Computers & Industrial Engineering 75 (2014) 102–115

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Self-oriented control charts for efficient monitoring of mean vectors ☆

D.A.O. Moraes [a,*], F.L.P. Oliveira [b], R.C. Quinino [c], L.H. Duczmal [c]

[a] Statistics Department, Universidade Federal de Santa Maria, Cidade Universitária, 97105-900 Santa Maria, Brazil
[b] Statistics Department, Universidade Federal de Ouro Preto, 35400-000 Ouro Preto, Brazil
[c] Statistics Department, Universidade Federal de Minas Gerais, Campus Pampulha, 31270-901 Belo Horizonte, Brazil

ARTICLE INFO

Article history:
Received 4 November 2013
Received in revised form 9 June 2014
Accepted 11 June 2014
Available online 21 June 2014

Keywords:
Quality control
Multivariate statistics
Mean vectors
Simulation
Average run length

ABSTRACT

This work presents a procedure for monitoring the centre of multivariate processes by optimising the noncentrality parameter with respect to the maximum separability between the in- and out-of-control states. Similarly to the Principal Component Analysis, this procedure is a linear transformation but using a different criterion which maximises the trace of two scatter matrices. The proposed linear statistic is self-oriented in the sense that no prior information is given, then it is monitored by two types of control charts aiming to identify small and intermediate shifts. As the control charts performances depend only on the noncentrality parameter, comparisons are made with traditional quadratic approaches, such as the Multivariate Cumulative Sum (MCUSUM), the Multivariate Exponentially Weighted Moving Average (MEWMA) and Hotelling's T^2 control chart. The results show that the proposed statistic is a solution for the problem of finding directions to be monitored without the need of selecting eigenvectors, maximising efficiency with respect to the average run length.

1. Introduction

Following the advancement of the field of automatic process control, many issues concerning machine learning and risk prevention have arisen. In the specific case of monitoring multivariate processes by means of control charts, machines must be calibrated to notify problems in a timely manner. In terms of controlling the quality of multivariate processes, increasing the data dimensionality implies that the likelihood function near the centre of the distribution decreases, and the majority of the data concentrate in the tails of the distribution (Jimenez & Landgrebe, 1998). As the presence of observations near the process centre becomes a very rare event, the industry seeks solutions that can minimise the lag between issues and warnings, which can be viewed as the inertial time from the root cause of the problem until detection.

Analysing this aspect of multivariate data, Haertel and Landgrebe (1999), mostly influenced by Fukunaga (1990), proposed classification methods for digital images by means of feature selection and extraction, eliminating the redundancy of the original variables and optimising the available information on the output variables. While some feature reduction techniques are already well known in the field of pattern recognition and digital image analysis, this methodology remains unexplored in time-varying single-hypothesis problems, which are extensively used in multivariate process control.

Specifically for mean vectors, recently, a proposal of Khoo, Sitt, Wu, and Castagliola (2013) makes use of run sum schemes to improve the performance of various charts with run rules schemes. Although their proposal is less sensitive than the Multivariate Exponentially Weighted Moving Average (MEWMA) chart toward small shifts, the sensitivity in detecting small shifts can be further enhanced by adding more regions and scores, so that this chart is as competitive as the MEWMA chart. Based on appropriate linear transformations of the multivariate data and utilising sequentially MEWMA and cumulative sum (CUSUM) schemes, the present work is also eligible on both objectives, small and large shifts of mean vectors.

Linear transformations are suggested by many authors such as Jackson (1991), Kourti and MacGregor (1996) and Choi, Martin, and Morris (2005), based on Principal Component Analysis (PCA), but these methods often do not provide a rule that optimises the separation of the processes. Additionally, selecting a reduced number of principal components in the multivariate charts often results in poor performance (Bersemis, Psarakis, & Paranetos, 2007) in terms of average run length (ARL). This phenomenon occurs because the orthogonal transformation of the PCA method do not

☆ This manuscript was processed by Area Editor Min Xie.

* Corresponding author. Tel.: +55 5532208000.
E-mail addresses: daltieri@smail.ufsm.br (D.A.O. Moraes), fernandoluiz@iceb.ufop.br (F.L.P. Oliveira), quinino@est.ufmg.br (R.C. Quinino), duczmal@est.ufmg.br (L.H. Duczmal).

http://dx.doi.org/10.1016/j.cie.2014.06.008
0360-8352/

optimise the separability among processes but rather better represent the data using a reduced number of features (Fukunaga, 1990). However, if the purpose is to classify the data, better features can be found using non-orthogonal transformations, which are specifically designed to maximise classification accuracy in terms of the Bayes error. In this work, we propose two control charts for mean vectors that can additionally provide optimum performance in terms of the ARL. Although the same type of linear transformation used in PCA is applied in the present work, a different criterion for optimisation is adopted.

In this paper a procedure is presented for monitoring the mean vector of a multivariate Gaussian process with individual observations by optimising the noncentrality parameter with respect to the maximum separability between the in-control (IC) and out-of-control (OC) states. In this optimal space, the combined control charts yield a faster indication of whether the process is moving from the IC to the OC state for a wide range of distances, from small to large shifts. The structure of the paper follows. Section 2 describes the methodology. Section 3 shows numerical simulations, and Section 4 presents the final remarks.

2. Methodology

2.1. Control charts for multivariate processes

The Statistical Process Control (SPC) approach for controlling process parameters consists of determining whether a given sample is more likely to belong to a known subjacent process or not. Then, the unknown sample could be tagged in one of two categories, the IC state or the OC state. A problem related to the design of control charts lies in recognising a shift from the IC state as soon as possible once it occurs.

This problem is a special case of sequential hypothesis testing in which only the parameters of the reference class, the IC process, are known. The second class, the OC process, is completely unknown until the process has shifted for an undetermined period. In this case, it is impossible to assign an alternative process *a priori*. Such problems are called single class hypotheses (Fukunaga, 1990), as there is no way to examine the wide variety of alternative states before designing a decision rule. For example, for selecting the most similar object from a list of objects, one possible simple decision rule could be to select the object with the features most similar to those of the target, representing the minimum distance in a global sense.

The SPC methodology for monitoring the parameters of a given process generally involves two steps (Montgomery, 2001). The first step is the Phase I stage, where the IC parameters are monitored to generate control boundaries, mostly based on the mean rate of false alarms and ARL. The second step is the Phase II stage, where the samples that arise from the running process are measured and their statistics compared with the IC boundaries. If the statistic computed using the new sample exceeds the boundaries, an alarm is triggered and the process is tagged as OC.

The most applied methods for monitoring mean vectors of a stable process employ Hotelling's T^2 (Hotelling, 1947), the Crosier's Multivariate Cumulative Sum (MCUSUM) (Crosier, 1988), and the Multivariate Exponentially Weighted Moving Average (MEWMA) chart of Lowry, Woodall, and Rigdon (1992). These three charts are directly comparable due to the common dependency of the ARL on the noncentrality parameter. The performance of MCUSUM and the MEWMA control charts when the parameters are estimated were studied in Mahmoud and Maravelakis (2010, 2011). To the monitoring of mean vectors, recent proposals includes the works of Yahav and Galit (2014), Ali and Amirhossein (2013), Lee (2013) and Wang and Reynolds (2013). Proposals for simultaneous monitoring of mean vectors and covariance matrices were developed by Zhang, Li, and Wang (2010), Khoo, The, and Wu (2010) and Costa and Machado (2013), with substantial improvements in the monitoring process. For the monitoring of covariance matrices, traditional control chart approaches include the use of moving ranges or generalised variance tests (Montgomery, 2001), while new approaches include the sample ranges method (Costa & Machado, 2011), the VMAX procedure (Costa & Machado, 2008) and the use of auxiliary information proposed by Riaz and Does (2008). In the case of autocorrelated data, the known proposals for monitoring mean vectors include the use of MEWMA control charts for monitoring small shifts, as well as time series modelling and controlling the residuals of the adjusted model as well (Montgomery, 2001).

In the next section, we link these three control charts with the Mahalanobis distance and propose a linear transformation to optimise the classification accuracy of new observed vectors in terms of the ARL.

2.2. The noncentrality parameter

With respect to the control of mean vectors with no prior information on the direction of the shifts, it is known that the performance of the Hotelling's T^2, MCUSUM, and MEWMA control charts depends only on the noncentrality parameter (d). If the shift is large, Shewhart-type charts, such as Hotelling's T^2, are known to have a good performance in terms of ARL. In contrast, if the shift is small, *non*-Shewhart charts, such as MCUSUM or MEWMA charts, have been proven to be more effective in terms of rapid change detection. Some authors have also developed efficient modifications of these methods, such as the dEWMA method proposed by Alkahtani and Schaffer (2012). For multistage processes, when there is prior information on the plausible directions of the shift, the directional MEWMA (DMEWMA) control chart (Zou & Tsung, 2008) that incorporate the generalised likelihood ratio test is a successful approach.

The squared noncentrality parameter is a distance with known distribution and properties (Fukunaga, 1990). The distance between the mean vectors of two populations is also known as the Mahalanobis (1936) distance. For the case of the normal distribution, the density is a Gamma distribution with mean n and variance $2n$, where n is the dimension of the data. Approaches based on this measure usually exhibit a good performance for very small distances and when the dimensionality is low (no higher than one or two) but deteriorate with increasing dimensions, as the separability between the two classes increase.

The Mahalanobis distance can be directly linked with other dissimilarity measures used in classification problems (Tou & Gonzalez, 1974; Therrien, 1989) in more general formulations, namely, the Bhattacharyya (1943) and Chernoff (1952) distances. These distances are perfect general measures that can be used to monitor both mean vectors and covariance matrices, as it is possible to optimise each factor separately. Fukunaga (1990) states that for problems based on the noncentrality parameter, the total error of classification, known as the Bayes error, increases significantly with the number of dimensions; the loss of this valuable information is the price we must pay for not knowing the alternative class.

