGEE TSUNG

TABLE 6. ARL Performance Study of a Process with a General Covariance Matrix (ARL$_0$ = 500, λ = 0.2)

Pattern Number	δ_1	δ_2	δ_3	δ_4	δ_5	MEWMA	LEWMA		VS-MEWMA				p_0
							$q = 3$	$q = 5$	$s = 1$	$s = 2$	$s = 3$	$s = 4$	
1	0.91	0	0	0	0	17.3	14.6	14.9	**14.4**	15.6	16.0	17.0	1
2	0	0.36	0	0	0	17.0	13.9	14.3	**13.2**	14.4	15.5	16.0	1
3	0	0	0.48	0	0	17.3	14.6	15.0	**13.8**	15.1	16.1	16.6	1
4	0	0	0	0.34	0	17.2	14.2	14.6	**13.4**	14.7	15.8	16.5	1
5	0	0	0	0	0.46	17.9	14.9	15.2	**13.9**	15.3	16.2	16.8	1
6	0.36	0.36	0	0	0	15.0	13.4	13.7	13.4	**13.2**	13.8	14.3	2
7	0.54	0	0.54	0	0	12.8	12.4	12.4	12.9	**12.2**	12.3	12.7	2
8	0.32	0	0	0.32	0	15.2	13.4	13.6	**13.1**	13.3	14.1	14.7	2
9	0.49	0	0	0	0.49	13.2	12.4	12.5	12.4	**12.0**	12.3	12.7	2
10	0	0.54	0.54	0	0	8.79	8.75	8.88	12.0	**8.21**	8.25	8.41	2
11	0	1.6	0	1.6	0	**3.48**	3.59	3.57	8.99	3.89	3.68	3.50	2
12	0	0.28	0	0	0.28	13.0	11.1	11.3	**10.3**	11.3	12.0	12.4	2
13	0	0	0.28	0.28	0	13.0	11.3	11.4	**10.6**	11.5	12.2	12.5	2
14	0	0	1.26	0	1.26	4.28	4.41	4.34	8.69	4.41	**4.20**	**4.20**	2
15	0	0	0	0.56	0.56	8.60	8.60	8.70	12.2	8.20	**8.19**	8.28	2
16	0.01	−0.15	0.07	0.17	−0.09	15.0	12.3	12.5	**11.5**	12.8	13.7	14.3	5 (2)*
17	0.07	−0.13	−0.4	0.19	0.35	10.1	10.0	10.1	12.3	**9.34**	9.43	9.66	5 (1)*
18	0.4	0.63	−0.57	0.47	−0.68	4.74	4.97	4.94	9.30	5.09	**4.65**	4.67	5 (5)*
19	−1.11	0.26	−0.17	0.34	−0.04	11.8	12.5	12.2	14.4	13.3	**11.3**	11.7	5 (4)*
20	2.51	7.11	7.05	7.11	7.08	**1.19**	2.88	1.30	7.44	3.02	2.19	1.47	5 (5)*

*Effective p_0 shown in parentheses.

variations (in contrast to the observed variables as discussed in this paper) so that we can simply monitor that particular source only. This will be further investigated in another paper.

Effect of Inaccurate Specification of the Number of Out-of-Control Variables

In the above simulations, we used a value of s that coincided with the value for p_0, which might not be possible in all practical applications. In some applications there may be several alternative out-of-control distributions having a variety of different values for p_0 so that it is impossible to achieve the ideal situation of $s = p_0$ for all special cases. Thus, it is useful to study the performance implications of the inferior matches where $s \neq p_0$.

With $p_0 = 2$ and $p = 10$ and 50, we used values of $s = 1$, 2, and 3 to simulate the cases where p_0 is slightly under- or overspecified. We found that the ARL was hardly affected by these choices for s. Plots of the ARL versus shift size (not shown) revealed that

the curves for the different values of s were visually indistinguishable.

Table 7 includes results for which s is far smaller than p_0, giving the results for the VS-MEWMA chart with $s = 2$, while p_0 is set to 2, $p/2$, and p, respectively. The MEWMA chart performance is also shown for comparison.

It is seen from Table 7 that the VS-MEWMA chart's performance is not much worse than that of the MEWMA chart, even when p_0 is greatly underestimated, especially with larger shift sizes. With smaller shifts, say $\delta < 0.8$, the penalty for underestimating p_0 may be confounded with a nonoptimal value of λ, since the same value of λ is used for both charts but the optimal value for the VS-MEWMA chart may be smaller.

This suggests that the VS-MEWMA chart is appealing for applications where most, but not all, of the out-of-control distributions have changes in only a few variables. It is more effective than the

TABLE 7. ARL Comparison Under Parameter Misspecification ($p = 10$, $s = 2$, $\lambda = 0.2$, $ARL_0 = 200$)

