



Simultaneous monitoring of process mean vector and covariance matrix via penalized likelihood estimation

Kaibo Wang^a, Arthur B. Yeh^b, Bo Li^{c,*}

^a Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

^b Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 43403, United States

^c School of Economics and Management, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:

Received 9 August 2013

Received in revised form 27 April 2014

Accepted 27 April 2014

Available online 9 May 2014

Keywords:

Likelihood ratio test

L_1 penalty function

Penalized likelihood estimation

Phase II monitoring

ABSTRACT

In recent years, some authors have incorporated the penalized likelihood estimation into designing multivariate control charts under the premise that in practice typically only a small set of variables actually contributes to changes in the process. The advantage of the penalized likelihood estimation is that it produces sparse and more focused estimates of the unknown population parameters which, when used in a control chart, can improve the performance of the resulting control chart. Nevertheless, the existing works focus on monitoring changes occurring only in the mean vector or only in the covariance matrix. Stemming from the ideas of the generalized likelihood ratio test and the multivariate exponentially weighted moving covariance, new control charts are proposed for simultaneously monitoring the mean vector and the covariance matrix of a multivariate normal process. The performance of the proposed charts is assessed by both Monte-Carlo simulations and a real example.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The statistical process control (SPC) has been a major tool in manufacturing for assignable cause detection and variation reduction (Montgomery, 2005). In a multivariate process, unexpected changes in either the process mean vector or the covariance structure among the variables can lead to an increase in process variability. Therefore, the joint monitoring of both the mean vector and the covariance matrix of a multivariate process becomes very important in ensuring the overall process quality.

In multivariate process monitoring, one of the major challenges comes from the high dimensionality. For a p -dimensional process, there are p mean components and $p(p + 1)/2$ variance/covariance components that any number of these components may go wrong. Therefore, the number of possible combinations of out-of-control (OC) scenarios is usually high, which makes the conventional general-purpose multivariate control charts ineffective. Similar phenomenon has been observed in many contemporary high-dimensional statistical problems. Recent statistical literature has witnessed a blossom of research on proposing penalized methods to deal with high-dimensional data, under the premise that only a sparse set of variables is relevant. See Hastie et al. (2009) and Bühlmann and Van de Geer (2011) for extensive discussion.

As pointed out by Wang and Jiang (2009) and Li et al. (2013), when a change in a multivariate process occurs, it is typically the case in practice that only a small set of the mean or variance/covariance components has changed. That is, a certain

* Corresponding author. Tel.: +86 10 6279 5143.

E-mail address: libo@sem.tsinghua.edu.cn (B. Li).

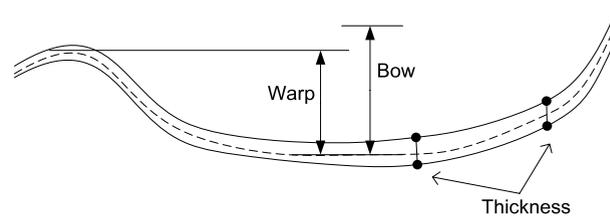


Fig. 1. Illustration of the shape and the thickness distribution of a wafer (sectional view).

sparsity exists in the shifted mean vector or covariance matrix. As a motivating example, we consider the monitoring of wafer quality in semiconductor manufacturing. Fig. 1 illustrates the shape and thickness distribution of a wafer. In industrial practice, the geometric quality of a wafer is characterized by indicators such as total thickness variation (TTV), total indicator reading (TIR), site TIR (STIR), Bow and Warp. More detailed definitions of these quality variables can be found in Li et al. (2013). Among these variables, TTV, TIR, and STIR are calculated from the thickness distribution, while Bow and Warp are calculated from the convex, concave or uneven shape. The thickness and shape of a wafer are affected by different engineering mechanisms. Therefore, the five quality variables could be classified into two groups. When the manufacturing process changes, it may lead to shifts in one of the groups, thus leading to sparse mean or correlation shifts.

In the setting of Multivariate SPC, when the sparsity assumption about the process shift patterns is reasonable and incorporated into the chart design, one can potentially improve the chart performance by adopting a more focused control charting mechanism. For example, Wang and Jiang (2009) designed a Shewhart-type chart, the variable-selection-based multivariate statistical process control (VS-MSPC) chart, for monitoring the mean vector. The VS-MSPC chart first employs the forward-variable-selection method to select a small set of potentially shifted variables, then calculates a T^2 -based charting statistic to detect mean changes in the small set of variables. Jiang et al. (2012) incorporated the variable-selection procedure with a MEWMA update equation and proposed the VS-MEWMA chart to further improve the chart performance in detecting small mean shifts. Zou and Qiu (2009) proposed to use the Lasso algorithm, instead of the forward-variable-selection procedure for variable selection.

As for monitoring the covariance matrix, Li et al. (2013) recently took advantage of the sparsity of the usual covariance matrix, and proposed a penalized likelihood ratio (PLR) chart. The PLR chart calculates the penalized likelihood ratio of a group of samples and signals an alarm when the likelihood ratio shifts to an abnormal value. The use of the penalized likelihood ratio has the effect of shrinking some of the components in the covariance matrix to zero, thus reducing the effective dimension of the parameters needed to be monitored. In two independent works, Yeh et al. (2012) and Maboudou-Tchao and Diawara (2013) applied the above idea for the covariance matrix monitoring to the case when only individual observations are available. Yeh et al. (2012) modified the penalty function used by Li et al. (2013) and shrank the sample precision matrix toward the in-control (IC) one, rather than to 0 as more commonly seen in the existing literature. This modification is meaningful in the SPC context. Maboudou-Tchao and Diawara (2013) proposed an effective accumulative method by penalizing the precision matrix per se in a slightly different EWMA propagation. Maboudou-Tchao and Agboti (2013) also studied the monitoring of the covariance matrix when the number of observations is fewer than the number of variables. The authors proposed to use the graphical Lasso algorithm to obtain a sparse estimate of the precision matrix, then used the sparse estimate for shift detection.

Although proven efficient, the variable-selection based or penalized likelihood estimation based charts are designed for detecting either just the mean shift or just changes in the covariance matrix. In this work, we are motivated to develop new penalized likelihood estimation based control charts for simultaneously monitoring the mean vector and the covariance matrix of a multivariate process using individual observations.

Some methods have been developed in the literature for the joint monitoring of the process mean and the variability. Hawkins and Zamba (2005) derived the likelihood ratio test for a change in mean and/or variance for normally distributed data, which formed the basis for a single chart thus developed. The advantages of this method include: (a) the chart does not rely on parameter estimates derived from the Phase I observations since the errors in Phase I estimation may lead to uncertain run length distribution of the chart; (b) the chart is in simple form since it uses only one instead of two charts to monitor both the mean and the variance; and (c) the chart can start quickly given a very small number (typically three) of Phase I observations. Therefore, the chart can be easily used to monitor short-run processes with no or limited historical data. Chen et al. (2001) proposed to monitor the mean and the variance of a univariate process using a single EWMA chart. Two EWMA statistics are first designed for standardized mean and transformed variance terms. Since the two EWMA statistics are independent and follow the same standardized normal distribution, the author suggested to monitor the maximum of these two statistics and the resulting control chart triggers an alarm when the maximum is larger than the control limit. This chart is named the MaxEWMA chart. Li et al. (2010) proposed a self-starting chart for the simultaneous monitoring of the process mean and the variance based on the likelihood-ratio statistic and the EWMA procedure. Nevertheless, all of the afore-mentioned charts are designed for univariate processes.