It is known that the performance of Hotelling's T^2, MEWMA and MCUSUM control charts for monitoring mean vectors depends only on the noncentrality parameter (d) and does not depend on the direction of the shift. The squared noncentrality value is exactly the Hotelling's distance of individual vectors to the process centre, given by

$$d^2 = (\mathbf{x}_i - \mathbf{M}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_i - \mathbf{M}_0), \quad (1)$$

where $\mathbf{x}_i$, $\mathbf{M}_0$ and Σ_0 are the observation vector, mean vector and covariance matrix of the reference process, respectively. The decision rule gives an out-of-control signal as soon as $d^2 > h$, where h is a specified threshold for achieving the desired false alarm rate, usually defined in terms of the in-control average run length (ARL_0).

Moreover, in addition to providing the standardised distance of the current observation to the reference distribution, the Mahalanobis distance (d_M^2) is a measure of dissimilarity between two populations mean vectors, given by

$$d_M^2 = (\mathbf{M}_2 - \mathbf{M}_1)' \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} (\mathbf{M}_2 - \mathbf{M}_1), \quad (2)$$

where $\mathbf{M}_i$, Σ_i, $i = 1, 2$, are the mean vectors and the covariance matrices of the two processes. Note that in SPC applications, both MEWMA and MCUSUM control charts make use of the Mahalanobis distance intrinsically. The main difference between these methods is the estimation of the current mean vector. In a recent study, Bersemis et al. (2007) discuss many technical aspects related to the Phase I and II applications of the multivariate control charts for mean vectors based on the noncentrality parameter.

Rao (1947) explained that the Mahalanobis distance is an explicit function of the proportion of overlapping individuals in the two populations. Rao (1949) also commented that Bhattacharyya had developed a perfectly general measure defined using the distance between two populations based on a Riemannian geometry metric, with the angular distance between points representing the populations in a unit sphere. Later, Atkinson and Mitchell (1981) studied Rao's paper in detail, providing the distances for a well-known family of distributions. Recently, Michelli and Noakes (2005) determined Riemannian distances for a large class of multivariate probability densities with the same mean by reducing the distances to quadratures and, in some cases, yielding closed-form expressions.

As it is possible to utilise the Mahalanobis distance for on-line monitoring by using the information available sequentially, a method for optimising it and extracting the optimal variable to discriminate between two populations with respect to the difference of mean vectors is presented. Thus, the transformation criteria adopted in the present work for extracting a more informative variable and application in statistical on-line process control is discussed in the following section.

2.3. Maximisation criteria

It is known that the number of available observational samples may be insufficient for estimating the scope of the parameters in the original space, given a moderate increase in the dimension of the data. Thus, the mapping of the original data ($\mathbf{Y}$) into another space ($\mathbf{X}$) in which most of the important information is optimally concentrated in a smaller number of variables may produce better results by reducing the total probability of misclassification without incurring any significant loss of information.

Among the two types of criteria that are frequently used in practice, one is based on a family of functions of scatter matrices, which are conceptually simple and for which systematic algorithms for choosing the best features are available. This type of criteria measures the class separability between classes, but does not provide information on the Bayes error directly. The second type of criteria provides upper bounds of the Bayes error, such as the Chernoff (1952) and Bhattacharyya (1943) distances.

When the objective is to effectively preserve class separability, the criteria for choosing the appropriate features are essentially independent of coordinate systems. Thus, these criteria are completely different from those utilised for signal representation as applied in PCA, which are based only on the correlation or covariance matrix.

Generally, in linear feature extraction for signal representation as utilised in PCA, the transformations are limited only to orthonormal spaces because the shape of the distribution has to be preserved. By changing the transformation criteria, the linear features do not need to be orthonormal, leading to more effective boundaries between the different classes involved for signal classification purposes.

In discriminant analysis, the main criterion frequently used in practice is based on a function of scatter matrices. Within-class, between-class and a mixture of scatter matrices are the three types of matrices used to formulate criteria for class separability, yielding a simple and systematic feature extraction algorithm. A within-scatter matrix (S_w) shows the scatter samples around their mean vectors, given by $S_w = \sum_{i=1}^{L} P_i \Sigma_i$, where P_i is the priori probability of each of the L classes. A between-scatter matrix (S_b) is the scatter of the expected vectors around the mixture means, $S_b = \sum_{i=1}^{L} P_i (\mathbf{M}_i - \mathbf{M}_0)(\mathbf{M}_i - \mathbf{M}_0)'$. The mixture scatter matrix is the covariance of all samples regardless of their class assignments, defined as $S_w + S_b$.

As an effective class separability criterion, these matrices need to be converted to a number that should increase when S_b increases or S_w decreases. A typical criterion for preserving class separability is $J = tr\left(S_w^{-1} S_b\right)$. Fukunaga (1990) shows that this criterion is invariant under any non-singular transformation, as the two scatter matrices can be simultaneously diagonalised using linear transformations, not altering the value of J regardless of the selection of the coordinate system. Thus, performance comparison can be conducted without any lack of generality by simulating only non-correlated processes.

For the out-of-control process, or ω_1 class, in addition to the estimation of the current mean vector, it is also possible to assume that the variance–covariance structure of the process is unknown, i.e., $\mathbf{Y}|\omega_1 \sim \text{MVN}(\mathbf{M}_1, \Sigma_1)$. Thus, the decision boundaries are adjusted according to the averaged covariance matrix instead of using only the prior information for Σ_0. The estimation of the covariance matrix generally requires a significantly larger number of samples than is necessary to estimate only the mean vector. Appropriately smoothing the covariance matrix at each instant for this case is a subject not described in detail in this study. As the main focus of this study is the on-line monitoring of mean vectors, the hypothesis assumed here is the equality of the covariance matrix in the processes, and only the mean vector is updated for each new observation vector.

In general form, a linear coordinate system transformation can be represented by

$$\mathbf{X} = \mathbf{A}'\mathbf{Y}, \quad (3)$$

where $\mathbf{A}$ is the transformation matrix that maps the original $\mathbf{Y}$ space into the transformed $\mathbf{X}$ space. However, contrary to the case of PCA utilised for signal representation, the column vectors of $\mathbf{A}$ do not need to be orthonormal when the purpose is classification. Thus, finding $\mathbf{A}$ that optimises J in the $\mathbf{Y}$ space is conducted as follows.

When $\Sigma_1 = \Sigma_2$, the Mahalanobis distance defined in Eq. (2) becomes the noncentrality parameter comparing two population mean vectors,

$$d^2 = (\mathbf{M}_2 - \mathbf{M}_1)' \Sigma^{-1} (\mathbf{M}_2 - \mathbf{M}_1). \quad (4)$$

Note that the product in the right-hand side of Eq. (4) has dimension 1×1 and therefore becomes

$$d^2 = tr\left((\mathbf{M}_2 - \mathbf{M}_1)' \Sigma^{-1} (\mathbf{M}_2 - \mathbf{M}_1)\right). \quad (5)$$

Given two matrices $\mathbf{A}$ and $\mathbf{B}$, and using $tr(AB) = tr(BA)$, then

$$d^2 = tr\left(\Sigma^{-1} (\mathbf{M}_2 - \mathbf{M}_1)(\mathbf{M}_2 - \mathbf{M}_1)'\right). \quad (6)$$

In the right-hand side of Eq. (6), observe that $(\mathbf{M}_2-\mathbf{M}_1)(\mathbf{M}_2-\mathbf{M}_1)'$ is a matrix with dimensions $n\times n$ generated by the product of a vector by itself; therefore, its rank equals 1. In this case, $(\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1)(\mathbf{M}_2-\mathbf{M}_1)')$ is also a $n\times n$ matrix with a rank equal to 1 and has only one eigenvalue other than zero, i.e.,

$$\lambda_1\neq 0,\ \lambda_2=\lambda_3=\ldots=\lambda_n=0. \tag{7}$$

This indicates that for two-class problems, only one feature is needed. It can be shown that the trace of a matrix is equal to the sum of its eigenvalues; thus, Eq. (6) becomes

$$d^2=\lambda_1, \tag{8}$$

with

$$\lambda_1=(\mathbf{M}_2-\mathbf{M}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1). \tag{9}$$

Therefore, in the original variable space $\mathbf{Y}$, d^2 is aligned with the eigenvector $\mathbf{e}_1$ of $\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1)(\mathbf{M}_2-\mathbf{M}_1)'$ and associated with the first eigenvalue λ_1. Observe that because λ_1 is exactly the noncentrality value, this fact guarantee the direct comparison of the proposed control charts with the selected ones.

Given a vector $\mathbf{Y}$ with dimensions $n\times 1$, the extracted feature ($\mathbf{X}$) can carry all the information of class separability due to the difference in mean vectors and is, therefore, a sufficient statistic, while the other features are redundant. In this transformed space, $\mathbf{X}$ can be obtained by projecting $\mathbf{Y}$ on e_1 according to Eq. (3)

$$\mathbf{X}=\mathbf{e}_1'\mathbf{Y}. \tag{10}$$

The direction of $\mathbf{e}_1$ can be obtained from the general problem of finding eigenvalues and eigenvectors in a generic matrix $\mathbf{M}$ such that

$$\mathbf{M}\mathbf{e}_1=\mathbf{e}_1\lambda_1. \tag{11}$$

In this case,

$$\mathbf{M}=\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1)(\mathbf{M}_2-\mathbf{M}_1)'. \tag{12}$$

Then,

$$\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1)(\mathbf{M}_2-\mathbf{M}_1)'\mathbf{e}_1=\mathbf{e}_1(\mathbf{M}_2-\mathbf{M}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1), \tag{13}$$

and therefore,

$$\mathbf{e}_1=\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1). \tag{14}$$

Thus, the extracted feature containing all the information with respect to the separation between the mean vectors of two multivariate normal processes is the transformed space given by

$$\mathbf{X}=\left(\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2-\mathbf{M}_1)\right)'\mathbf{Y}. \tag{15}$$

Then, to project two distributions onto one vector, this vector must be perpendicular to the optimum hyperplane between the two distributions. The above argument suggests that by projecting two distributions into the vector oriented along $\mathbf{e}_1$, all the classification information is preserved as long as the class separability is effectively measured using the selected criterion.

2.4. The Lin-MEWMA control chart

The procedure described above is defined for the classification of two classes that are given or estimated a priori. Thus, each new observation is classified as belonging to one of the two known classes, according to the decision criteria previously established. As the parameters or estimates of both classes are given, most commonly, the maximum likelihood rule is adopted for the classification procedure.