	Steady-State ARL						Zero-State ARL					
	MEWMA			VS- MEWMA			MEWMA			VS- MEWMA		
δ	$p_0 = 2$	5	10	2	5	10	2	5	10	2	5	10
0.2	127	77.0	43.2	132	86.9	52.9	131	79.6	44.7	135	88.4	54.9
0.4	52.9	21.2	10.7	54.9	25.6	14.3	54.8	22.3	11.5	56.8	26.7	15.3
0.6	23.6	9.76	5.53	23.7	11.3	7.32	24.9	10.3	6.02	24.4	12.1	7.99
0.8	13.16	6.05	3.75	12.9	7.03	4.92	14.0	6.54	4.16	13.5	7.53	5.45
1.0	8.84	4.41	2.90	8.46	5.10	3.71	9.38	4.83	3.24	8.92	5.51	4.24
1.5	4.69	2.73	1.94	4.54	3.10	2.40	5.17	3.04	2.16	4.82	3.41	2.83
2.0	3.29	2.05	1.54	3.17	2.29	1.85	3.64	2.27	1.89	3.35	2.54	2.18
2.5	2.58	1.70	1.22	2.49	1.87	1.54	2.87	1.99	1.38	2.61	2.09	1.97
3.0	2.15	1.45	1.05	2.09	1.61	1.30	2.40	1.80	1.03	2.19	1.94	1.81
3.5	1.87	1.23	1.01	1.83	1.39	1.11	2.09	1.41	1.00	1.98	1.73	1.44
4.0	1.68	1.09	1.00	1.63	1.21	1.03	1.98	1.09	1.00	1.82	1.40	1.11
4.5	1.53	1.02	1.00	1.47	1.09	1.00	1.88	1.01	1.00	1.58	1.12	1.01
5.0	1.38	1.00	1.00	1.32	1.03	1.00	1.69	1.00	1.00	1.31	1.02	1.00

MEWMA chart with small values of s and p_0, of course, but is still effective even with $s < p_0$.

Figure 4 shows the steady-state ARL contour plots for the VS-MEWMA chart with different values of p_0 and s. A process with $p = 10$ is simulated; a small shift of $\delta = 0.5$ and a moderate shift of $\delta = 1.5$ are studied. The vertical parallel lines on the left of the graphs show that when p_0 is small, the value of s does not affect ARL performance significantly. However, if p_0 is large (>5 or 6), the ARL performance deteriorates quickly if s is smaller than 5. This implies that if p_0 is small, the ARL performance is quite robust against misspecification of s; while, if $s < p_0$, the ARL performance may deteriorate.

Statistical Properties and Design of VS-MEWMA Charts

The VS-MEWMA chart is different from traditional charts, since the VS procedure is incorporated in process monitoring. Therefore, in the following, we further study its performance under different shift directions and discuss its design.

Directionally Variant Property of the VS-MEWMA Chart

Both Hotelling's T^2 and MEWMA charts enjoy the property of being directionally invariant. That is, their ARL performance is only influenced by
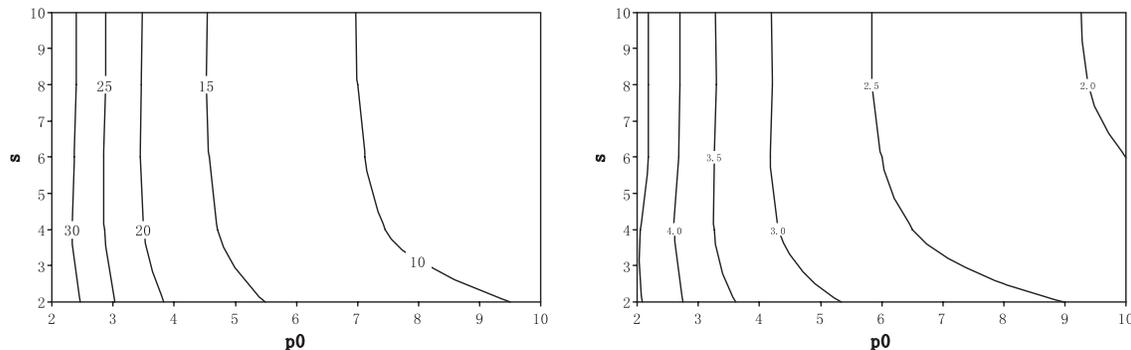

FIGURE 4. Steady-State ARL Contour Plot for the VS-MEWMA Chart under Parameter Misspecification. $p = 10$, $\lambda = 0.2$, $ARL_0 = 200$. (left) $\delta = 0.5$; (right) $\delta = 1.5$.
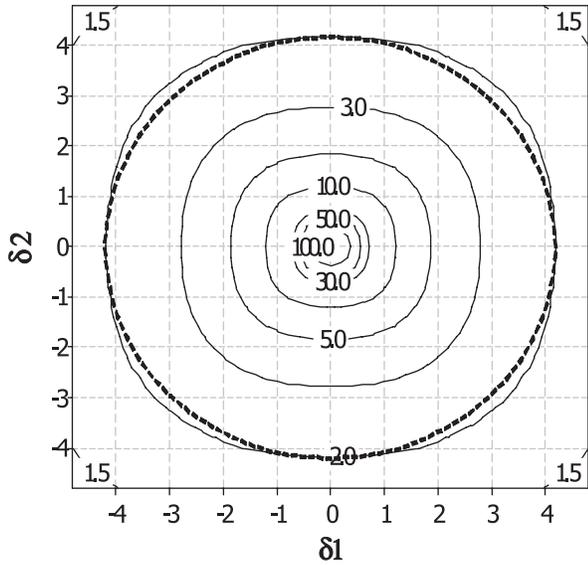
FIGURE 5. ARL Contour Plot of the Steady-State ARL of the VS-MEWMA Chart ($p = 10$, $s = 2$, $\lambda = 0.2$, in-control ARL $= 200$).

the shift magnitude via a noncentrality parameter (NCP), $(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta})^{1/2}$. The VS-MEWMA chart with $s = p$, for which all variables may be assumed to be out-of-control, reduces to the MEWMA chart and is thus directionally invariant. With $s < p$, the directionally invariant property does not hold. As an illustration, we studied a VS-MEWMA chart with $p = 10$, $s = 2$, and $\lambda = 0.2$; the process variables are still assumed mutually independent, i.e., $\boldsymbol{\Sigma} = \mathbf{I}_p$. By varying the shift magnitude of the first two out-of-control variables, δ_1 and δ_2, we obtained the ARL values corresponding to different combinations of shifts. It was found that for all shifts having the same value of the NCP, the out-of-control ARL

values show slight variation with direction. For example, for all $\boldsymbol{\delta}$ such that NCP $= 1$, the out-of-control steady-state ARL varied between 14.3 and 16.3.