Zhou et al. (2010) proposed to use the generalized likelihood ratio test (GLRT) to monitor patterned mean and variance (non-constant time-varying) changes. In this chart, a likelihood ratio test statistic was derived based on the process mean,

which is used for monitoring process mean shifts with unknown patterns. Another likelihood ratio test statistic was also derived for detecting variance changes. Finally, these two likelihood ratio test statistics are put together to form a new vector, which is then monitored to detect changes in both the mean and the variance. However, in this method, only changes in the diagonal variance components can be detected. In addition, in the GLRT test, the shift pattern information in the alternative hypothesis must be completely known.

In this work, we are motivated to develop a novel charting statistic for the joint monitoring of the process mean vector and the covariance matrix of a multivariate process. It is assumed that only individual observations are available at each sampling period. That is, the subgroup size is equal to one. The charting statistic first tries to estimate the process mean vector and the covariance matrix using a penalized likelihood estimation method. The charting statistic is then derived based on a likelihood ratio test. The sparse estimates obtained in this procedure are also helpful to process diagnosis. It is worthwhile to note that in this paper, we modified the penalty for the covariance matrix to be the Frobenius norm of the difference between the estimated covariance matrix and the in-control covariance matrix. While most earlier works concentrated on the precision matrix partly for the sake of the ease of computation, we think the covariance matrix per se is more relevant for industrial practice and gives more informative clues for diagnostics. On the computational side, we modified the algorithm recently proposed by [Bien and Tibshirani \(2011\)](#) by using a shifted version of the soft-thresholding.

The rest of this paper is organized as follows. Section 2 first presents the likelihood ratio and the existing charts for the joint monitoring of the mean vector and the covariance matrix. Two new control charts are then developed based on the penalized likelihood estimates of the process mean vector and the covariance matrix. Adaptive versions of the proposed charts are proposed to circumvent the need of tuning parameter selection. In Section 3, the performance of the proposed charts are studied and compared with the existing charts. Some guidelines for designing the proposed charts are also discussed. Finally, Section 4 concludes this work with suggestions for future research.

2. The proposed control charting mechanism

In this section, we first present the likelihood ratio for testing the simultaneous changes of process mean/variance/correlation. We then derive two versions of the control chart based on the penalized likelihood estimate.

2.1. The log likelihood ratio and the existing charts

Let \mathbf{x} represent a p -dimensional quality characteristic to be monitored and we assume that \mathbf{x} follows a p -dimensional normal distribution denoted as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and the covariance matrix of the distribution, respectively. When the process is IC, we assume that $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$. Here $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the known IC mean vector and covariance matrix. Therefore, to detect changes in either $\boldsymbol{\mu}$ or $\boldsymbol{\Sigma}$, it is equivalent to testing the hypothesis that $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ v.s. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ or $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$.

If only an individual observation, \mathbf{x}_t , is available at any given sampling period t , it is impossible to calculate the sample covariance matrix or test the above hypothesis. As an alternative, [Huwang et al. \(2007\)](#) first proposed to accumulate at each time period t the term $(\mathbf{x}_t - \boldsymbol{\mu}_0)(\mathbf{x}_t - \boldsymbol{\mu}_0)^T$ using an EWMA equation. The same idea was also proposed in [Hawkins and Maboudou-Tchao \(2008\)](#). If the process mean, $\boldsymbol{\mu}_0$ is known, the MEWMC chart proposed by [Hawkins and Maboudou-Tchao \(2008\)](#) for detecting changes in a covariance matrix operators as follows, for $t \leq 1$,

$$\begin{cases} \mathbf{S}_t^0 = \omega(\mathbf{x}_t - \boldsymbol{\mu}_0)(\mathbf{x}_t - \boldsymbol{\mu}_0)^T + (1 - \omega)\mathbf{S}_{t-1}^0 \\ -2LR_t = -\ln |\mathbf{S}_t^0| + \text{tr}(\mathbf{S}_t^0), \end{cases}$$

where $\mathbf{S}_0^0 = \mathbf{I}_p$, a p -dimensional identity matrix, $0 < \omega < 1$ is a smoothing constant, and $-2LR_t$ is the log-likelihood ratio up to some constant. Note that the observation \mathbf{x}_t is already standardized to have an identity covariance matrix in [Hawkins and Maboudou-Tchao \(2008\)](#).

The above charting statistic $-2LR_t$ is derived from the GLR statistic for detecting changes in the covariance matrix. That is, the MEWMC chart is designed under the assumption that the process mean vector does not shift. Since we are interested in detecting changes in both the mean vector and the covariance matrix in this work, we modify the above procedure to derive a GLR chart that can jointly monitor the mean vector and the covariance matrix:

$$\begin{cases} \boldsymbol{\mu}_t = \alpha \mathbf{x}_t + (1 - \alpha)\boldsymbol{\mu}_{t-1} \\ \mathbf{S}_t^0 = \omega(\mathbf{x}_t - \boldsymbol{\mu}_0)(\mathbf{x}_t - \boldsymbol{\mu}_0)^T + (1 - \omega)\mathbf{S}_{t-1}^0 \\ \mathbf{S}_t^\mu = \omega(\mathbf{x}_t - \boldsymbol{\mu}_t)(\mathbf{x}_t - \boldsymbol{\mu}_t)^T + (1 - \omega)\mathbf{S}_{t-1}^\mu \\ -2LR_t = -\ln |\mathbf{S}_t^\mu| + \text{tr}(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}_t^0). \end{cases} \quad (1)$$

Here $\boldsymbol{\mu}_0 = \mathbf{0}$, $\mathbf{S}_0^0 = \mathbf{I}_p$ and $0 < \alpha, \omega < 1$ are both smoothing parameters. Note that, as shown in [Huwang et al. \(2007\)](#), the accumulated \mathbf{S}_t^0 is also sensitive to mean shifts, while the accumulated \mathbf{S}_t^μ is not affected by the mean shifts. [Huwang et al. \(2007\)](#) used \mathbf{S}_t^0 and \mathbf{S}_t^μ to develop the MEWMS and the MEWMC charts, respectively, for monitoring the changes in the covariance matrix. It is interesting to point out that when $p = 1$, the \mathbf{S}_t^0 and \mathbf{S}_t^μ reduce to, respectively, the EWMS and EWMV

statistics studied by MacGregor and Harris (1993). We name this chart the GLR chart whose performance will be compared to that of our proposed charts in Section 3.

Although the MEWMC chart proposed by Hawkins and Maboudou-Tchao (2008) is designed only for detecting changes in the covariance matrix, the authors also suggested combining the conventional MEWMA chart, together with the MEWMC chart, for the joint monitoring of changes in the mean vector and the covariance matrix. Hawkins and Maboudou-Tchao (2008) called this combined scheme the MEWMA-C chart whose performance will also be compared to that of our proposed charts in Section 3.

2.2. The proposed control charts

Both of the afore-mentioned GLR chart and MEWMC chart are derived from the likelihood ratio test statistic for testing the hypothesis that $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ v.s. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ or $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$. That is, when the process changes, the mean vector or the covariance matrix may shift to any direction. Therefore, the maximum likelihood estimators (MLE) for the mean vector and the covariance matrix are used. However, in a multivariate process, as Wang and Jiang (2009) and Li et al. (2013) pointed out, it is more common that the shift occurs among a subset of the variables. That is, when a mean vector or a covariance matrix of a multivariate process changes, some of the elements in the mean vector or the covariance matrix remain unchanged. Taking this information into consideration, we modify the GLR chart and derive a penalized version of the charting statistic.