The difference between the MEWMA scheme and the MEWMA control chart must be considered first. For the application of this method to the on-line process control of mean vectors, it is first necessary to estimate the current mean vector. For this purpose, we tested sliding window schemes and the MEWMA scheme. In the present work, we differentiate the MEWMA scheme from the MEWMA chart in that the former is a method for on-line mean vector estimation and that the latter is a Mahalanobis distance classifier, applying a MEWMA scheme.

In fact, many authors (Hwarng and Hubele, 1993a, 1993b; Guh & Shiue, 2005) have suggested a sliding window approach as the essential tool for on-line pattern identification. However, additionally to the problem of choosing the appropriate window size, there are issues about addressing unnatural patterns, i.e., when a misalignment of the pattern in time may occur. Also, the identified pattern could be different from the training pattern (Guh & Shiue, 2008; Hachicha & Ghorbel, 2012), and dynamic window sizes may be more appropriate. Our comparison of the sliding window schemes are not presented in this work because the results have shown that the MEWMA scheme is more simple and efficient than the tested sliding window approaches, and it is the scheme adopted for the current purpose. The main reason for not choosing sliding windows was that it tends to delay change detection. This is known as the inertial effect of non-Shewhart control charts and are still augmented by the use of sliding windows. Finally, the MEWMA scheme has the advantage of allowing a direct comparison of the proposed chart with the quadratic MEWMA control chart by utilising the same smoothing values.

The actual mean vector $\mathbf{M}_t$ is estimated with the inclusion of the most recent observed vector, $\mathbf{Y}_t$, such that

$$\mathbf{M}_t=(1-\lambda)\mathbf{M}_{t-1}+\lambda\mathbf{Y}_t,\quad t=1,2,3,\ldots \tag{16}$$

where $\mathbf{M}_0=\mathbf{0}$ and $0>\lambda\geqslant 1$. Although different weights could be set for each variable in the vector, observe that maintaining the weights of all variables as fixed renders the proposed control charts directionally independent. If different weights are given for the different variables, the control charts becomes directionally dependent and their performances cannot be compared with those of the selected non-directional control charts.

Performance efficiency is measured by comparing the ARL of the competing control charts after the calibration process. As the proposed control charts for monitoring mean vectors are based on the linear X statistic and utilises the MEWMA and CUSUM schemes sequentially, they are called the Linear MEWMA (Lin-MEWMA) and Cumulative Sum of X (CUSUM-Lin) control charts.

The Lin-MEWMA control chart involves simply monitoring the one-dimensional variable by selecting symmetrical in-control limits that result in a pre-specified ARL_0. The observations generated from the in-control process are referred to as class ω_0 such that $\mathbf{Y}|\omega_0\sim \mathrm{MVN}(\mathbf{M}_0,\boldsymbol{\Sigma}_0)$. The transformed variable triggers an out-of-control signal as soon as

$$\mathbf{X}_t=\left(\boldsymbol{\Sigma}_0^{-1}(\mathbf{M}_t-\mathbf{M}_0)\right)'\mathbf{Y}_t\geqslant\pm h_1. \tag{17}$$

Here, $\mathbf{M}_t$ is the current mean vector, and h_1 is chosen to achieve a pre-specified ARL_0.

2.5. The CUSUM-Lin control chart

For investigating very small shifts in the mean value of a one-dimensional normal variable, the CUSUM and EWMA control charts are recognised as efficient tools for quality improvement. The performance of these methods is often compatible and depend only on the individual parameters k and λ, respectively, for the CUSUM and EWMA charts. The method for performing a CUSUM procedure over a one-dimensional statistic for multivariate processes was first proposed by Crosier (1988), called the CUSUM of T control chart. In fact, applying a CUSUM and EWMA schemes over the quadratic statistic was attempted also in many

experiments by the authors but the results presented performance equal or worse that of the MCUSUM and MEWMA charts. Although the CUSUM of T procedure was less effective than the vectorial CUSUM (MCUSUM) chart, the CUSUM of X (CUSUM-Lin) shows great performance in detecting very small shifts in the mean vector.

Beyond its simple rule for the choice of the k parameter, the standardised CUSUM procedure has two main advantages, as shown by Montgomery (2001). First, many CUSUM charts can now accommodate the same values of k and h, and the choice of these parameters does not depend on the standard deviation of the variables. Second, a standardised CUSUM naturally controls the process variability, as it becomes more sensitive to the process dispersion around the mean vector.

Thus, for the CUSUM-Lin control chart, first compute

$$z_t = \frac{X_t - \mu_X}{\sigma_X}, \tag{18}$$

the standardised value of X_t. As a closed form for μ_X and σ_X is not provided yet, the estimation of those parameters was carried out by means of numerical simulation. A closed-form expression for the first and second order moments of the statistic remains a topic for future research. In order to overcome this problem, one may apply the non-standardised CUSUM Montgomery (2001) and expect the same performance. As the presented standardised version is a simple scale transformation, it does not affect the overall performance of the technique.

Then, the positive and negative standardised CUSUMs are defined as

$$C_t^+ = \max\left[0, z_t - k + C_{t-1}^+\right] \tag{19}$$

$$C_t^- = \max\left[0, -k - z_t + C_{t-1}^-\right], \tag{20}$$

where the starting values are $C_t^- = C_t^+ = 0$. If either C_t^+ or C_t^- exceed the decision threshold h chosen to achieve a specific ARL_0, the process is considered to be out-of-control.

The condition $k = 0.5$ was selected for both CUSUM-Lin and MCUSUM charts because it is expected to yield the best performance for a shift $d = 1$ by utilising the rule $k = d/2$ (Montgomery, 2001). As the CUSUM-Lin chart monitors the proposed variable after smoothing by the MEWMA scheme, the performance of this method can be viewed as a double-filtering procedure, as it is affected not only by the k parameter but also by the λ smoothing factor employed first by the MEWMA scheme in the linear transformation.

3. Results and discussion

This section describes the calibration of the control chart and the experiments for performance comparison. The differences between the proposed control charts and traditional multivariate approaches of Hotellings T^2, MEWMA and MCUSUM charts are analysed for mean vector shifts in the 0–5 range of the noncentrality parameter. The experimental scenarios described below are established to illustrate the sensitivity of the λ and k parameters. For the MEWMA scheme, as λ in the range 0.1 to 0.4 leads to effective detection of shifts in the mean vector with magnitude $d \approx 1$ or larger (Lowry et al., 1992), values of λ in this range are applied for the proposed control charts. The k parameter for the MCUSUM chart was set to 0.5 (Crosier, 1988), indicating that the chart is calibrated to detect differences in the noncentrality parameter when it is equal to 1. Although the CUSUM-Lin chart is primarily filtered using the λ factor, the value $k = 0.5$ is kept fixed for the sake of comparison with the MCUSUM chart.

3.1. Control chart calibration procedure

In this work, the control chart calibration is performed by specifying a sequence of thresholds for each control chart and recording the resulting ARL_0. The thresholds of all competing control charts were adjusted to obtain $ARL_0 = 200$. The control chart calibration procedure is performed in two steps. The first step utilises a linear regression model of d^2 as a function of the $ln(ARL_0)$. All the regression models presented a goodness of fit of at least 0.95. The second step in the calibration procedure iteratively refines the desired threshold using linear interpolation. A sequence of 20 thresholds with 5000 Monte Carlo simulations were performed to obtain the first ARL_0 estimates. For the final control chart comparisons, 50,000 Monte Carlo simulations were performed.

Mahmoud and Maravelakis (2010, 2011) show that utilising estimated parameters instead of known parameters delays the change detection of the multivariate control charts. Without loss of generality, for performance comparison, all control charts were calibrated using Phase I sample estimates of limited sizes, to amplify the performance differences. Therefore, to illustrate the control chart efficiency in different dimensions, the n-dimensional multivariate processes are simulated using $n = 2$ and 4.

For the bivariate case, the in-control process parameters were estimated using a Phase I stage with $m = 25$ observations. In the four-dimensional case, the mean vector and covariance matrix of the in-control process were estimated using a Phase I stage with $m = 50$ observations. Thus, the estimated in-control limits must address a bigger variation than expected for the case of known parameters. While the in-control limit utilising known parameters for the MCUSUM chart in Crosier's (Crosier, 1988) paper is $h = 5.5$, in this work, with unknown parameters, we found $h = 6.64$ using our new methodology. The same was observed using the MEWMA chart; the in-control threshold in Lowry et al. (1992) paper was $h = 8.79$, and we obtained $h = 11.43$ by utilising unknown parameters.

3.2. A numerical example

A numerical example with a simulated dataset to illustrate the MCUSUM and MEWMA charts in both respective original papers is shown in Table 1. The first five observations of the data are generated from a bivariate normal in-control process with mean vector (0, 0), unitary variances and a correlation of 0.5. The last five observations are generated from an out-of-control process with the mean vector shifted to $\mathbf{M}_1 = (1, 2)$. The shift correspond to a noncentrality value $d = 2$, which is considered as a median distance value.