Figure 5 shows the contour plot under different combinations of (δ_1, δ_2). The contour plots almost form concentric circles, but with a slightly increased ARL when the shift sizes are nearly equal. A circle is shown (with a dashed line) as reference, which coincides with the 2.0 contour when only one direction is shifted. With nearly equal shifts, the circle is inside the ARL contour, indicating that a slightly larger NCP is required for the same out-of-control ARL when the shift is equally distributed. This seems reasonable, since one variable carrying the entire shift is more likely to be correctly identified by the VS step compared with several variables carrying a shift having the same NCP value. Furthermore, the benefit of the VS method lies in dimensionality reduction, and, as the number of shifted dimensions increases, that benefit is eroded.

Designing a VS-MEWMA Chart

Besides the number of suspicious variables, the most critical choice to be made in designing a VS-MEWMA chart is the value of the smoothing parameters, λ. As we noted above, this parameter affects the monitoring and diagnosis properties of the control chart directly. In the following, we study the steady-state ARL performance of the VS-MEWMA chart with different values of λ and δ, which is shown in Figure 6. For illustration, the process is still assumed to have $\boldsymbol{\Sigma} = \mathbf{I}_p$.

To quickly detect shifts as small as 0.2 or 0.4, we can see from Figure 6 that a small value of λ is beneficial. For a shift of magnitude 1.0, $\lambda \approx 0.13$ in Figure
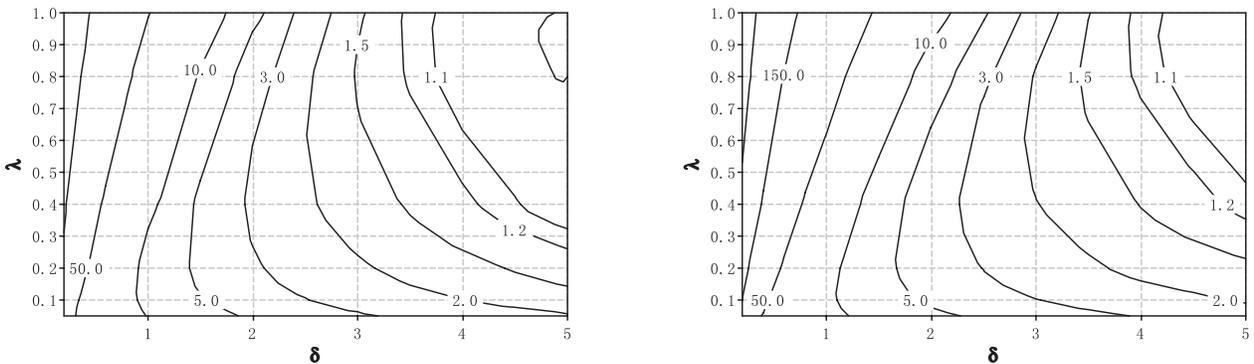

FIGURE 6. Steady-State ARL Contour Plots of VS-MEWMA ($s = 2$). (left) $p = 10$; (right) $p = 50$.

6(a) and $\lambda \approx 0.1$ in Figure 6(b) give the lowest out-of-control ARL. If the shift magnitude increases to 2.0, the optimal values of λ move to around 0.45 and 0.3, respectively. A further increase of shift size to 3.0 drives the optimal values of λ to 0.8 and 0.6. If the shift size is as large as 5.0, the optimal value of λ approaches 1.0, which means that the VS-MSPC chart using only the most recent observations is favored. The above phenomenon also resembles the traditional univariate EWMA chart in choosing λ, for which small values are often recommended for detecting small shifts. As the choice of λ largely depends on the values of p, s, and δ, a simple recommendation that satisfies all needs is hard to obtain. In general, a small value of λ is suitable for detecting small shifts and a large value is recommended for large shifts.

Once the smoothing parameter value is chosen, we must determine control limits for a specified in-control ARL. In this research, as a close-form equation is difficult to obtain, we designed a program and used Monte Carlo simulation to obtain control limits given all other design parameters. (The simulation program is available from the authors upon request.)

A Footwear Manufacturing Example

In this section, we use a real example from a footwear manufacturing process to illustrate the implementation of the VS-MEWMA chart for quality improvement. Shoe making essentially assembles different parts, including inner sole, middle sole, outer sole, vamp, heel, etc., in a production line. Normally, an inner sole and a vamp are first attached to a form with approximately the shape of a human foot, called a shoe *last*. A reference circle is then drawn on the bottom of the last using a pen, which will be used in later steps such as polishing and glue spreading. Specifically, a shoe-shaped wooden mask is first pressed against the last manually, and then a worker uses a pen to draw a circle on the last along the wooden mask edge. That circle is called a *reference circle* and is used in subsequent production procedures. Finally, the middle and outer soles are glued to the inner sole to form a shoe.