Compared to the MLE, the benefits of using sparse estimates of the mean vector and the covariance matrix are two-fold. Firstly, a reduction of the degrees of freedom is attained, which is expected to lead to an improvement in the detecting power. For a p -dimensional process, the usual likelihood estimator would need to estimate p elements for the mean vector, and $p(p+1)/2$ elements for the covariance matrix. However, sparse estimates for the mean vector and the covariance matrix contain zeros, which reduce the degrees of freedom. The sparse estimation rules out some or many candidate parameters under monitoring and makes the resulting charting statistic concentrate more on those parameters which are more likely to incur changes. Secondly, the penalized chart is helpful to root cause diagnosis. If the chart triggers an alarm, the non-zero elements in the sparse estimates are more likely to cause the alarm. Such information would guide the practitioners with a more focused search for the root causes in the process. In the following, we propose two penalized charts, both of which are derived from the sparse estimates of the unknown mean vector and covariance matrix.

The first proposed chart originates from the GLR statistic in Eq. (1). The usual GLR testing statistic uses no information of the alternative hypothesis. However, based on the discussion about the sparsity of potential shifts and utilizing the penalized likelihood estimation method, we modify the GLR chart as follows. For $t \geq 1$, the observations are first accumulated using an EWMA equation. A sparse estimate of the mean vector is then derived. To obtain a sparse estimate of the covariance matrix using individual observations, we also need to accumulate the samples in certain ways. Here we obtain two versions, one using the constant IC mean vector and the other using the sparse mean vector estimate, to obtain the sample covariance. The corresponding sparse estimates are then calculated. Finally, the likelihood ratio test statistic using the sparse mean vector and covariance matrix estimates is calculated as the charting statistic. The detailed steps are outlined as follows. For $t \geq 1$,

1. Obtain the MEWMA estimate of the mean vector,

$$\boldsymbol{\mu}_t = \alpha \mathbf{x}_t + (1 - \alpha) \boldsymbol{\mu}_{t-1}, \quad (2)$$

where $0 < \alpha < 1$ is a smoothing constant.

2. Solve for $\hat{\boldsymbol{\mu}}_{\lambda t}$, which is a penalized and sparse estimate of the process mean vector, by minimizing

$$\hat{\boldsymbol{\mu}}_{\lambda t} = \arg \min_{\boldsymbol{\mu}} \left\{ (\boldsymbol{\mu}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_t - \boldsymbol{\mu}) + \lambda_1 \sum_{i=1}^p |\mu_i| \right\}, \quad (3)$$

where λ_1 is a penalty coefficient, and μ_i is the i th element of vector $\boldsymbol{\mu}$.

3. Obtain the MEWMC for estimating the covariance matrix by incorporating $\hat{\boldsymbol{\mu}}_{\lambda t}$,

$$\mathbf{S}_t^\lambda = \omega (\mathbf{x}_t - \hat{\boldsymbol{\mu}}_{\lambda t})(\mathbf{x}_t - \hat{\boldsymbol{\mu}}_{\lambda t})^T + (1 - \omega) \mathbf{S}_{t-1}^\lambda, \quad (4)$$

where $0 < \omega < 1$ is a smoothing constant.

4. Solve for $\hat{\mathbf{S}}_t^\lambda$ to obtain a sparse estimate of the covariance matrix by minimizing

$$\hat{\mathbf{S}}_t^\lambda = \arg \min_{\mathbf{S}} \left\{ -\ln |\mathbf{S}| - \text{tr}(\mathbf{S}^{-1} \mathbf{S}_t^\lambda) + \lambda_2 \|\mathbf{S} - \boldsymbol{\Sigma}_0\|_1 \right\}, \quad (5)$$

where λ_2 is a penalty coefficient and $\|\mathbf{A} - \mathbf{B}\|_1$ is the L_1 -norm of the difference between two matrices \mathbf{A} and \mathbf{B} , which is equal to the sum of the absolute values of all element-wise differences of the two matrices.

5. Obtain the MEWMC for estimating the covariance matrix using $\boldsymbol{\mu}_0$,

$$\mathbf{S}_t^0 = \omega (\mathbf{x}_t - \boldsymbol{\mu}_0)(\mathbf{x}_t - \boldsymbol{\mu}_0)^T + (1 - \omega) \mathbf{S}_{t-1}^0.$$

6. Calculate the following charting statistic using the above penalized estimates

$$PGLR_t = -\ln |\hat{\mathbf{S}}_t^\lambda| - \text{tr}(\mathbf{S}_t^\lambda (\hat{\mathbf{S}}_t^\lambda)^{-1}) + \text{tr}(\boldsymbol{\Sigma}_0^{-1} \mathbf{S}_t^0). \quad (6)$$

We will call the control chart which uses $PGLR_t$ as the charting statistic the pGLR-chart (the penalized GLR-chart), and it triggers an alarm if $PGLR_t > h$, where h is a predetermined control limit.

The first two terms in Eq. (6) correspond to the log-likelihood function under the alternative hypothesis and the last term corresponds to the log-likelihood function under the null hypothesis subject to a constant. Note that the penalty coefficients λ_1 and λ_2 affect the sparsity of $\hat{\boldsymbol{\mu}}_{\lambda t}$ and $\hat{\mathbf{S}}_t^\lambda$. The penalized estimates are more sparse and hence are expected to improve the chart performance. Similar to the GLR chart, the pGLR chart is capable of detecting changes in both the mean vector and the covariance matrix.

It is worth noting that in step 4 (Eq. (5)), a sparse estimate of the covariance matrix is obtained by penalizing the difference between an estimate \mathbf{S} and the IC covariance matrix $\boldsymbol{\Sigma}_0$. This treatment is quite different from that used by Yeh et al. (2012) and Li et al. (2013). Li et al. (2013) penalized $\|\mathbf{S}^{-1}\|_1$, which has the effect of forcing all the elements of \mathbf{S}^{-1} to shrink toward zero, including the diagonal variance components. Yeh et al. (2012) penalized $\|\mathbf{S}^{-1} - \boldsymbol{\Sigma}_0^{-1}\|_1$, which has the effect of forcing all the elements of \mathbf{S}^{-1} to move toward those of $\boldsymbol{\Sigma}_0^{-1}$. The target of both works is the precision matrix. Yeh et al. (2012) and Li et al. (2013) mentioned that zeros in the precision matrix indicate conditional independence of the original variables, which have certain engineering implications in practice. Bien and Tibshirani (2011) provided an algorithm for estimating the covariance matrix when $\|\mathbf{S}\|_1$ is penalized. In this work, we modify the algorithm due to Bien and Tibshirani (2011) so that the penalty is meaningful in the SPC context. That is, the difference between the estimated and the IC covariance matrix, $\|\mathbf{S} - \boldsymbol{\Sigma}_0\|_1$, is penalized directly, which could give better diagnostic information if an alarm is triggered.

The second proposed chart is motivated by the MEWMA scheme used in Hawkins and Maboudou-Tchao (2008). The MEWMA scheme consists of two separate charts, the MEWMA and the MEWMC charts. The MEWMA chart is expected to detect mean shifts, and the MEWMC chart for covariance changes. Based on the assumption of sparse changes in both the mean vector and the covariance matrix, we re-design the MEWMA scheme, using the sparse estimates of the mean vector and the covariance matrix in the control chart. More specifically, we first obtain a sparse estimate of the process mean vector, $\hat{\boldsymbol{\mu}}_{\lambda t}$, using the same steps 1 and 2 in constructing the pGLR chart. Then, instead of accumulating the observations for estimating $\boldsymbol{\Sigma}$ by accounting for possible mean shifts as in step 3 of the pGLR-chart, we simply use \mathbf{S}_t^0 to find a sparse estimate of $\boldsymbol{\Sigma}$, since when calculating \mathbf{S}_t^0 , the process mean vector is assumed to be IC and thus $\boldsymbol{\mu}_0$ is used in the update equation. Finally, the penalized MEWMA chart assumes that the covariance matrix is IC, and employs the T^2 statistic for detecting mean changes, while the penalized MEWMC chart monitors the likelihood ratio for detecting changes in the covariance matrix which assumes that the mean vector is IC.