First of all, notice that both Lin-MEWMA and CUSUM-Lin charts were calibrated for a process with unknown parameters (Phase I

Table 1
A numerical example of bivariate quality-control schemes ($\lambda = 0.1$).

t	y_1	y_2	e_{11}	e_{12}	X_t	C_i^+	C_i^-
1	−1.19	0.59	−0.20	0.16	0.33	0.00	0.00
2	0.12	0.90	−0.22	0.25	0.20	0.00	0.16
3	−1.69	0.40	−0.45	0.39	0.92	0.43	0.00
4	0.30	0.46	−0.40	0.40	0.06	0.00	0.47
5	0.89	−0.75	−0.19	0.20	−0.32	0.00	1.78
6	0.82	0.98	−0.13	0.25	0.15	0.00	2.06
7	−0.30	2.28	−0.31	0.55	1.35	1.39	0.00
8	0.63	1.75	−0.31	0.69	1.01	2.02	0.00
9	1.56	1.58	−0.17	0.73	0.88	2.35	0.00
10	1.46	3.05	−0.17	0.96	2.70*	6.72	0.00
h					±1.84	15.2	

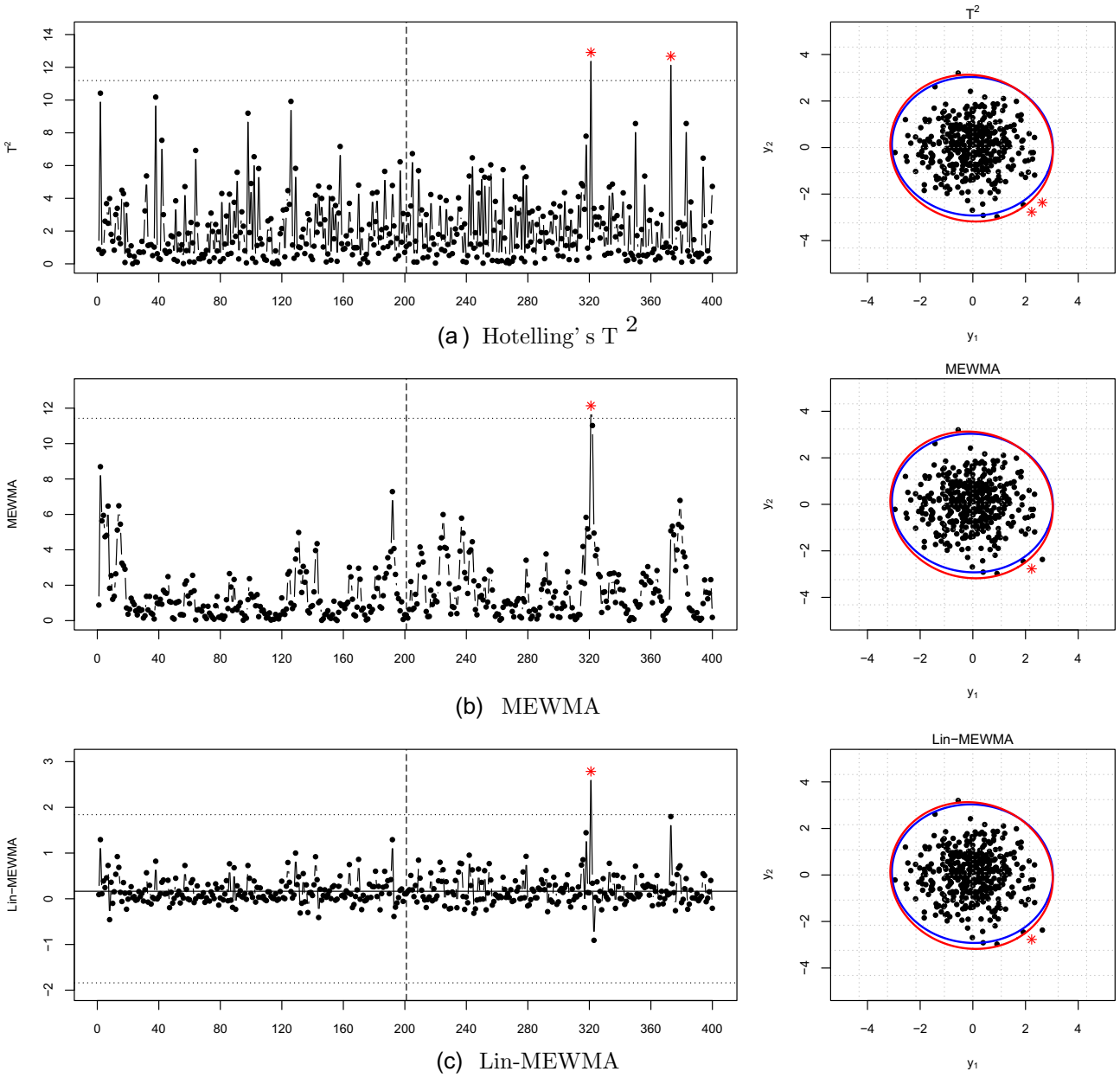

Fig. 1. Hotelling's T^2, MEWMA and Lin-MEWMA control chart patterns with scatter plots ($n = 2$, $\lambda = 0.1$, $d = 0$).

with $m = 25$). Then, in the same way that the MEWMA and MCUSUM charts have delayed detection when estimated parameters are employed to delimit the thresholds (Mahmoud & Maravelakis, 2010, 2011), a relative delay in change detection observed for both charts is expected for the proposed charts. In practice, as the shift size is unknown, the λ factor must be chosen a priori. Following the example of the original MEWMA's (Lowry et al., 1992) paper, λ is set to 0.1 for comparison purposes.

As seen in the next section, the same value for the smoothing factor λ affects the charts performance differently. Therefore, in the simulated example, the Lin-MEWMA control chart shows a performance compatible with that of the MEWMA chart, signalling the shift after the ninth observation. Additionally, the CUSUM-Lin chart with $\lambda = 0.1$ and $k = 0.5$ is sensitive for smaller shifts in the mean vector, when $d \leqslant 2$. In this way, and considering that the bounds are set for a process with unknown parameters, the simulated change was not detected by the CUSUM-Lin chart before the tenth observation.

3.3. Control chart pattern analysis

The illustrations that follow compare the common patterns of the control chart for mean vector shifts. Fig. 1 presents the simulation of an extended run length of 400 observations with no change in the process ($d = 0$) monitored using Hotelling's T^2 and the MEWMA and Lin-MEWMA control charts. Red dots are out-of-control observations; the horizontal line indicates the estimated threshold (Table 2), and the vertical dashed line in the middle of the control chart indicates a change point for the process.

As stated before, PCA finds directions that best represent the multivariate process. Therefore, the estimations of confidence ellipsoids (axis directions and sizes) were conducting using PCA. Fig. 1 show those estimated PCA boundaries, seen in the scatter plots on the right side of each respective control chart in blue and red. The blue circle is the estimated ellipsoid with a confidence level of 99.5% for the IC process, and the red circle is the estimated ellipsoid for the OC process.

Table 2
Control chart performance comparison ($n = 2$, $\lambda = 0.1$).

d	T^2	MCUSUM	MEWMA	Lin-MEWMA	CUSUM-Lin
0.0	200.7	200.3	200.9	200.5	200.1
0.5	123.7	57.3	54.8	58.9	37.2
1.0	51.5	12.7	13.3	14.5	11.0
1.5	20.9	6.9	7.1	7.3	6.5
2.0	8.8	4.8	5.0	4.6	4.6
2.5	4.4	3.8	3.9	3.3	3.6
3.0	2.5	3.1	3.2	2.5	2.9
3.5	1.7	2.6	2.8	2.0	2.5
4.0	1.3	2.3	2.5	1.6	2.2
4.5	1.1	2.1	2.2	1.4	1.9
5.0	1.1	1.9	2.0	1.2	1.7
h	11.19	6.64	11.43	1.84	15.19

Fig. 2 shows the MCUSUM and CUSUM-Lin charts for monitoring the simulated IC process. The first aspect to be observed when comparing all the control charts is that the Lin-MEWMA chart plots an statistic varying around a mean value. This behaviour of the proposed linear statistic allows for the successful implementation of the standardised CUSUM procedure. This is not possible in low-dimensional spaces because the quadratic statistic is very asymmetrical and does not satisfy the required assumptions of the CUSUM procedure. The experiments showed that the standardised CUSUM with the MEWMA statistic was equally or less effective than the MEWMA control chart itself. When the quadratic statistic is applied in the CUSUM chart, the transformation indicates very erratic behaviour, without good stability in the control chart signal.

From the simulated experiment of Fig. 3, it can be seen that a shift in the mean vector ($d = 2$) can be detected by all compared control charts. This change is not well perceived using the Hotelling's T^2 chart because too few observations are present in the external region of the IC process.

The MEWMA statistic does not cover the IC region in the same way as the Hotelling's T^2 and the Lin-MEWMA statistics. Instead of denoting observations situated outside the IC region, these observations are denoted as OC according to whether the current mean vector is considered OC. However, in Fig. 3c, the linear statistic monitored using the Lin-MEWMA chart displays OC observations outside of the linear projection in the direction of the shift. As $\mathbf{M}_1 = (2, 0)$, OC observations are clearly delimited by linear boundaries perpendicular to the mean vector's shift.

This characteristic aspect, the projected linear boundaries, of the Lin-MEWMA control chart is a very important result of the proposed statistic. Observe that any observation vectors situated near the IC process centre were tagged as OC. In real applications, not discarding non-defective units generated by an OC process may result in cost or time savings in stopping the process too early. Even if the process mean vector is actually OC, when all variables of the particular observation vector are IC, some applications may see no reason to discard a perfectly manufactured unit.

This type of selective criterion to trigger an OC signal was used before only in Shewhart-type methods such as the Hotelling's T^2 chart. As seen if Fig. 3, a non-Shewhart control chart, i.e., utilising