Among all these steps, the drawing step that plots the reference circle is critical to shoe quality. A displaced reference circle may cause misalignment among the sole, vamp, and last, which further leads to defects such as glue overflow, glue deficiency, and shape distortion in later production steps. Moreover, mask shifts, rotations, and pen tilts may make the reference circle shift out of control. Therefore, moni-
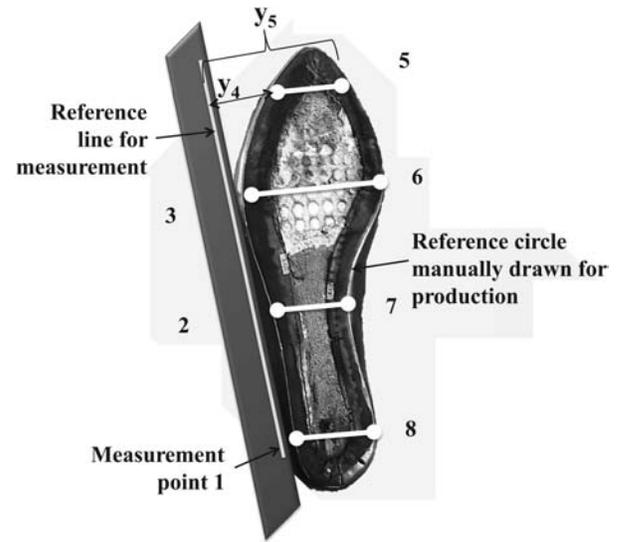

FIGURE 7. Measurement Points for Characterizing a Reference Circle.

toring the circle position is important to quality control in shoe production.

To characterize the reference circle, a measurement baseline (called a reference line) is defined, and shoes are put against the baseline. For each shoe sample, eight points are identified on the circle, and their distances to the reference line are measured, which forms an eight-dimensional observation vector for each shoe. Figure 7 illustrates the definition of the measurement point and variables on a shoe.

In order to monitor this process, 20 samples were collected from the production line (as shown in Table 8 and Figure 8). The sample variance-covariance matrix is shown in Table 9. Four types of multivariate control charts, T^2, VS-MSPC, MEWMA, and VS-MEWMA, were applied to monitor the process. Both MEWMA and VS-MEWMA charts share a smoothing parameter value of 0.2 for a fair comparison. Since the production is essentially labor intensive and the production pace is rather slow, the in-control ARL for all four charts is set to 25. Control limits were calculated via numerical simulations. Since we have no other historical observations available, the sample mean vector and variance-covariance matrix were treated as in-control values of the process. Once the control limits were obtained, the 20 samples were plotted on each chart and compared against the limits. The control charts are shown in Figure 9; detailed charting statistics and control limits are reported in Table 10.

TABLE 8. Measurements of Reference Circles in Footwear Manufacturing

Samples	Measurement Point (mm)							
	1	2	3	4	5	6	7	8
1	100.0	115.8	107.8	344.2	743.6	849.2	529.6	567.8
2	108.0	120.6	111.0	337.2	739.0	852.4	531.8	561.8
3	103.2	112.8	114.0	322.8	747.6	840.0	529.8	568.0
4	90.6	96.2	159.8	413.2	724.0	833.4	551.8	559.0
5	94.6	110.0	158.0	395.6	724.0	835.4	552.6	564.6
6	96.0	99.8	148.8	421.4	727.8	823.6	533.2	565.4
7	90.2	110.8	122.2	363.8	742.2	832.4	522.4	550.0
8	90.4	104.2	142.2	391.4	731.2	828.2	530.0	557.6
9	96.2	112.8	104.6	335.8	752.8	839.0	521.6	549.2
10	90.0	111.0	162.0	406.0	720.8	830.2	550.8	569.0
11	86.4	109.2	145.6	382.4	726.6	840.6	543.4	558.6
12	98.8	106.8	163.6	430.4	726.6	831.4	551.2	567.6
13	94.6	107.6	161.8	416.4	734.0	832.8	543.4	566.6
14	112.8	120.0	96.0	345.6	747.8	843.6	523.6	565.2
15	88.0	107.4	103.2	353.0	738.8	835.6	517.8	558.2
16	76.4	111.0	108.6	349.2	747.0	860.0	518.2	543.6
17	87.4	105.8	156.4	419.0	730.4	841.0	536.6	568.4
18	82.8	110.6	117.0	424.6	744.0	837.4	523.6	556.2
19	96.4	114.4	92.2	330.2	753.2	848.8	518.0	555.2
20	87.2	105.8	166.8	410.0	732.4	832.6	543.8	562.4

Figure 9 shows that, among the four charts, the VS-MEWMA chart is the only one that triggers an alarm, which happens at sample 20. Figure 9 also shows the suspicious variables suggested by the VS-MSPC and VS-MEWMA methods. It is interesting to note that variables y_5 and y_6 have been selected each time as suspicious. We also noticed that sample #11 and sample #20 have the lowest and largest charting statistics on the VS-MEWMA chart, respectively. A plot for these two samples is shown in Figure 10. Since the variables are highly correlated and the absolute values do not show meaningful information, their PCA-standardized distances are shown. It is seen that for these two samples y_5 and y_6 have almost opposite values. This indicates that one failure pattern, e.g., mask rotation, may have happened when the operator drew the circle.

Conclusions and Future Research

On the basis of the FVS algorithm, this paper extends the VS-MSPC chart by incorporating information from recent observations, in addition to the lat-

TABLE 9. Variance-Covariance Matrix of the Measurement Points

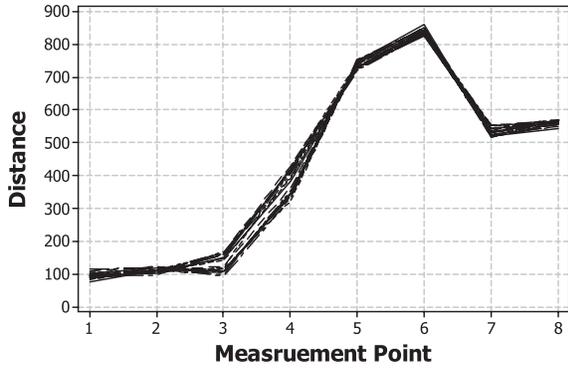|	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
y_1	72.5	25.6	−62.1	−122.3	18.9	4.9	0.2	29.7
y_2	25.6	35.2	−105.6	−153.6	37.4	34.6	−31.4	−1.7
y_3	−62.1	−105.6	699.6	833.7	−241.1	−150.8	287.9	87.0
y_4	−122.3	−153.6	833.7	1372.9	−289.1	−216.7	300.3	92.8
y_5	18.9	37.4	−241.1	−289.1	103.9	54.3	−107.9	−35.4
y_6	4.9	34.6	−150.8	−216.7	54.3	79.6	−50.6	−21.9
y_7	0.2	−31.4	287.9	300.3	−107.9	−50.6	150.0	51.7
y_8	29.7	−1.7	87.0	92.8	−35.4	−21.9	51.7	52.0