The procedures for the second proposed chart are listed as follows. For $t \geq 1$,

1. Obtain the MEWMA for estimating the mean vector,

$$\boldsymbol{\mu}_t = \alpha \mathbf{x}_t + (1 - \alpha) \boldsymbol{\mu}_{t-1},$$

where $0 < \alpha < 1$ is a smoothing constant.

2. Solve for $\hat{\boldsymbol{\mu}}_{\lambda t}$ by minimizing

$$\hat{\boldsymbol{\mu}}_{\lambda t} = \arg \min_{\boldsymbol{\mu}} \left\{ (\boldsymbol{\mu}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_t - \boldsymbol{\mu}) + \lambda_1 \sum_{i=1}^p |\mu_i| \right\},$$

where λ_1 is a penalty coefficient.

3. Obtain the MEWMC for estimating the covariance matrix using $\boldsymbol{\mu}_0$,

$$\mathbf{S}_t^0 = \omega (\mathbf{x}_t - \boldsymbol{\mu}_0)(\mathbf{x}_t - \boldsymbol{\mu}_0)^T + (1 - \omega) \mathbf{S}_{t-1}^0,$$

where $0 < \omega < 1$ is a smoothing constant.

4. Solve for $\hat{\mathbf{S}}_{\lambda t}^0$ by minimizing the following penalized likelihood ratio

$$\hat{\mathbf{S}}_{\lambda t}^0 = \arg \min_{\mathbf{S}} \left\{ -\ln |\mathbf{S}| - \text{tr}(\mathbf{S}^{-1} \mathbf{S}_t^0) + \lambda_2 \|\mathbf{S} - \boldsymbol{\Sigma}_0\|_1 \right\},$$

where λ_2 is a penalty coefficient.

5. Finally, calculate the following charting statistics:

$$T_1^2 = (\hat{\boldsymbol{\mu}}_{\lambda t} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\hat{\boldsymbol{\mu}}_{\lambda t} - \boldsymbol{\mu}_0)$$

and

$$T_2^2 = -\ln |\hat{\mathbf{S}}_{\lambda t}^0| - \text{tr}(\mathbf{S}_t^0 (\hat{\mathbf{S}}_{\lambda t}^0)^{-1}) + \text{tr}(\boldsymbol{\Sigma}_0^{-1} \mathbf{S}_t^0).$$

The proposed chart monitors the two charting statistics T_1^2 and T_2^2 and it triggers an alarm if $T_1^2 > h_1$ or $T_2^2 > h_2$, where h_1 and h_2 are predetermined control limits. Since the above T_1^2 and T_2^2 are essentially the penalized versions of the MEWMA and MEWMC charts, we name it the penalized MEWMAC (pMEWMAC) chart. By combining these two separate charts, we expect that the proposed pMEWMAC chart can help remove noises in the observations and thus will result in better chart performance.

2.3. The adaptive charting scheme

To implement the charts proposed above, we will need to choose two penalizing tuning parameters λ_1, λ_2 . The choice of the tuning parameters is a delicate issue. In practice, the performance of the control charts varies when the tuning parameters are differently chosen. Li et al. (2013) used fixed penalty parameters in their chart, but studied the effect of different choices in monitoring process variability. Zou and Qiu (2009) utilized an adaptive procedure and suggested selecting the tuning parameter from the Lasso solution path. To address this issue, here we follow Zou and Qiu (2009) to adopt an adaptive tuning parameter selection scheme. The adaptive strategy searches a set of tuning parameter combinations and picks the most suitable one automatically adapting to the potential out-of-control pattern.

Specifically, let $T_\lambda, \lambda = (\lambda_1, \lambda_2)$ be any one of the above charting statistics with tuning parameter vector λ . Note that T_λ represents a generic charting statistic, it could be the pGLR statistic in the pGLR method, or either the T_1^2 or T_2^2 in the pMEWMAC method. We only manifest its dependence on the tuning parameters λ_1, λ_2 . Denote q as the size of the candidate set for the tuning parameter combination $\lambda = (\lambda_1, \lambda_2)$. The newly defined adaptive version of the charting statistic is

$$T^{(a)} = \max_{j \in \{1, \dots, q\}} \frac{T_{\lambda^{(j)}} - E(T_{\lambda^{(j)}})}{\sqrt{\text{Var}(T_{\lambda^{(j)}})}} \quad (7)$$

where $\lambda^{(j)} = (\lambda_1^{(j)}, \lambda_2^{(j)})$, $j \in \{1, \dots, q\}$ are a candidate set of the smoothing parameter combination over which the maximum of the charting statistic is calculated, $\lambda_1^{(j)}$ and $\lambda_2^{(j)}$ are chosen from a list of pre-specified values, and $E(T_{\lambda^{(j)}})$ and $\text{Var}(T_{\lambda^{(j)}})$ are the mean and variance of the charting statistic $T_{\lambda^{(j)}}$, respectively. In practice, these quantities are estimated by simulations.

The idea of searching a range of charting statistics using different tuning parameters was originated in Horowitz and Spokoiny (2001). In essence, the adaptive method uses different “optimal” tuning parameters in different scenarios automatically. The theory developed by Horowitz and Spokoiny (2001), in an analogous setting, shows that the adaptive scheme judiciously uses advisable tuning parameters in the sense that the failure pattern can be detected in a timely manner by using the implicitly selected tuning parameters.

Generally speaking, there are no detailed guidelines in the literature on how to choose the candidate set. Ideally, we want the candidate set of tuning parameters of λ_1 and λ_2 to contain the “oracle” pairs such that the resulting penalized solution leads us to the true out-of-control signal. Therefore, in practice, without considering computational burden, we would choose a dense grid over a wide range blindly. By doing that, it is highly likely that the resulting procedures will be adaptive to the true failure pattern. In our numerical studies, we simply chose 9 grid points partly motivated by preliminary simulations for the sake of saving computation time.

3. Performance study and chart design guidelines

In this section, we study the performance of the proposed pGLR and pMEWMAC charts and compare it with that of the existing charts. The chart performance is based on the average run length (ARL), where the run length is defined as the number of observations taken before the first OC signal shows up on a control chart. Since the key idea here is the use of the sparse estimates in constructing the proposed charts, we first demonstrate the performance of the sparse estimators.

3.1. Performance of the lasso penalty

Without loss of generality, we assume that when the process is IC, each observation follows a normal distribution, $\mathbf{x}_t \sim N(0, \mathbf{I}_p)$, where $p = 5$ is the dimension of the process. We sequentially generate 50 random observations from the IC distribution, then use Eqs. (2)–(5) to accumulate the observations and obtain the sparse estimates defined in Eqs. (3) and (5). The smoothing parameters are set to $\alpha = \omega = 0.2$. Note that the desired sparsity resulted from the penalized estimation hinges on proper selection of the tuning parameters. To better demonstrate the sparsity-hunting effect of the lasso estimation, we arbitrarily use a fixed combination of $\lambda_1 = 0.2$ and $\lambda_2 = 0.5$ for illustration. The adaptive choice of the tuning parameters will be implemented when investigating the ARL performance of the charts in the next section.