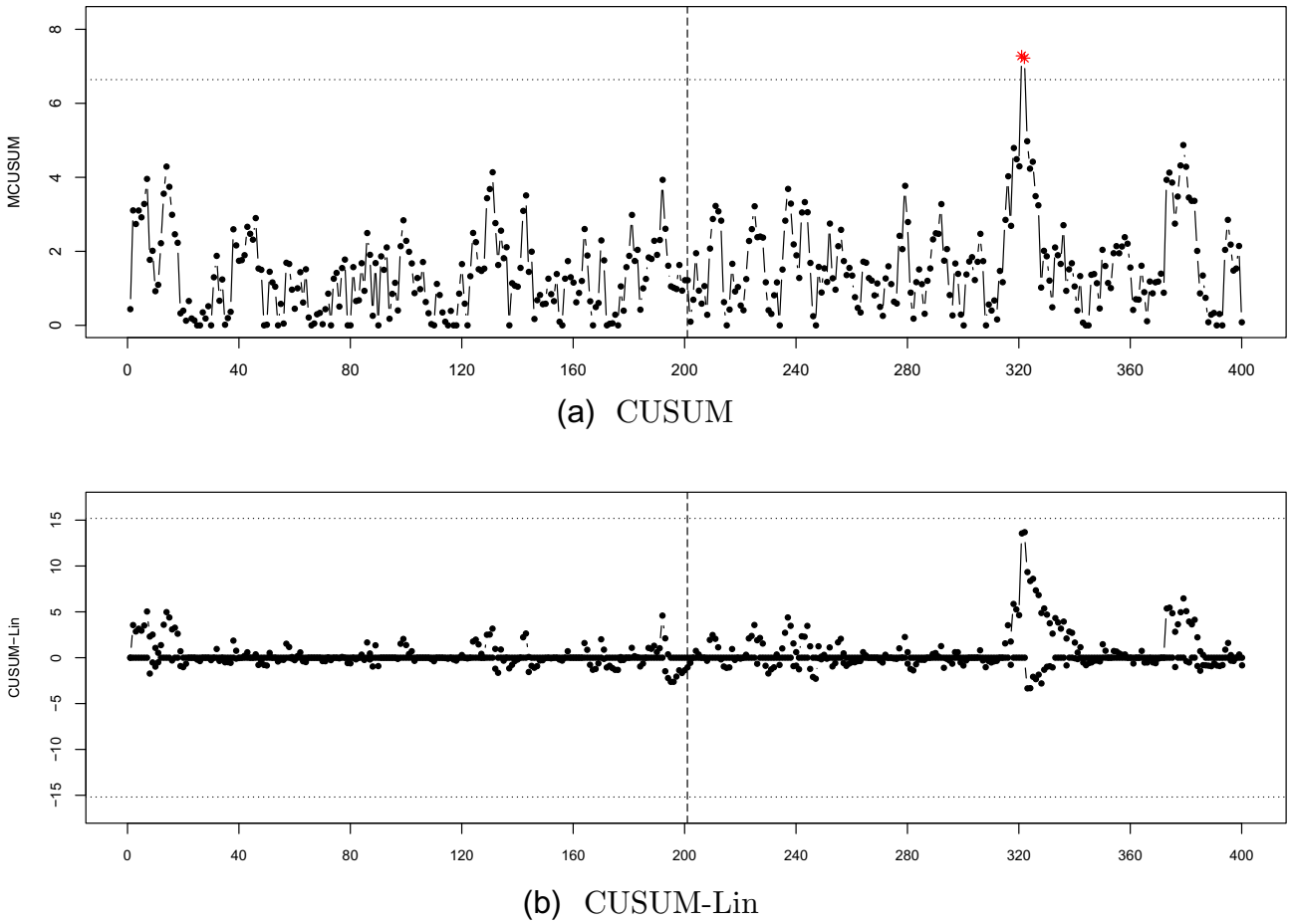

Fig. 2. MCUSUM and CUSUM-Lin control chart patterns ($n = 2$, $\lambda = 0.1$, $k = 0.5$, $d = 0$).

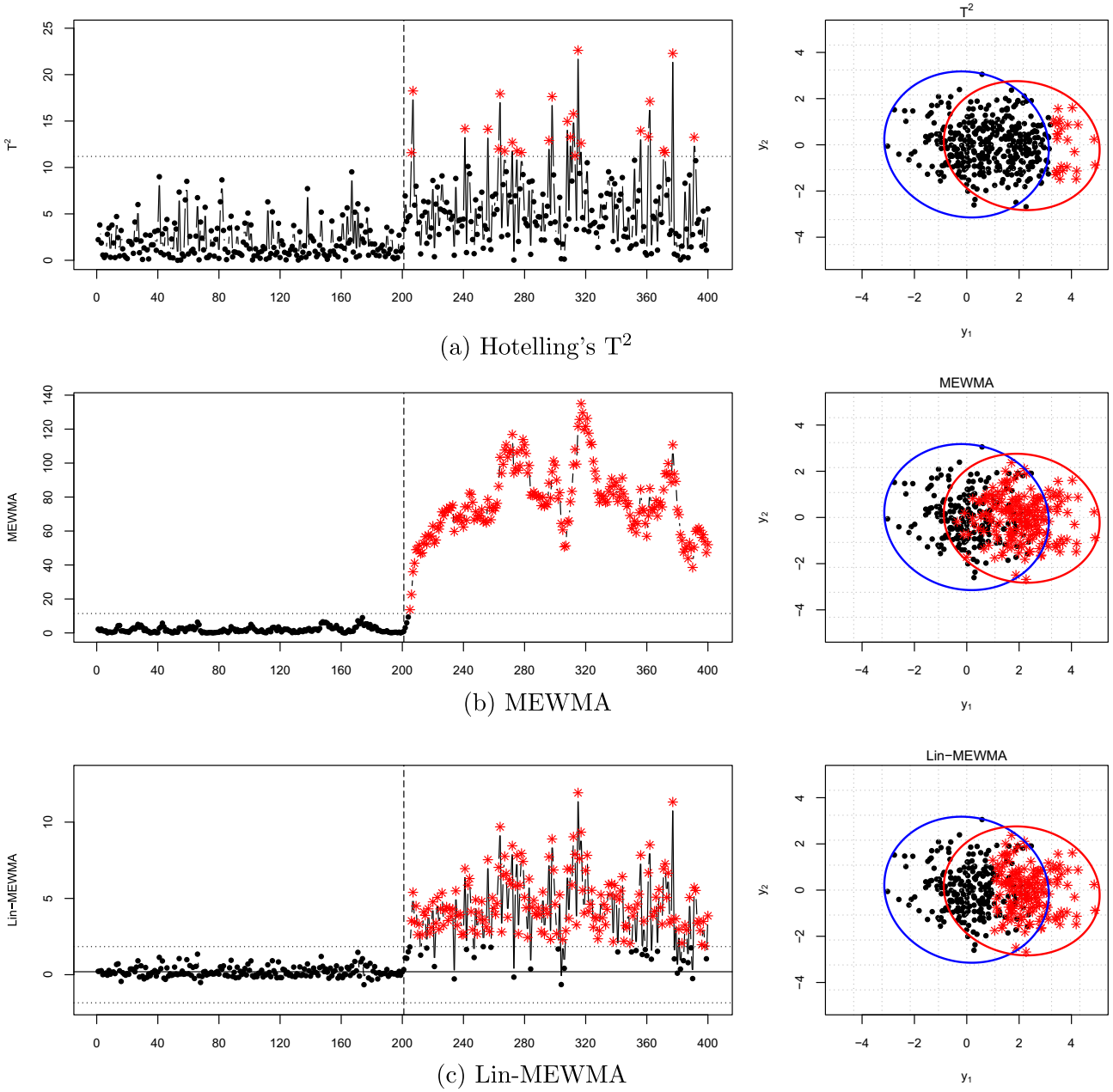

Fig. 3. Hotelling's T^2, MEWMA and Lin-MEWMA control chart patterns with scatter plots ($n = 2$, $\lambda = 0.1$, $d = 2$).

the past process information, can also selectively tag observations vectors even if the current mean vector is already OC.

While the Hotellings T^2 chart depicts too few observations as OC, the MEWMA chart discards observations that are clearly IC as OC observations. Here, the use of $\lambda = 0.1$ for the MEWMA scheme in the Lin-MEWMA chart seems to be a reasonable choice as a compromise for the two others, identifying and protecting units that are clearly IC situated inside the projected linear boundaries.

As seen in Fig. 3, with $\lambda = 0.1$, a linear boundary is projected at $d \cong 1$. Increasing λ to 0.4 in the MEWMA scheme extends the projected boundaries to larger distances ($d \cong 2.5$). Therefore, the control chart becomes less sensitive to smaller shifts in the mean vector and has a performance that is similar to that of the Hotelling's T^2 chart. This indicates that if the overlap between the processes is high and the smoothing factor tends to 1, more observations will be considered IC and the chart will be sensitive only to very large shifts in the mean vector.

Considering multivariate Gaussian distributions, higher values for the smoothing factor should be considered only when the process dimensionality is very high because the changes occurs generally at higher distances. In such cases, considering that many variables could change simultaneously, the observed distances will be naturally larger than what would be observed if only one variable has a shift.

Fig. 4 illustrates a simulated example with a reduced run length of 50 observations. In this case, a change occurs at position $t = 51$, and the CUSUM-Lin chart detects the change before the MCUSUM chart. As usual, for the CUSUM procedure, the CUSUM-Lin chart ensures a continuous change, triggering alarms after detecting a significant change in the mean vector. Standardising the linear statistic through a CUSUM procedure renders the statistic to be a