FIGURE 8. Plots of Sample Observations.

est one, for multivariate process monitoring and fault diagnosis, leading to a new VS-MEWMA chart. The new chart utilizes recent observations for both identifying suspicious variables and increasing the control chart sensitivity by focusing detection only on those "suspicious" variables. The rationale behind the VS-MEWMA chart is to first reduce dimensionality (and therefore improve detection sensitivity) and then monitor the reduced-dimension pro-

cess. The statistical efficiency and robustness of the VS-MEWMA chart are studied via simulation. The proposed chart has superior performance in detecting shifts in a small set of variables in a high-dimensional process; it is also robust to parameter misspecification if the number of shifted variables is small, and it still performs satisfactorily even if a shift affects all variables.

Moreover, note that VS-MEWMA charts proposed in this paper and in Zou and Qiu (2009) use a bound (s and q, respectively) to limit the number of shifted variables selected in each step. The parameter s is a hard bound, whereas q is an upper bound. In fact, instead of viewing the VS step as a penalized likelihood method applied to the EWMA statistic $\mathbf{w}_t$, one may consider a *prior* distribution for the probability of a variable to shift and update the *posterior* distribution after observing $\mathbf{y}_t$, similar to the Box-Meyer Bayesian approach for identifying active factors in the design of experiments (Box and Meyer (1993)). Such priors mimic the penalty imposed on the number of shifted variables but with soft-margins. It is also expected that such
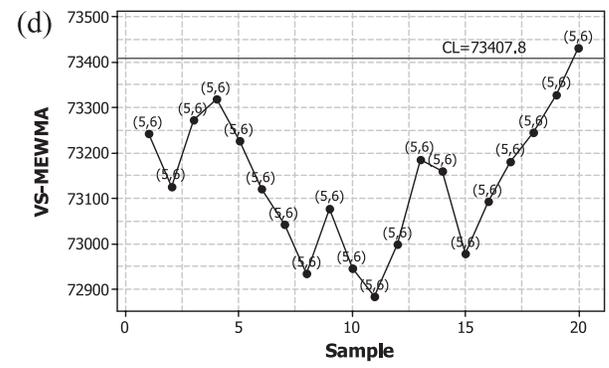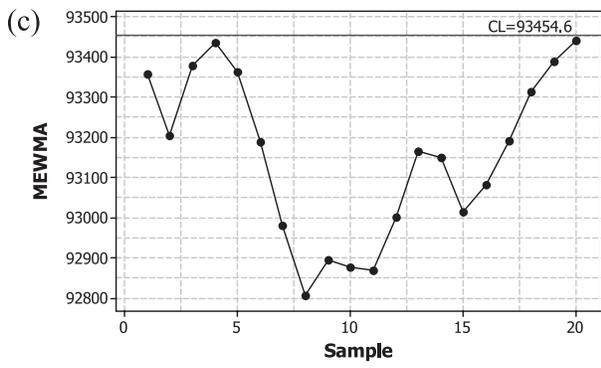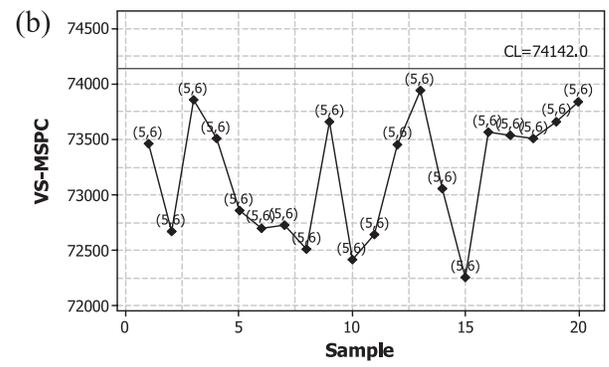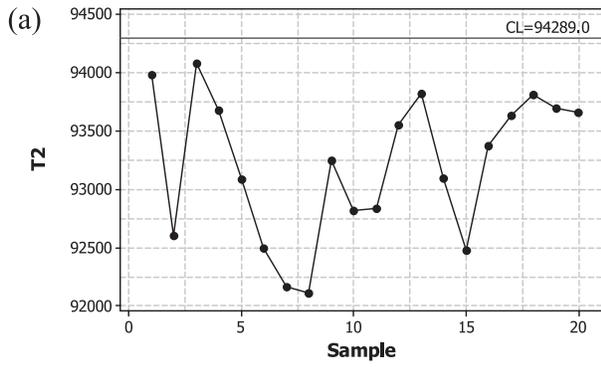

FIGURE 9. Control Chart Comparison for the Shoe Production Process. (a) T^2 chart; (b) VS-MSPC chart ($s = 2$, $\lambda = 0.2$); (c) MEWMA chart ($\lambda = 0.2$); (d) VS-MEWMA chart ($s = 2$, $\lambda = 0.2$).