Based on the EWMA accumulation defined in Eq. (2), we obtain

$$\boldsymbol{\mu}_t = (-0.272, -0.111, 0.023, -0.480, 0.275)^T.$$

After applying the penalty in the estimation, a sparse estimate is obtained as follows:

$$\hat{\boldsymbol{\mu}}_{\lambda_1 t} = (-0.222, -0.061, \mathbf{0}, -0.430, 0.225)^T,$$

which has the third element being set to exactly zero.

The MEWMC when incorporating $\hat{\boldsymbol{\mu}}_{\lambda_1 t}$ from Eq. (4) is:

$$\hat{\mathbf{S}}_t^{\lambda_1} = \begin{bmatrix} 1.165 & -0.123 & -0.145 & 0.128 & -0.122 \\ -0.123 & 0.237 & -0.042 & 0.058 & -0.046 \\ -0.145 & -0.042 & 0.414 & -0.028 & -0.109 \\ 0.128 & 0.058 & -0.028 & 0.203 & -0.148 \\ -0.122 & -0.046 & -0.109 & -0.148 & 0.478 \end{bmatrix}.$$

After applying the penalty, a sparse estimate is obtained as follows:

$$\hat{\mathbf{S}}_t^{\lambda_2} = \begin{bmatrix} 1.000 & -0.020 & \mathbf{0} & 0.039 & \mathbf{0} \\ -0.020 & 0.256 & \mathbf{0} & 0.041 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0.561 & \mathbf{0} & \mathbf{0} \\ 0.039 & 0.041 & \mathbf{0} & 0.190 & -0.119 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -0.119 & 0.575 \end{bmatrix},$$

where a number of elements were being shrunk to exactly zero. The sparse estimate is also closer to the true IC covariance matrix \mathbf{I}_p .

Now assume that the process is shifted to an OC distribution having

$$\boldsymbol{\mu}_{OC} = (0.5, 0.5, 0.5, 0, 0)^T$$

and

$$\boldsymbol{\Sigma}_{OC} = \begin{bmatrix} 1.5 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 1.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 1.5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

By generating 50 random observations and treating them using Eqs. (2)–(5), the ordinary and penalized estimates of the mean vector and the covariance matrix are

$$\boldsymbol{\mu}_t = (-0.143, 0.637, 0.252, -0.111, 0.023)^T$$

$$\hat{\boldsymbol{\mu}}_{\lambda_1 t} = (-0.093, 0.587, 0.202, -0.061, \mathbf{0})^T,$$

and

$$\hat{\mathbf{S}}_t^{\lambda_1} = \begin{bmatrix} 0.708 & 0.233 & 0.498 & 0.073 & 0.093 \\ 0.233 & 0.862 & -0.246 & 0.098 & 0.220 \\ 0.498 & -0.246 & 0.907 & -0.062 & 0.024 \\ 0.073 & 0.098 & -0.062 & 0.237 & -0.042 \\ 0.093 & 0.220 & 0.024 & -0.042 & 0.414 \end{bmatrix},$$

$$\hat{\mathbf{S}}_t^{\lambda_2} = \begin{bmatrix} 0.686 & 0.197 & 0.508 & 0.049 & \mathbf{0} \\ 0.197 & 1.000 & -0.171 & \mathbf{0} & \mathbf{0} \\ 0.508 & -0.171 & 0.932 & -0.011 & \mathbf{0} \\ 0.049 & \mathbf{0} & -0.011 & 0.256 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0.564 \end{bmatrix},$$

respectively. It is found that both the sparse mean vector and covariance matrix estimates preserve a certain number of zeros corresponding to the set of unchanged mean and variance/covariance components. By incorporating the sparse and more targeted estimates of the unknown population parameters into designing the proposed charts, it is expected that the performance of the proposed charts will be improved.

3.2. The ARL performance study and comparison

In this section, we study and compare the ARL performance of the two penalized charts, the pGLR and the pMEWMC charts, with that of their counterparts, the GLR and the MEWMC charts. For illustrative purpose, the IC parameters of the process are set to $\boldsymbol{\mu}_0 = \mathbf{0}_p$, and $\boldsymbol{\Sigma}_0 = \mathbf{I}_p$, with $p = 10$. It should be noted that the proposed charts can be applied to processes with any general IC mean vector and covariance matrix. A real example with such a general setting will be demonstrated later in Section 3.4.

Table 1
Out-of-control shift patterns for performance comparison.

Pattern number	Shifted elements
OC1	μ_1
OC2	μ_1, μ_2, μ_3
OC3	μ_1, \dots, μ_5
OC4	σ_{11}, σ_{22}
OC5	$\sigma_{11}, \dots, \sigma_{13}, \sigma_{21}, \dots, \sigma_{23}, \sigma_{31}, \dots, \sigma_{33}$
OC6	$\sigma_{11}, \dots, \sigma_{15}, \sigma_{21}, \dots, \sigma_{25}, \dots, \sigma_{51}, \dots, \sigma_{55}$
OC7	OC1 and OC4 together
OC8	OC2 and OC5 together
OC9	OC3 and OC6 together

Table 2
The OC ARL performance of the GLR chart.

δ	Shift patterns								
	1	2	3	4	5	6	7	8	9
0.2	180.6	153.0	127.8	156.1	112.8	70.0	140.3	87.1	48.4
0.4	141.1	74.0	44.0	108.1	55.2	24.9	80.1	27.6	11.9
0.6	91.5	29.8	12.9	74.6	29.7	12.2	42.0	10.6	4.3
0.8	57.1	11.6	3.9	52.2	18.4	7.3	22.5	4.8	2.0
1.0	34.1	4.4	1.5	36.5	12.3	4.9	12.8	2.6	1.3
1.2	19.1	2.0	1.0	26.8	9.0	3.5	7.4	1.6	1.1
1.4	11.1	1.2	1.0	20.2	6.7	2.7	4.6	1.2	1.0
1.6	6.5	1.0	1.0	15.8	5.1	2.2	3.1	1.1	1.0
1.8	3.8	1.0	1.0	12.5	4.1	1.8	2.1	1.0	1.0
2.0	2.3	1.0	1.0	10.1	3.5	1.6	1.6	1.0	1.0
2.2	1.5	1.0	1.0	8.2	2.9	1.4	1.3	1.0	1.0
2.4	1.2	1.0	1.0	7.0	2.5	1.3	1.2	1.0	1.0
2.6	1.1	1.0	1.0	5.9	2.2	1.2	1.1	1.0	1.0
2.8	1.0	1.0	1.0	5.1	2.0	1.2	1.0	1.0	1.0
3.0	1.0	1.0	1.0	4.3	1.8	1.1	1.0	1.0	1.0

For a fair comparison, the IC ARL of all charts is set to 200. As both MEWMAC and pMEWMAC contain two separate charts, the IC ARL of each individual chart is approximately 380, such that the overall IC ARL of the combined charts is still 200. The smoothing parameters of the charts are set to $\alpha = \omega = 0.2$. As only individual observations are available at each step and the EWMA smoothing is used for mean vector and covariance matrix estimation, we use 50 IC observations for warm up and study the steady-state ARL of all charts. As was indicated in Eq. (7), the tuning parameters, λ_1 and λ_2 , will be adaptively chosen to maximize the charting statistic. We here set $\lambda_1 = \{0.05, 0.1, 0.3\}$ and $\lambda_2 = \{0.1, 0.2, 0.4\}$. The possible combinations of (λ_1, λ_2) include $(0.05, 0.1), (0.05, 0.2), (0.05, 0.4), (0.2, 0.1), \dots, (0.3, 0.4)$. Thus, in total, $q = 9$ combinations are utilized in Eq. (7) to search for the maximum.