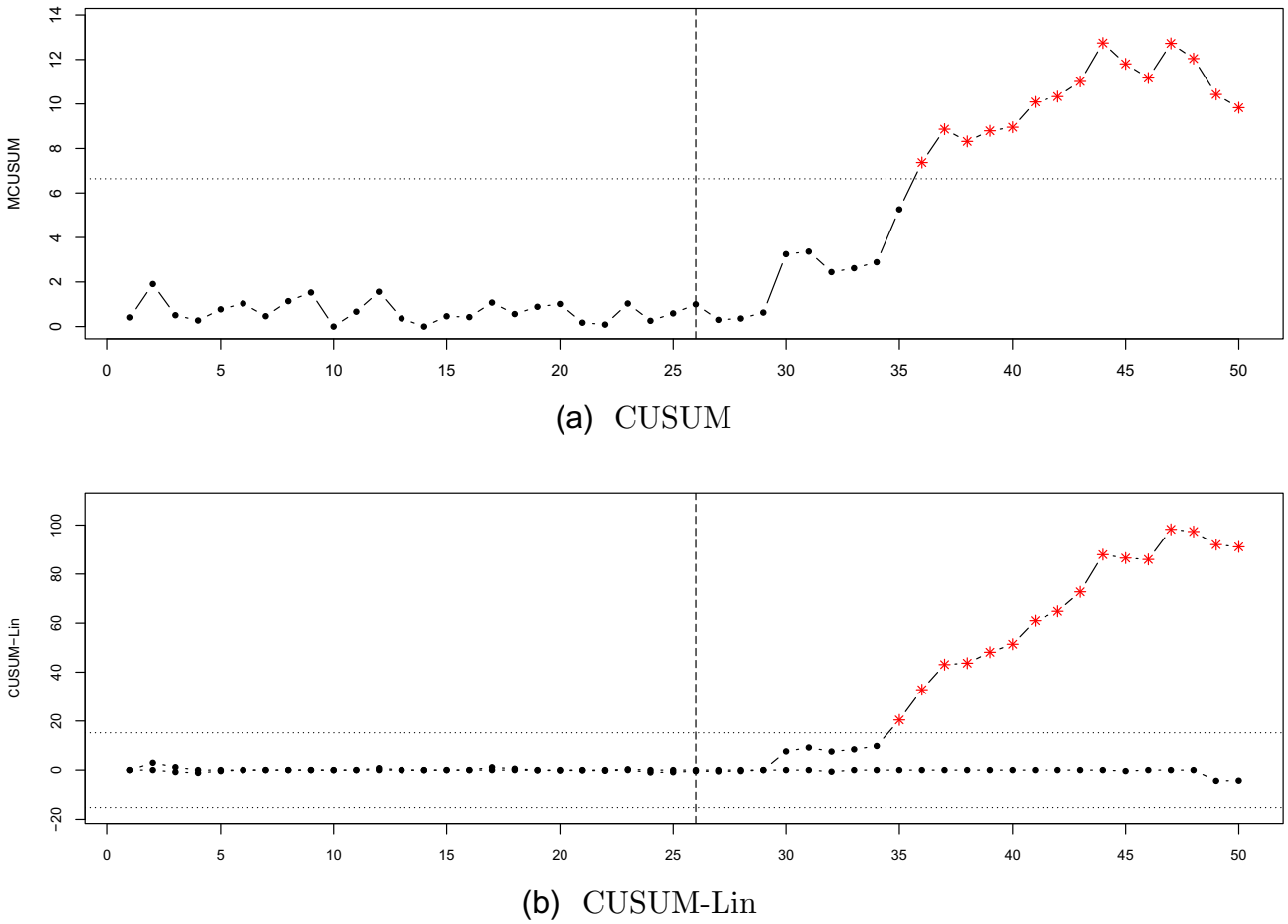

Fig. 4. MCUSUM and CUSUM-Lin control chart patterns ($n = 2$, $\lambda = 0.1$, $k = 0.5$, $d = 1$).

quadratic measure again, resulting in the loss of some important information on the vector location, as described above.

Although in- and out-of-control processes becomes more separable, there is no apparent reason for differences in performance by the concurrent control charts, unless the difference is due to a known inertial effect of the MEWMA control chart (Lowry et al., 1992) when compared to the Hotelling's T^2 chart. In the next section, the performances of the control charts are compared. The combined effect of double-filtering the process using the CUSUM-Lin chart after the smoothed linear transformation is also studied by the following ARL comparison.

3.4. Performance comparison

To effectively compare the performances of the control charts, the charts were calibrated to $ARL_0 = 200$. For the control chart calibration, the parameter estimates of $(\mathbf{M}_0, \mathbf{\Sigma}_0)$ were calculated using a Phase I stage with $m = 25$ observations of the IC process in the two-dimensional case ($n = 2$) and with $m = 50$ for $n = 4$. The effect of the unknown parameters creates a natural delay in change detection for all the control charts. Regarding the methodological differences, comparisons were made utilising the same λ factor in the MEWMA scheme for both the Lin-MEWMA and MEWMA charts.

The values of the weighting factor λ were fixed at 0.1 and 0.4 for both simulated dimensions. For the comparative analysis of the MEWMA-based charts, the values selected for λ represent low and intermediate smoothing levels. Additionally, when $\lambda = 1$, the proposed statistic displays an exactly equal performance for both the Hotellings T^2 and the MEWMA chart.

As the experiments demonstrate, using $k = 0.5$ for the MCUSUM chart results in its performance being similar to that of the MEWMA chart with $\lambda = 0.1$. Thus, this fixed k value for the MCUSUM chart and the Hotelling's T^2 chart for individual observation vectors acts as a benchmark for relatively small and large shifts, respectively, in the mean vector.

Notice that very small shifts in the mean vector are usually considered to be situated in the acceptance region. Although the ARL_0 is fixed to 200, it may be desirable to detect the out-of-control signal (ARL_1) in less than 20 observations in the maximum case. The following comparison is intended to understand where and how the proposed methodology is more efficient than the alternative approaches. In Table 2, for example, shifts bigger than $d > 0.5$ are detected using all the non-Shewhart methods in less than 15 observations.

The experiments of Table 2 were designed to detect large shifts, and the best performance for the Hotelling's T^2 chart is achieved when $d \geqslant 3$. Both the MEWMA and MCUSUM charts present the same performance, given the parameter settings. Comparing only the *non*-Shewhart methods, the Lin-MEWMA chart presents the best performance for $d \geqslant 2$ and behaves most similarly to the MEWMA chart when $d = 1.5$.

By transforming the ARLs in a logarithmic scale it is possible to obtain a good ARL differentiation between small and large shifts with a decrease in the scale difference. This behaviour indicates that all control charts present a compromise between small and large shifts, as shown by the log(ARL) in Fig. 5. For both the MCU-

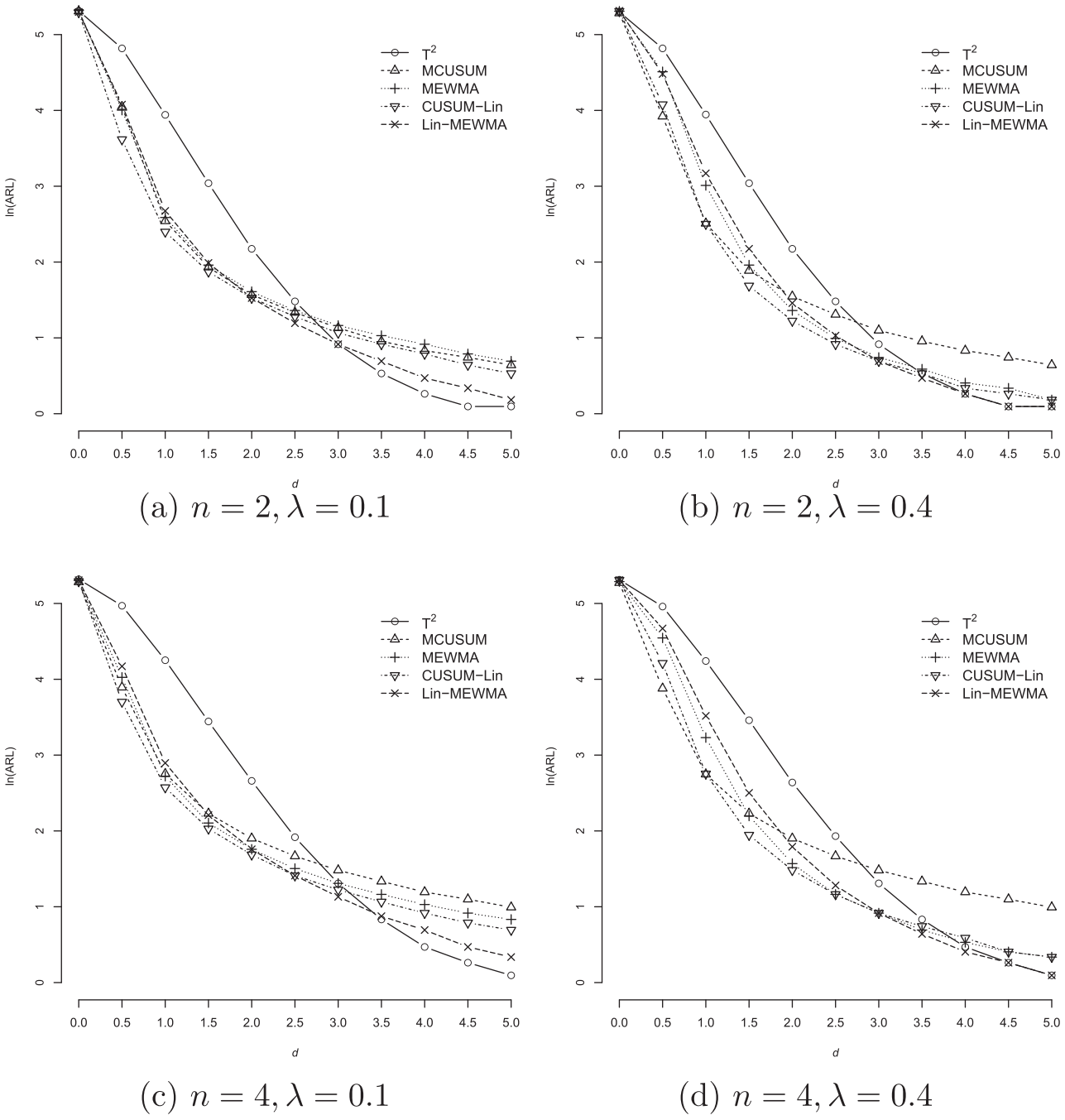

Fig. 5. Control chart ARL comparison on the logarithmic scale.

SUM and MEWMA charts, the ARL curves in Fig. 5a show that to be effective for very small shifts, the control chart must compensate with a loss of performance at higher shifts. By analysing the log(ARL) curves, the Lin-MEWMA chart is the most optimal approach in terms the overall performance for all shifts in the 0–5 range.

While the CUSUM-Lin chart performs similar to the Lin-MEWMA chart for $d = 2$, its performance is the best when $d < 2$. As stated before, the advantage of increasing the alarm rate for $d = 0.5$ can be viewed as a pseudo true alarm rate, and its use may be controversial. Therefore, the advantages of applying this chart may depend on the particular application. The results indicate that for large shifts, the CUSUM-Lin tends to suffer problems related to inertia, but with less impact than the alternative quadratic MEWMA chart. Based on this fact, comparing only the non-Shewhart methods, it can be clearly seen in Fig. 5a that the Lin-MEWMA and even the CUSUM-Lin charts are less affected by the problem of inertia than the MCUSUM and MEWMA charts.

The observed average difference in the proposed methods for triggering an OC signal can represent great savings of cost in real applications. This means that too many alarms for a very small shift in the process can make the process stop too early, leading to the fix of small problems. The utilisation of symmetrical in-control limits for the Lin-MEWMA control chart has been shown to have good properties, as it highlights out-of-control observations that are situated at both sides of the out-of-control process with respect to the direction of the shift.