TABLE 10. Charting Statistics
and Control Limit (CL) of the Example

	T^2	VS-MSPC	MEWMA	VS-MEWMA
			CL	
Sample	94,289.0	74,142.0	93,454.6	73,407.8
1	93,977.7	73,458.0	93,355.8	73,241.3
2	92,604.1	72,667.8	93,203.9	73,125.9
3	94,075.4	73,856.2	93,376.5	73,271.3
4	93,675.1	73,501.7	93,433.7	73,317.3
5	93,083.4	72,858.3	93,363.0	73,225.3
6	92,495.6	72,699.8	93,187.8	73,120.0
7	92,160.3	72,726.5	92,980.8	73,041.0
8	92,108.0	72,506.6	92,805.7	72,933.9
9	93,249.8	73,654.7	92,893.3	73,077.5
10	92,814.4	72,413.9	92,875.8	72,944.4
11	92,834.6	72,642.5	92,867.0	72,883.9
12	93,545.7	73,452.4	93,001.8	72,997.4
13	93,817.6	73,940.0	93,164.1	73,185.4
14	93,093.6	73,054.7	93,148.6	73,159.2
15	92,478.6	72,252.6	93,012.7	72,977.4
16	93,371.8	73,558.2	93,082.0	73,092.6
17	93,630.1	73,535.9	93,190.2	73,180.8
18	93,812.1	73,501.8	93,312.2	73,244.5
19	93,695.6	73,661.0	93,388.3	73,327.8
20	93,654.8	73,838.7	93,440.5	73,429.7

a Bayesian updating procedure would resemble an EWMA scheme that can capture recent information (Girshick and Rubin (1952)). In future work, we plan to explore Bayesian methods in adaptively determining the number of shifted variables for multivariate process monitoring and diagnosis.
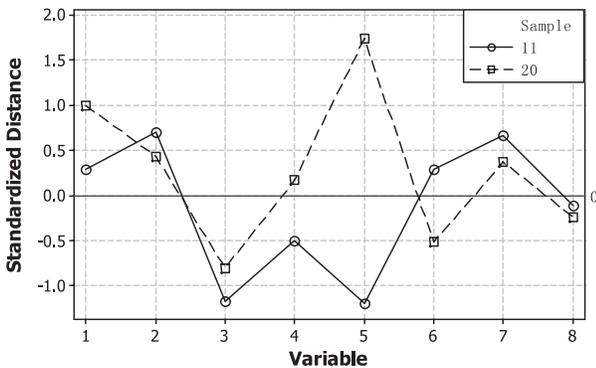

FIGURE 10. Plot of Samples #11 and #20 (standardized).

The VS-MEWMA chart proposed in this paper is for Phase II monitoring only. Developing a VS chart for Phase I analysis would be useful to pursue in future research. Moreover, the VS-MEWMA chart accumulates all recent observations to estimate the process mean. As suggested by one referee, it is expected to be more efficient if a change point could be first identified, and then only subsequently, the suspected out-of-control samples could be used to select variables and estimate process shifts. Zou et al. (2011) developed a VS method for identifying responsible variables following a signal. A similar approach could be applicable in the Phase I analysis as well.

Appendix A

Wang and Jiang (2009) proposed a VS-MSPC chart that monitors $\Lambda(\mathbf{y}_t) = 2\mathbf{y}_t^\mathrm{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_t^* - \boldsymbol{\mu}_t^{*\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_t^*$. Letting $\Lambda_1(\mathbf{y}_t) = \mathbf{y}_t^\mathrm{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_t^*$, we now prove that the control chart statistic $\Lambda(\mathbf{y}_t)$ is equivalent to $\Lambda_1(\mathbf{y}_t)$. Note that $\boldsymbol{\mu}^*$ is obtained from Equation (2); i.e., only s components are nonzero. Without loss of generality, assume that the first s components are nonzero and $\boldsymbol{\mu}^*$ can be represented as $(\boldsymbol{\delta}', \mathbf{0}')'$ where 0 denotes a zero vector of length $p - s$. Accordingly, we can partition $\mathbf{y}_t$ and $\boldsymbol{\Sigma}^{-1}$ as

$$\mathbf{y}_t = (\mathbf{u}, \mathbf{v})', \quad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

respectively. It follows that

$$\Lambda = \left(\mathbf{u} - \frac{\boldsymbol{\delta}}{2}\right)' \mathbf{A}_{11}\boldsymbol{\delta} + \mathbf{v}'\mathbf{A}_{12}\boldsymbol{\delta}$$
$$\Lambda_1 = \Lambda + \frac{1}{2}\boldsymbol{\delta}'\mathbf{A}_{11}\boldsymbol{\delta}.$$

Note that $\boldsymbol{\delta}$ is obtained from Equation (2), i.e.,

$$\boldsymbol{\delta} = \arg \min_{\boldsymbol{\delta}}[(\mathbf{u} - \boldsymbol{\delta})'\mathbf{A}_{11}(\mathbf{u} - \boldsymbol{\delta})$$
$$+ 2(\mathbf{u} - \boldsymbol{\delta})'\mathbf{A}_{12}\mathbf{v} + \mathbf{v}'\mathbf{A}_{22}\mathbf{v}].$$

Taking the first-order derivative and setting it to $\mathbf{0}$, the above equation gives $\boldsymbol{\delta} = \mathbf{u} + \mathbf{A}_{11}^{-1}\mathbf{v}\mathbf{A}_{12}$. Plugging this estimate into Λ and Λ_1, it follows that

$$\Lambda = \left(\mathbf{u} - \frac{\boldsymbol{\delta}}{2} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{v}\right)' \mathbf{A}_{11}\boldsymbol{\delta}$$
$$= \frac{1}{2}\boldsymbol{\delta}'\mathbf{A}_{11}\boldsymbol{\delta} = \frac{1}{2}\Lambda_1,$$

i.e., monitoring Λ is equivalent to monitoring Λ_1.