Nine different OC scenarios, defined in Table 1, are tested against each chart. When a shift occurs, each shifted element in the mean vector or/and the covariance matrix is increased by an amount of δ . Among the OC patterns, OC1–OC3 are for mean shifts, OC4–OC6 are for covariance matrix changes, and OC7–OC9 are for changes in both. We tried a sequence of δ values: 0.2, 0.4, 0.6, . . . , 2.6, 2.8, 3 to examine the performance pattern of the various charts under comparison.

The simulated OC ARLs of the GLR, pGLR, MEWMAC and pMEWMAC charts are summarized in Tables 2–5, respectively. Based on the results shown in Tables 2–5, a number of observations can be made.

1. A comparison between the GLR and the pGLR charts shows that the pGLR-chart consistently outperforms the GLR-chart for all the scenarios we considered (OC1–OC9). This shows that the added penalty term is effective and enhances the performance of the proposed pGLR-chart. The improved performance of the pGLR-chart for detecting the joint shifts makes it more attractive as a control charting mechanism for simultaneously monitoring the mean vector and the covariance matrix of a multivariate normal process.
2. Between the MEWMAC and pMEWMAC charts, the pMEWMAC-chart performs slightly better than the MEWMAC chart for most of the mean shifts (OC1–OC3), but is slightly worse for variance changes (OC4–OC6). Under the joint shifts of OC7–OC9, the two charts perform quite closely. This shows that penalized version of the MEWMAC chart does not improve the performance significantly. One possible explanation is that the MEWMAC chart already efficiently used information in the observations; the gain added by using a penalized method in identifying shift signals is limited in the pMEWMAC chart.
3. A comparison between the pGLR and the pMEWMAC charts indicates that the former is more effective in detecting variance changes in OC4 and OC5, while the later is more powerful in detecting mean and joint shifts.

In summary, applying a penalty to the GLR chart significantly improves its charting performance, while adding a penalty to the MEWMAC chart does not improve its performance significantly. Between the GLR and the pGLR charts, the latter is recommended due to its improved performance. Between the MEWMAC and the pMEWMAC charts, the former is

Table 3
The OC ARL performance of the pGLR chart.

δ	Shift patterns								
	1	2	3	4	5	6	7	8	9
0.2	184.5	142.7	108.1	153.2	108.3	63.1	140.6	77.5	39.2
0.4	122.7	55.1	26.8	100.3	47.8	19.0	73.3	18.7	7.6
0.6	83.4	18.7	6.4	66.3	25.0	9.5	32.9	6.8	2.6
0.8	39.0	5.7	1.9	43.5	15.4	5.4	16.3	3.2	1.5
1.0	23.9	2.3	1.1	27.5	8.8	3.7	8.2	1.7	1.2
1.2	12.9	1.2	1.0	22.1	6.6	2.4	4.8	1.2	1.0
1.4	6.6	1.0	1.0	15.8	4.5	2.0	3.0	1.1	1.0
1.6	3.8	1.0	1.0	12.7	3.6	1.8	2.1	1.0	1.0
1.8	1.9	1.0	1.0	9.7	3.1	1.5	1.4	1.0	1.0
2.0	1.6	1.0	1.0	7.2	2.5	1.4	1.3	1.0	1.0
2.2	1.1	1.0	1.0	6.4	2.1	1.2	1.1	1.0	1.0
2.4	1.0	1.0	1.0	5.0	2.2	1.2	1.1	1.0	1.0
2.6	1.0	1.0	1.0	4.1	1.8	1.1	1.0	1.0	1.0
2.8	1.0	1.0	1.0	3.5	1.6	1.1	1.0	1.0	1.0
3.0	1.0	1.0	1.0	3.1	1.5	1.0	1.0	1.0	1.0

Table 4
The OC ARL performance of the MEWMAC chart.

δ	Shift patterns								
	1	2	3	4	5	6	7	8	9
0.2	169.6	126.2	95.3	144.3	108.3	66.0	123.4	67.2	36.9
0.4	109.8	41.5	19.6	100.0	53.2	24.3	58.5	17.9	8.3
0.6	55.6	11.7	4.0	67.9	29.4	12.4	26.9	6.8	3.1
0.8	27.2	3.6	1.3	47.7	17.9	7.4	13.4	3.2	1.7
1.0	13.4	1.5	1.0	34.2	12.2	4.8	7.4	1.8	1.2
1.2	6.8	1.1	1.0	25.6	8.6	3.5	4.4	1.3	1.1
1.4	3.4	1.0	1.0	19.5	6.8	2.7	2.8	1.1	1.0
1.6	1.9	1.0	1.0	15.4	5.3	2.2	1.9	1.0	1.0
1.8	1.3	1.0	1.0	12.3	4.2	1.9	1.5	1.0	1.0
2.0	1.1	1.0	1.0	9.8	3.5	1.6	1.2	1.0	1.0
2.2	1.0	1.0	1.0	8.1	3.0	1.5	1.1	1.0	1.0
2.4	1.0	1.0	1.0	6.8	2.6	1.3	1.1	1.0	1.0
2.6	1.0	1.0	1.0	5.8	2.2	1.2	1.0	1.0	1.0
2.8	1.0	1.0	1.0	5.0	2.0	1.2	1.0	1.0	1.0
3.0	1.0	1.0	1.0	4.4	1.8	1.1	1.0	1.0	1.0

Table 5
The OC ARL performance of the pMEWMAC chart.

δ	Shift patterns								
	1	2	3	4	5	6	7	8	9
0.2	160.9	123.9	91.3	150.3	124.2	74.8	124.1	64.6	39.5
0.4	104.9	37.9	21.7	108.7	56.6	27.0	52.6	18.8	8.5
0.6	48.1	11.7	5.1	69.3	30.0	14.4	22.2	6.6	3.4
0.8	22.9	3.5	1.4	48.8	19.6	8.6	11.9	3.2	1.8
1.0	11.1	1.5	1.0	34.5	13.6	5.6	6.1	1.9	1.2
1.2	4.6	1.1	1.0	26.8	9.6	3.7	4.0	1.4	1.1
1.4	2.5	1.0	1.0	20.4	6.9	3.0	2.3	1.1	1.0
1.6	1.5	1.0	1.0	14.2	5.7	2.3	1.7	1.0	1.0
1.8	1.1	1.0	1.0	11.9	4.3	2.0	1.3	1.0	1.0
2.0	1.0	1.0	1.0	10.6	4.1	1.6	1.2	1.0	1.0
2.2	1.0	1.0	1.0	8.2	3.2	1.4	1.1	1.0	1.0
2.4	1.0	1.0	1.0	6.3	2.8	1.4	1.0	1.0	1.0
2.6	1.0	1.0	1.0	5.5	2.5	1.2	1.0	1.0	1.0
2.8	1.0	1.0	1.0	5.4	2.1	1.2	1.0	1.0	1.0
3.0	1.0	1.0	1.0	4.5	2.0	1.2	1.0	1.0	1.0

recommended since it has a relatively simpler form with competitive performance. A selection between the pGLR and the MEWMAC charts is subjective. The MEWMAC chart has better overall performance, while the pGLR chart is simpler in that it contains a single chart but the MEWMAC chart contains two separate charts. In addition, the pGLR chart has the capability of suggesting diagnostic information when a shift is detected, which is attractive to practitioners in identifying and removing root causes in practice. Therefore, no chart is uniformly better than all others. This work has proposed a new alternative for the joint monitoring of process mean vector and covariance matrix. From a practitioner's perspective, it is suggested that

a selection should be made by comprehensively evaluating the importance of charting performance, diagnostic capability, and computational complexity.