The projected linear boundaries do not allow for alarms for perfect observations, which are situated near the IC process centre. Although not perfectly centred, if the OC process is not perfectly

Table 3
Control chart performance comparison ($n = 2$, $\lambda = 0.4$).

d	T^2	MCUSUM	MEWMA	Lin-MEWMA	CUSUM-Lin
0.0	200.2	200.5	200.8	200.3	200.7
0.5	123.7	50.5	90.9	88.3	59.0
1.0	51.7	12.3	20.3	23.8	12.2
1.5	20.9	6.6	7.1	8.8	5.4
2.0	8.8	4.7	3.9	4.3	3.4
2.5	4.4	3.7	2.7	2.8	2.5
3.0	2.5	3.0	2.1	2.0	2.0
3.5	1.7	2.6	1.8	1.6	1.7
4.0	1.3	2.3	1.5	1.3	1.4
4.5	1.1	2.1	1.4	1.1	1.3
5.0	1.1	1.9	1.2	1.1	1.2
h	11.19	6.64	11.43	4.96	7.42

Table 4
Control chart performance comparison ($n = 4$, $\lambda = 0.1$).

d	T^2	MCUSUM	MEWMA	Lin-MEWMA	CUSUM-Lin
0.0	200.2	200.1	201.0	201.6	199.5
0.5	144.0	49.1	56.0	64.8	40.5
1.0	70.2	15.7	15.2	18.1	13.1
1.5	31.3	9.3	8.2	9.1	7.6
2.0	14.3	6.7	5.8	5.8	5.4
2.5	6.8	5.3	4.5	4.1	4.1
3.0	3.7	4.4	3.7	3.1	3.4
3.5	2.3	3.8	3.2	2.4	2.9
4.0	1.6	3.3	2.8	2.0	2.5
4.5	1.3	3.0	2.5	1.6	2.2
5.0	1.1	2.7	2.3	1.4	2.0
h	17.02	10.88	17.09	2.65	12.08

Table 5
Control chart performance comparison ($n = 4$, $\lambda = 0.4$).

d	T^2	MCUSUM	MEWMA	Lin-MEWMA	CUSUM-Lin
0.0	201.1	199.8	199.6	200.2	200.9
0.5	142.5	48.4	94.2	106.7	67.5
1.0	69.5	15.6	25.3	33.7	15.7
1.5	31.8	9.3	9.0	12.2	7.0
2.0	14.0	6.7	4.8	6.0	4.4
2.5	6.9	5.3	3.2	3.6	3.2
3.0	3.7	4.4	2.5	2.5	2.5
3.5	2.3	3.8	2.0	1.9	2.1
4.0	1.6	3.3	1.7	1.5	1.8
4.5	1.3	3.0	1.5	1.3	1.5
5.0	1.1	2.7	1.4	1.1	1.4
h	17.0	10.9	17.1	7.4	7.5

Table 6
Comparison of the performance of the CUSUM-Lin and Lin-MEWMA control charts ($n = 2$).

d	CUSUM-Lin				Lin-MEWMA			
	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$
0.0	199.8	199.9	200.1	200.2	200.0	199.7	200.3	200.3
0.5	37.2	41.6	51.8	59.0	58.9	65.2	80.0	88.3
1.0	11.0	10.4	11.2	12.2	14.5	16.1	20.4	23.8
1.5	6.5	5.5	5.4	5.4	7.3	6.9	7.7	8.8
2.0	4.6	3.8	3.6	3.4	4.6	4.1	4.1	4.3
2.5	3.6	2.9	2.7	2.5	3.3	2.8	2.8	2.8
3.0	2.9	2.3	2.1	2.0	2.5	2.1	2.0	2.0
3.5	2.5	2.0	1.8	1.7	2.0	1.7	1.6	1.6
4.0	2.2	1.7	1.5	1.4	1.6	1.4	1.3	1.3
4.5	1.9	1.5	1.4	1.3	1.4	1.2	1.2	1.1
5.0	1.7	1.3	1.2	1.2	1.2	1.1	1.1	1.1

Table 7
Comparison of the performance of the CUSUM-Lin and Lin-MEWMA control charts ($n = 4$).

d	CUSUM-Lin				Lin-MEWMA			
	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$
0.0	199.5	200.5	200.3	201.1	201.6	201.2	200.1	200.4
0.5	40.5	50.3	58.1	67.5	64.8	80.2	96.9	106.7
1.0	13.1	13.0	14.0	15.7	18.1	22.1	27.5	33.7
1.5	7.6	6.9	6.7	7.0	9.1	9.1	10.4	12.2
2.0	5.4	4.6	4.4	4.4	5.8	5.3	5.5	6.0
2.5	4.1	3.5	3.3	3.2	4.1	3.6	3.5	3.6
3.0	3.4	2.8	2.6	2.5	3.1	2.6	2.5	2.5
3.5	2.9	2.4	2.2	2.1	2.4	2.1	1.9	1.9
4.0	2.5	2.1	1.9	1.8	2.0	1.7	1.6	1.5
4.5	2.2	1.8	1.6	1.5	1.6	1.4	1.3	1.3
5.0	2.0	1.6	1.4	1.4	1.4	1.2	1.2	1.1

linearly separable, i.e., if there is an overlap between the IC and OC distributions, the process does not produce defective units at all and may possibly keep running. When the mean vector is OC for intermediate shifts, perfect units are saved from being discarded or inspected, and real defective units are tagged OC sooner than with the competing *non*-Shewhart methods.

To avoid the problem of inertia, the results in Table 3 for $\lambda = 0.4$ indicate that the Lin-MEWMA control chart has a performance superior to that of the MEWMA chart for $d \geqslant 3$. For this selected parameter, the Lin-MEWMA chart does not show an inferior performance when compared to the Hotelling's T^2 chart for large shifts, which is a significant result because the proposed chart is as effective or more so than the MEWMA chart for small and intermediate shifts in the mean vector. As the k parameter for the MCUSUM chart is fixed, its performance cannot be directly compared with that of the MEWMA chart at $\lambda = 0.4$ shown in Table 3, but it can serve as a comparison basis for the inertial effect at large shifts.

To analyse how the performances of the competing multivariate control charts deteriorate when the process dimensionality increases, experiments with $n = 4$ are shown in Tables 4 and 5. With a reminder of the properties of multivariate spaces, note that the increase in the dimensionality is reflected in the increase in the observed distances. This phenomenon arises because the data spread toward the tails of the distribution, with fewer observation vectors appearing near the process centre. Comparing Tables 2 and 4, the increase in the dimensionality reflects a natural delay in change detection in all the control charts, especially for the very small shifts.

Similar to the bivariate case, the shifts in the higher dimensionality are detected using the Lin-MEWMA chart when $\lambda = 0.1$ as soon as they are using the MEWMA chart when $d = 2$ and faster when $d > 2$ (Table 4). Fig. 5c and d present the ARL results of Tables 2 and 5 on the logarithmic scale. When $\lambda = 0.4$ (Table 5), the projected linear boundaries result in the Lin-MEWMA control chart performing best in detecting large shifts, while the CUSUM-Lin procedure shows superior performance for the very small shifts. Again, the MCUSUM performance of Table 5 cannot be directly compared with the performance of the other chartsbecause it is designed for smaller shifts and serves as a basis of comparison of the inertial effect.

The experiments demonstrated that the choice of parameters always have to include a compromise between earlier alarms for small shifts and a relative delay for large shifts, when compared to the Hotelling's T^2 chart. For the MEWMA control chart, smaller values of λ result in the proposed control chart being more efficient for the detection of smaller shifts at the cost of losing sensitivity for large shifts in the mean vector. If the chart is to detect regular and large shifts early, the ARL curve incurs a loss of performance for

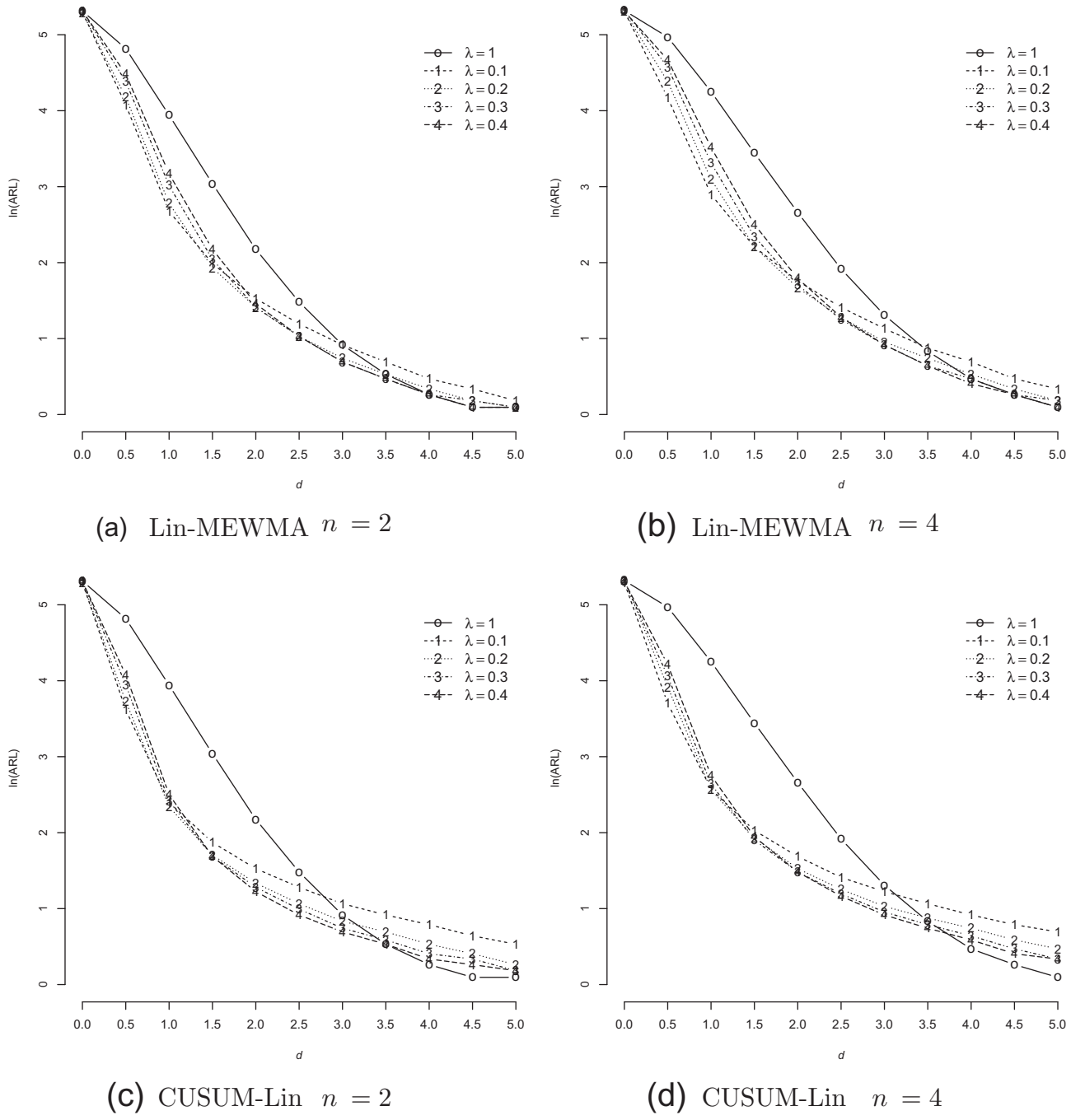

Fig. 6. ARL comparison of the Lin-MEWMA and CUSUM-Lin control charts on the logarithmic scale.

very small shifts. When the chart is configured to detect small shifts, it will tend to suffer more of the inertial effect in detecting larger shifts.