Appendix B

The following pseudo-code illustrates how the solution to Equation (7) is obtained.

1. Let $s_0 = 0$ // number of selected variables

2. Let $\mathbf{a}$ be a list of length s_0 // list of selected variables

3. for $i = 1$ to p

   (i) Add the ith variable to the list of selected variables if it is not in the list yet (move to the next variable if this one is already in the list)

   (ii) Fit model using all selected variables

   (iii) Calculate R^2 of the model

4. end of the "for" loop

5. Append the variable that has the largest R^2 value to list $\mathbf{a}$

6. $s_0 = s_0 + 1$

7. if $(s_0 < s)$, go to step 3 to select next variable; otherwise, a model with exactly s variables is selected, so the process terminates.

At step 3, with s_0 tentatively selected variables, we must obtain the solution $\boldsymbol{\mu}_t^*$, which has s_0 nonzero elements, to minimize the weighted likelihood function in Equation (5). With the consideration of $s_0 \leq s$, the objective is given by:

$$\begin{cases} \min_{\boldsymbol{\mu}_t} \left(\sum_{i=0}^{t} \omega_i (\mathbf{z}_i - \mathbf{R}\boldsymbol{\mu}_t)^{\mathrm{T}} (\mathbf{z}_i - \mathbf{R}\boldsymbol{\mu}_t) \right) \\ \text{s.t.} \quad \sum_j I(|\mu_{t(j)}| \neq 0) \leq s_0. \end{cases}$$

By rearranging variable sequence and moving all nonzero coefficients in $\boldsymbol{\mu}_t$ together, the design matrix and coefficients can be partitioned as follows:

$$\mathbf{R} = \left[\underbrace{\mathbf{R}_1}_{p \times s_0} \quad \underbrace{\mathbf{R}_2}_{p \times (p-s_0)} \right], \quad \boldsymbol{\mu}_t = \left[\begin{array}{c} \boldsymbol{\mu}_{t,1} \\ 0 \end{array} \right].$$

Now, $\mathbf{R}_2$ can be safely removed from the model to reduce the number of variables to s_0. With simple manipulation, we can get the estimate that minimizes the weighted likelihood function, as follows (the nonzero part)

$$\boldsymbol{\mu}_{t,1}^* = \frac{\left(\mathbf{R}_1^{\mathrm{T}}\mathbf{R}_1\right)^{-1} \sum_{i=1}^{t} \omega_i \mathbf{R}_1^{\mathrm{T}} \mathbf{z}_i}{\sum_{i=1}^{t} \omega_i}$$

$$= \left(\mathbf{R}_1^{\mathrm{T}}\mathbf{R}_1\right)^{-1} \mathbf{R}_1^{\mathrm{T}} \mathbf{R} \sum_{i=1}^{t} \left(\frac{\omega_i}{\sum_{i=1}^{t} \omega_i} \mathbf{y}_i \right).$$

Let $\mathbf{v}_t = \mathbf{R}\mathbf{w}_t$. Alternatively, if we choose to minimize Equation (7), which uses EWMA smoothed observations directly, the optimal solution $\boldsymbol{\mu}_t^*$ (to save notations, we use the same notation) should satisfy:

$$\begin{cases} \min_{\boldsymbol{\mu}_t} \left((\mathbf{v}_t - \mathbf{R}\boldsymbol{\mu}_t)^{\mathrm{T}} (\mathbf{v}_t - \mathbf{R}\boldsymbol{\mu}_t) \right) \\ \text{s.t.} \quad \sum_j I(|\mu_{t(j)}| \neq 0) \leq s. \end{cases}$$

Again, we rearrange variable sequence as above, and the nonzero part of the solution is, therefore, given by

$$\boldsymbol{\mu}_{t,1}^* = \left(\mathbf{R}_1^{\mathrm{T}}\mathbf{R}_1\right)^{-1} \mathbf{R}_1^{\mathrm{T}} \mathbf{R} \mathbf{w}_t$$

$$= \left(\mathbf{R}_1^{\mathrm{T}}\mathbf{R}_1\right)^{-1} \mathbf{R}_1^{\mathrm{T}} \mathbf{R} \sum_{i=1}^{t} \lambda(1-\lambda)^{t-i} y_i.$$

Therefore, setting

$$\omega_i = \lambda(1-\lambda)^{t-i} \Bigg/ \sum_{i=1}^{t} \lambda(1-\lambda)^{t-i}$$

makes the above two solutions equal, when $t \to \infty$.

This shows that by using the above weighting scheme, minimizing Equation (5) is equivalent to minimizing Equation (7). Therefore, in this work, we use the simple EWMA form in Equation (6) to accumulate recent observations and then use the above VS algorithm to obtain $\boldsymbol{\mu}_t^*$.

Acknowledgments

References

APLEY, D. W. and LEE, H. Y. (2003). "Identifying Spatial Variation Patterns in Multivariate Manufacturing Processes: A Blind Separation Approach". *Technometrics* 45(3), pp. 220–234.

APLEY, D. W. and LEE, H. Y. (2010). "Simultaneous Identification of Premodeled and Unmodeled Variation Patterns". *Journal of Quality Technology* 42(1), pp. 36–51.

BOX, G. E. P. and MEYER, R. D.(1993). "Finding the Active Factors in Fractionated Screening Experiments". *Journal of Quality Technology* 25(2), pp. 94–105.

CHENOURI, S.; STEINER, S. H.; and VARIYATH, A. M. (2009). "A Multivariate Robust Control Chart for Individual Observations". *Journal of Quality Technology* 41(3), pp. 259–271.