3.3. Discussion on control chart design and signal diagnosis

In the proposed pGLR and pMEWMA charts, two sets of parameters need to be determined before using the charts: the smoothing constants for the EWMA calculation, and the penalty coefficients for the penalized likelihood estimation. Similar to any EWMA or MEWMA charts, the smoothing constant used for the EWMA calculation of the proposed charts has the effect of smoothing the data series. Typically, a small value would lead to a better chart performance for detecting smaller shifts, while a large value leads to a faster detection of larger shifts. In the conventional EWMA chart design, a small value, usually around 0.1 or 0.2, is suggested. In implementing the proposed charts, similar smoothing constants are recommended.

The effect of the penalty coefficient on the actual sparse estimate thus obtained has been widely discussed in the statistics literature. A larger penalty coefficient is expected to generate more sparsity, and a smaller penalty coefficient corresponds to less sparsity in the resulting mean vector and covariance matrix estimates. Yeh et al. (2012) studied the effect of the penalty coefficient for monitoring the covariance matrix. The authors suggested that the best choice of the penalty coefficient varies with shift magnitudes and shift patterns. The performance of the proposed pGLR and pMEWMA charts is affected by many factors, including the design parameters and the process itself. Therefore, if the shift patterns are known in advance, the penalty coefficient should be chosen to optimize the chart's performance against the potential shift patterns. In addition, engineering knowledge could also be incorporated so that the designed charts could be more focused and perform as expected.

Alternatively, Maboudou-Tchao and Agboto (2013) suggested applying the K -fold cross-validation to the in-control sample, which is also a feasible choice for situations in which not much information about process faults is known. Denote the k th fold subsample by F_k , $k = 1, \dots, K$. One can calculate the penalized mean estimator $\hat{\boldsymbol{\mu}}_{-k}^{\lambda_1}$ and the penalized covariance matrix estimator $\hat{\mathbf{S}}_{-k}^{\lambda_1, \lambda_2}$ accounting for the mean estimate $\hat{\boldsymbol{\mu}}_{-k}^{\lambda_1}$, both of which are estimated based on the in-control sample excluding the k th subgroup. The K -fold cross validation score is then defined as

$$CV(\lambda_1, \lambda_2) = \sum_{k=1}^K \left(n_k \log |\hat{\mathbf{S}}_{-k}^{\lambda_1, \lambda_2}| - \sum_{i \in F_k} (\mathbf{x}^{(i)} - \hat{\boldsymbol{\mu}}_{-k}^{\lambda_1})' (\hat{\mathbf{S}}_{-k}^{\lambda_1, \lambda_2})^{-1} (\mathbf{x}^{(i)} - \hat{\boldsymbol{\mu}}_{-k}^{\lambda_1}) \right),$$

where n_k is the sample size of the k th subgroup F_k and $\mathbf{x}^{(i)}$ is the i th observation in F_k . The λ_1 and λ_2 can be chosen to maximize $CV(\lambda_1, \lambda_2)$.

In this work, we suggest providing a collection of tuning parameters and let the chart select the best one adaptively. In this way, the parameters are determined in a data-driven manner by the observations. In this paper, we designed the candidate set to contain both small, moderate and large parameter values. Although the size of the candidate set is still small, it already suggests quite competitive performance. With faster computing power, a larger candidate set could be used, which is expected to result in better chart performance.

It is worth noting that one additional benefit of using the proposed chart is their capability in assisting in root cause diagnosis. In practical SPC applications, an alarm is expected to be followed by a diagnostic procedure. The conventional GLR, MEWMA or MEWMC charts do not provide direct information regarding which variables are more likely to be responsible for the alarm. In Wang and Jiang (2009), the authors treated the non-zero coefficients obtained from a variable-selection procedure as potential shift direction, and designed a directionally-variant chart that improves the detecting power along the identified shift direction. When an OC signal is triggered, these selected variables having non-zero coefficients are suspects responsible for the signal. In this work, a similar procedure is followed. At step t , $\hat{\boldsymbol{\mu}}_{\lambda_t}$ in Eq. (3) and $\hat{\mathbf{S}}_t^{\lambda}$ in Eq. (5) are the last estimates of process mean vector and covariance matrix. Therefore, if an OC alarm is triggered, the variables having non-zero coefficients and elements in $\hat{\boldsymbol{\mu}}_{\lambda_t}$ and $\hat{\mathbf{S}}_t^{\lambda}$ should be considered responsible for the signal. Such information could benefit the engineers for physical root cause identification and removal.

Zou et al. (2011) proposed a diagnostic framework using the penalized method by suggesting a BIC type tuning parameter selection scheme. Their method is based on the recently developed extended BIC scheme for tuning parameter selection in the high dimensional penalized method (Chen and Chen, 2008). Such a method can be used for tuning parameter selection in a Shewhart chart. As it requires a set of Phase II samples, it is not suitable for the EWMA on-line monitoring scenario we mainly considered. However, it will be useful for diagnostics if an alarm is triggered in the monitoring process. As long as an alarm is triggered, we can use the collected Phase II samples to estimate the mean vector and the covariance matrix, possibly adding lasso type penalty to obtain sparse and more focused failure pattern. We can use the extended BIC criterion suggested by Zou et al. (2011) to select the tuning parameter in this step.

3.4. Application to a real example

In studying the MEWMC chart, Hawkins and Maboudou-Tchao (2008) applied the MEWMA chart, which combines separate MEWMC and MEWMA charts, to a real data set collected from ambulatory monitoring. Four physiological variables, mean systolic blood pressure (SBP), mean diastolic blood pressure (DBP), mean heart rate (HR) and overall mean arterial

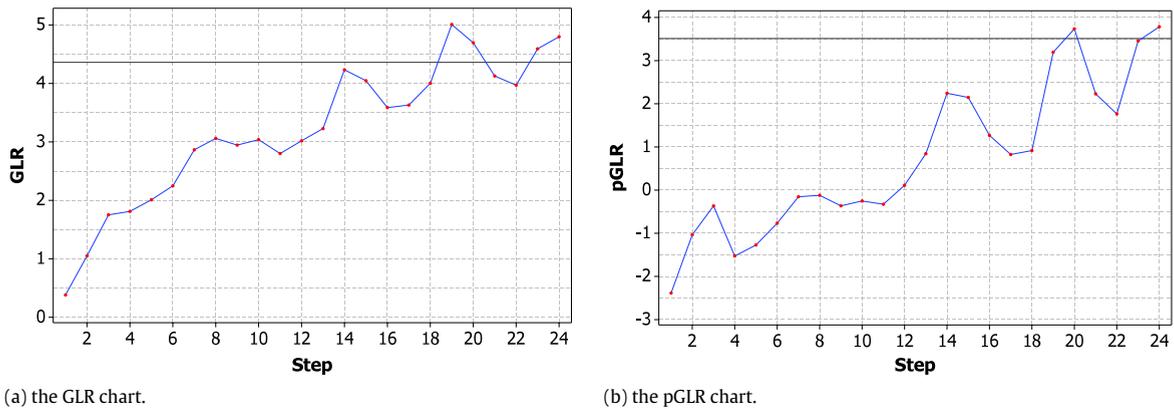


Fig. 2. The applications of the GLR chart and the pGLR chart to the real example.

pressure (MAP), of 24 samples are measured (please refer to Hawkins and Maboudou-Tchao (2008) for a more detailed description of the data set). The IC mean vector and covariance matrix of the process are given as follows:

$$\mu_0 = (126.61, 77.48, 80.95, 97.97)^T$$

and

$$\Sigma_0 = \begin{pmatrix} 15.04 & 8.66 & 10.51 & 12.04 \\ 8.66 & 5.83 & 5.56 & 7.50 \\ 10.51 & 5.56 & 15.17 & 8.79 \\ 12.04 & 7.50 & 8.79 & 10.57 \end{pmatrix}.$$

In the following, for demonstration purpose, we apply the GLR chart and the proposed pGLR chart to the same data set and compare their performance in detecting the process changes (the MEWMAC and pMEWMAC charts could be applied in the same way). Both charts use $\alpha = \omega = 0.2$ in accumulating historical observations. The candidate set of tuning parameters used in the this example is the same as that used in the simulation study. That is, $\lambda_1 = \{0.05, 0.1, 0.3\}$ and $\lambda_2 = \{0.1, 0.2, 0.4\}$, and 9 combinations in total are used. We first use simulation to obtain the control limits of both charts so that the IC ARL is approximately 200, then monitor the 24 samples using the two charts.

The GLR chart and the pGLR chart are shown in Fig. 2. Both charts show a similar trend when the process evolves. For this particular data set, the GLR chart triggered the first alarm at sample 19, while the pGLR chart first signaled at sample 20.

As mentioned earlier, the penalized estimates can help provide clues for process diagnosis. Therefore, we extract the sparse estimates of process mean and covariance matrix at sample 20, which is when the first OC signal showed up on the pGLR chart, and compare them with their in-control values. The sparse estimate of the mean difference is

$$\mu_{\text{diff}} = \hat{\mu}_{\lambda,20} - \mu_0 = (0, 0, 0, -0.645)^T,$$

and the difference between the estimated covariance matrix and the IC covariance matrix is

$$\Sigma_{\text{diff}} = \hat{\Sigma}_{20}^{\lambda} - \Sigma_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.006 & 0 & -0.139 \\ 0 & 0 & 0 & 0 \\ 0 & -0.139 & 0 & 0.101 \end{pmatrix}.$$

The sparse estimates of the mean vector and the covariance matrix suggest two possible shift sources: the mean shift of x_4 , and/or the variance/covariance changes of the variables having non-zero elements in Σ_{diff} . In explaining the OC signal, Hawkins and Maboudou-Tchao (2008) analyzed the regression-adjusted variables and suggested that x_3 (corrected for x_1 and x_2) became less variable, whereas x_4 (corrected for x_1 , x_2 and x_3) became more variable. From Σ_{diff} , it can be seen that an increase in the variance of x_4 is observed. In addition, a decrease in the covariance between x_1 and x_3 is also detected. The observations suggested by the pGLR chart are generally consistent with those observed in Hawkins and Maboudou-Tchao (2008). Please note that Hawkins and Maboudou-Tchao (2008) paid special efforts after the signal to obtain diagnostic information. The pGLR chart, on the other hand, can readily provide such information after variable selection, and is therefore helpful to practitioners for further root cause diagnosis.

4. Conclusions

In a multivariate process, it is common that when a process change occurs, only a small set of variables are affected, which leads to sparse change patterns in the mean vector or/and the covariance matrix. Taking such information into consideration,

we develop, in this work, two control charts for simultaneously monitoring the mean vector and the covariance matrix in a multivariate process in which only individual observations are available.

Through the simulation studies, we demonstrated that using the penalized estimates is helpful for removing the noises in the observations and generating sparse estimates. When incorporating these sparse estimates into constructing the proposed charts, the chart performance is further improved in most cases. In addition, the penalized versions of the GLR chart and the MEWMAC chart, the pGLR chart and the pMEWMAC chart, respectively, can assist in identifying the shifted variables or components, which are helpful for further root-cause diagnosis.

The current work only demonstrates the potential of incorporating the sparse estimates to improve the chart performance. However, more research efforts are needed along this line. Unlike the conventional charts which have been extensively studied in the literature, the robustness and the flexible design of the penalized chart, and its use in diverse SPC applications, are all worthy of further investigations in the future.

Acknowledgments

The authors wish to thank the Associate Editor and the anonymous referees for numerous insightful comments which improved the paper greatly.

Dr. Wang's work was supported by the National Natural Science Foundation of China under Grant No. 71072012 and Tsinghua University Initiative Scientific Research Program. Dr. Li's work was supported by the National Natural Science Foundation of China under Grant No. 71272029 and the Beijing Higher Education Young Elite Teacher Project under Grant No. YETP0135.

References

- Bien, J., Tibshirani, R.J., 2011. Sparse estimation of a covariance matrix. *Biometrika* 98, 807–820.
- Bühlmann, P.L., Van de Geer, S., 2011. *Statistics for High-dimensional Data*. Springer.
- Chen, J., Chen, Z., 2008. Extended bayesian information criteria for model selection with large model spaces. *Biometrika* 95, 759–771.
- Chen, G., Cheng, S.W., Xie, H., 2001. Monitoring process mean and variability with one EWMA chart. *J. Qual. Technol.* 33, 223–233.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *Linear Elements of Statistical Learning*, second ed., Springer.
- Hawkins, D.M., Maboudou-Tchao, E.M., 2008. Multivariate exponentially weighted moving covariance matrix. *Technometrics* 50, 155–166.
- Hawkins, D.M., Zamba, K., 2005. Statistical process control for shifts in mean or variance using a changepoint formulation. *Technometrics* 47, 164–173.
- Horowitz, J.L., Spokoiny, V.G., 2001. An adaptive, rate optimal test of a parametric mean regression model against a nonparametric alternative. *Econometrica* 69, 599–631.
- Huwang, L., Yeh, A.B., Wu, C.W., 2007. Monitoring multivariate process variability for individual observations. *J. Qual. Technol.* 39, 258–278.
- Jiang, W., Wang, K., Tsung, F., 2012. A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis. *J. Qual. Technol.* 44, 209–230.
- Li, B., Wang, K., Yeh, A.B., 2013. Monitoring the covariance matrix via penalized likelihood estimation. *IE Trans.* 45, 132–146.
- Li, Z., Zhang, J., Wang, Z., 2010. Self-starting control chart for simultaneously monitoring process mean and variance. *Int. J. Prod. Res.* 48, 4537–4553.
- Maboudou-Tchao, E.M., Agboto, V., 2013. Monitoring the covariance matrix with fewer observations than variables. *Comput. Statist. Data Anal.* 99–112.
- Maboudou-Tchao, E.M., Diawara, N., 2013. A lasso chart for monitoring the covariance matrix. *Qual. Technol. Quantitative Manag.* 10, 95–114.
- MacGregor, J., Harris, T., 1993. The exponentially weighted moving variance. *J. Qual. Technol.* 25, 106–118.
- Montgomery, D.C., 2005. *Introduction to Statistical Quality Control*, fifth ed. John Wiley, Hoboken, NJ.
- Wang, K., Jiang, W., 2009. High-dimensional process monitoring and fault isolation via variable selection. *J. Qual. Technol.* 41, 247–258.
- Yeh, A.B., Li, B., Wang, K., 2012. Monitoring multivariate process variability with individual observations via penalized likelihood estimation. *Int. J. Prod. Res.* 50, 6624–6638.
- Zhou, Q., Luo, Y., Wang, Z., 2010. A control chart based on likelihood ratio test for detecting patterned mean and variance shifts. *Comput. Statist. Data Anal.* 54, 1634–1645.
- Zou, C., Jiang, W., Tsung, F., 2011. A lasso-based diagnostic framework for multivariate statistical process control. *Technometrics* 53, 297–309.
- Zou, C., Qiu, P., 2009. Multivariate statistical process control using lasso. *J. Amer. Statist. Assoc.* 104, 1586–1596.