CHOI, S. W.; MARTIN, E. B.; MORRIS, A. J.; and LEE, I. B. (2006). "Adaptive Multivariate Statistical Process Control for Monitoring Time-Varying Processes". *Industrial & Engineering Chemistry Research* 45(9), pp. 3108–3118.

DING, Y.; ELSAYED, E. A.; KUMARA, S.; LU, J. C.; NIU, F.; and SHI, J. (2006). "Distributed Sensing for Quality and Productivity Improvements". *IEEE Transactions on Automation Science and Engineering* 3(4), pp. 344–359.

DUNIA, R.; QIN, S.; EDGAR, T.; and MCAVOY, T. (1996). "Identification of Faulty Sensors Using Principal Component Analysis". *AIChe Journal* 42(10), p. 2797.

GIRSHICK, M. A. and RUBIN, H. (1952). "A Bayes Approach to a Quality Control Model". *Annals of Mathematical Statistics* 23(1), pp. 114–125.

GONZALEZ, I. and SANCHEZ, I. (2008). "Principal Alarms in Multivariate Statistical Process Control". *Journal of Quality Technology* 40(1), pp. 19–30.

HAN, D.; TSUNG, F.; HU, X.; and WANG, K. (2007). "CUSUM and EWMA Multi-Charts for Detecting a Range of Mean Shifts". *Statistica Sinica* 17(3), pp. 1139–1164.

HAWKINS, D. M. (1991). "Multivariate Quality-Control Based on Regression-Adjusted Variables". *Technometrics* 33(1), pp. 61–75.

HAWKINS, D. M. (1993). "Regression Adjustment for Variables in Multivariate Quality Control". *Journal of Quality Technology* 25(3), pp. 170–182.

HAWKINS, D. M. and MABOUDOU-TCHAO, E. M. (2008). "Multivariate Exponentially Weighted Moving Covariance Matrix". *Technometrics* 50(2), pp. 155–166.

HUWANG, L. C.; YEH, A. B.; and WU, C. W. (2007). "Monitoring Multivariate Process Variability for Individual Observations". *Journal of Quality Technology* 39(3), pp. 258–278.

JIANG, W.; AU, T.; and TSUI, K. (2007). "A Statistical Process Control Approach to Business Activity Monitoring". *IIE Transactions* 39(3), pp. 235–249.

JIANG, W. and TSUI, K.-L. (2008). "A Theoretical Framework and Efficiency Study of Multivariate Control Charts". *IIE Transactions on Quality and Reliability Engineering* 40(7), pp. 650–663.

LUCAS, J. and SACCUCCI, M. (1990). "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements". *Technometrics* 32, pp. 1–12.

MASON, R. L.; TRACY, N. D.; and YOUNG, J. C. (1995). "Decomposition of T^2 for Multivariate Control Chart Interpretation". *Journal of Quality Technology* 27(2), pp. 99–108.

MASON, R. L. and YOUNG, J. C. (1999). "Improving the Sensitivity of the T^2 Statistic in Multivariate Process Control". *Journal of Quality Technology* 31(2), pp. 155–165.

MASON, R. L. and YOUNG, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia, PA: American Statistical Association and Society for Industrial Mathematics.

MASTRANGELO, C. M.; RUNGER, G. C.; and MONTGOMERY, D. C. (1996). "Statistical Process Monitoring with Principal Components". *Quality and Reliability Engineering International* 12, pp. 203–210.

MONTGOMERY, D. C. (2005). *Introduction to Statistical Quality Control*. Hoboken, NJ: John Wiley.

PIGNATIELLO, J. J. and RUNGER, G. C. (1990). "Comparisons of Multivariate CUSUM Charts". *Journal of Quality Technology* 22, pp. 173–186.

RUNGER, G. C. and ALT, F. B. (1996). "Choosing Principal Components for Multivariate Statistical Process Control". *Communications in Statistics—Theory and Methods* 25(5), pp. 909–922.

SHAN, X. M. and APLEY, D. W. (2008). "Blind Identification of Manufacturing Variation Patterns by Combining Source Separation Criteria". *Technometrics* 50(3), pp. 332–343.

SONESSON, C. and BOCK, D. (2003). "A Review and Discussion of Prospective Statistical Surveillance in Public Health". *Journal of the Royal Statistical Society, Series A: Statistics in Society* 166, pp. 5–21.

SPANOS, C. J. (1992). "Statistical Process Control in Semiconductor Manufacturing". *Proceedings of the IEEE* 80(6), pp. 819–830.

STOUMBOS, Z. G. and SULLIVAN, J. H. (2002). "Robustness to Non-Normality of the Multivariate EWMA Control Chart". *Journal of Quality Technology* 34(3), pp. 260–276.

WANG, K. and JIANG, W. (2009). "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection". *Journal of Quality Technology* 41(3), pp. 247–258.

WANG, K. and TSUNG, F. (2008). "An Adaptive T^2 Chart for Monitoring Dynamic Systems". *Journal of Quality Technology* 40, pp. 109–123.

ZHU, Y. Z. and JIANG, W. (2009). "An Adaptive T^2 Chart for Multivariate Process Monitoring and Diagnosis". *IIE Transactions* 41(11), pp. 1007–1018.

ZOU, C.; JIANG, W.; and TSUNG, F. (2011). "A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control". *Technometrics* 53(3), pp. 297–309.

ZOU, C. and TSUNG, F. (2008). "Directional MEWMA Schemes for Multistage Process Monitoring and Diagnosis". *Journal of Quality Technology* 40(4), pp. 407–427.

ZOU, C. L. and QIU, P. H. (2009). "Multivariate Statistical Process Control Using LASSO". *Journal of the American Statistical Association* 104(488), pp. 1586–1596.

∼