The same behaviour can be observed for all the control charts, as indicated by the ARL values in Tables 6, 7 and in Fig. 6 on a logarithmic scale. It is to be emphasised that the ARL curve of the our new scheme best fits the expected behaviour, which delays change detection for *non*-significant shifts and accelerates the detection for significant shifts.

Upon comparing the ARL curves of the quadratic classifiers and the proposed methodologies, the choice of λ depends on the dimensionality of the problem. As discussed earlier in this work, on increasing the dimension of the multivariate process, the changes will be significant at higher distances. Therefore, λ must be increased for better detection of large and meaningful shifts. The experiments demonstrate that for small dimensions such as those tested in this work, $\lambda < 0.2$ is recommended. If different values of k are set in the CUSUM-Lin chart, for example, increasing values, the ARL is expected to be more sensitive to large shifts in the mean vector at the cost of performing worse for small shifts.

From the results, it is clear that there is a correlation between the smoothing parameter and the process dimensionality. As the control chart's performance depends on the noncentrality parameter, the amount of change in the process that is significant must also be considered in terms of the process dimension to select appropriate values for the smoothing parameter λ.

4. Conclusion

The problem of finding the best directions by means of reducing the data dimensionality with linear methods such as PCA often results in a poorer performance than that obtained using the traditional quadratic approaches. This occurs because there is no definitive rule for selecting a reduced number of principal components to effectively monitor. In fact, the linear projection involved in the PCA is limited to orthonormal transformations and are not indicated for classification purposes. When the direction of change is not known *a priori*, selecting a reduced number of dimensions to be monitored can be a complicated task, resulting in an exhaustive search for several univariate and multivariate control charts. Additionally, several studies have shown that monitoring all the transformed principal components with a multivariate control chart results in the same performance obtained with monitoring the original variables.

Therefore, the proposed method presents advantages in addition to having its performance dependent only on the noncentrality parameter. In the proposed linear transformation, a reduced space dimension need not be chosen because the non-orthonormal transformation always leads to one-dimensional spaces, or a first-order statistic, irrespective of the original space dimensionality.

By maximising the trace criterion between two scatter matrices, the Lin-MEWMA control chart bounds the IC process in a projected axis oriented along the direction of the shift. It was observed that this procedure yields more efficient shift detection in terms of the ARL for intermediate and large shifts than the MEWMA chart for the same smoothing parameter. Additionally, the Lin-MEWMA control chart is selective with respect to the individual observation vectors and does not misclassify perfect units. This behaviour is not observed with the *non*-Shewhart methods as they discard perfect observations as often as when the current mean vector is considered to be OC.

Among the main issues to be covered with the use of control charts is whether it is possible to identify the causes that led to an alarm, or at least to recognise the direction of change. In the same manner of the MEWMA control chart, the components of the current mean vector can provide an indication of the shift. Moreover, in many automated production processes, operators prefer to wait for a second alarm before completely stop the process for inspection. In this sense, while the self-oriented control chart can provide useful information for the inference of the shift direction with the presence of a second observation, the quadratic MEWMA chart control does not provide such useful information. Observe that connecting the in-control mean vector, the actual estimated mean vector and plus two individual observations allows a very informative indication on the shift's direction, which cannot be obtained only by the MEWMA control chart.

A question remains regarding the size of the shift to be monitored. As described in the discussion section, the performance of the Lin-MEWMA control chart is affected by the smoothing factor λ similar to the MEWMA chart. Depending on the choice of λ, the performance of the proposed control charts lies midway between that of the MEWMA and Hotelling's T^2 charts. The experiments have shown that it is possible to achieve the same alarm rate as that of the MEWMA control chart for very small shifts in the mean vector and still maintain performance comparable to the Shewhart-type chart if the actual shift is large.

The k factor is also an issue in terms of the CUSUM-Lin chart, but based on the extensive CEP literature, $k = 0.5$ represent a reasonable choice for this parameter in low dimensional problems. As demonstrated in the experiments for both dimensions ($n = 2, 4$), even though performance is diminished for large shifts, the CUSUM-Lin chart is as effective or more so than the MCUSUM and MEWMA charts for a wide range of distances.

Because the overlap between the in- and out-of-control states increases with the dimensionality, future research on this topic may include comparisons with respect to the influence of the λ factor for hyperspaces ($n > 100$). It was observed that as the transformed variable allows for the implementation of the standardised CUSUM procedure, it is also suitable for double filtering using the univariate EWMA procedure. A detailed comparison between different values of k in the CUSUM-Lin chart and different values of λ for a EWMA-based control chart is also to be undertaken in a future performance comparison study.

Acknowledgements

The authors thanks to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support for this project.

References

Ali, M., & Amirhossein, A. (2013). Monotonic change point estimation in the mean vector of a multivariate normal process. *International Journal of Advanced Manufacturing Technology, 69*(5–8), 1895–1906.

Alkahtani, S., & Schaffer, J. (2012). A double multivariate exponentially weighted moving average (dMEWMA) control chart for a process location monitoring. *Communications in Statistics – Simulation and Computation, 41*(2), 238–252.

Atkinson, C., & Mitchell, A. F. S. (1981). Rao's distance measure. *Sankhya: The Indian Journal of Statistics, 4*, 345–365.

Bersemis, S., Psarakis, S., & Paranetos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engeneering International, 23*, 517–543.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of Calcutta Mathematical Society, 35*, 99–109.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics, 23*(4), 493–507.

Choi, S. W., Martin, E. B., & Morris, A. J. (2005). Fault detection based on a maximum-likelihood principal component analysis mixture. *Industrial and Engineering Chemistry Research, 44*, 2316–2327.

Costa, A., & Machado, M. (2008). A new chart for monitoring the covariance matrix of bivariate processes. *Communications in Statistics – Simulation and Computation, 37*, 1453–1465.

Costa, A., & Machado, M. (2011). A control chart based on sample ranges for monitoring the covariance matrix of the multivariate processes. *Journal of Applied Statistics, 38*(2), 233–245.

Costa, A., & Machado, M. (2013). A single chart with supplementary runs rules for monitoring the mean vector and the covariance matrix of multivariate processes. *Computers and Industrial Engineering, 66*, 431–437.

Crosier, R. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics, 30*(3), 291–303.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Boston: Academic Press.

Guh, R., & Shiue, Y. (2005). On-line identification of control chart patterns using self-organizing approaches. *International Journal of Production Research, 43*(6), 1225–1254.

Guh, R., & Shiue, Y. (2008). An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts. *Computers and Industrial Engineering, 55*(2), 475–493.

Hachicha, W., & Ghorbel, A. (2012). A survey of control-chart pattern-recognition literature (1991–2010) based on a new conceptual classification scheme. *Computers and Industrial Engineering, 63*, 204–222.

Haertel, V., & Landgrebe, D. (1999). On the classification of classes with nearly equal spectral response in remote sensing hyperspectral image data. *IEEE Transactions on Geoscience and Remote Sensing, 37*(5), 2374–2386.

Hotelling, H. (1947). Multivariate quality control – illustrated by the air testing of sample bombsights. *Techniques of Statistical Analysis*, 111–184.

Hwarng, H., & Hubele, N. (1993a). Back-propagation pattern recognizers for control charts: Methodology and performance. *Computers and Industrial Engineering, 24*(2), 219–235.

Hwarng, H., & Hubele, N. (1993b). $\bar{X}$ control chart pattern identification through efficient off-line neural network training. *IIE Transactions, 25*(3), 27–40.

Jackson, J. E. (1991). *A user guide to principal components*. Wiley-Interscience.

Jimenez, L., & Landgrebe, D. (1998). Supervised classification in high dimensional space: Geometrical, statistical and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, 28*, 39–54.

Khoo, M., Sitt, C., Wu, Z., & Castagliola, P. (2013). A run sum Hotelling's χ^2 control chart. *Computers & Industrial Engineering, 64*, 686–695.

Khoo, M., The, S., & Wu, Z. (2010). Monitoring process mean and variability with one double ewma chart. *Communications in Statistics – Theory and Methods, 39*(2), 3678–3694.

Kourti, T., & MacGregor, J. F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology, 28*, 409–428.
Lee, M. H. (2013). Variable sampling rate multivariate exponentially weighted moving average control chart with double warning lines. *Quality Technology & Quantitative Management, 10*(3), 353–368.
Lowry, C., Woodall, W., & Rigdon, S. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics, 34*(1), 46–53.
Mahalanobis, P. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India, 2*(1), 49–55.
Mahmoud, M., & Maravelakis, P. (2010). The performance of the MEWMA control chart when parameters are estimated. *Communications in Statistics – Simulation and Computation, 39*(9), 1803–1817.
Mahmoud, M., & Maravelakis, P. (2011). The performance of multivariate cusum control charts with estimated parameters. *Journal of Statistical Computation and Simulation, 1*, 1–18.
Michelli, C. A., & Noakes, L. (2005). Rao distances. *Journal of Multivariate Analysis, 92*, 97–115.
Montgomery, D. (2001). *Introduction to statistical quality control*. New York: Wiley.
Rao, C. R. (1947). The problem of classification and distance between two populations. *Nature*, 30.
Rao, C. R. (1949). On the distance between two populations. *Sankhya: The Indian Journal of Statistics, 9*, 246–248.
Riaz, M., & Does, R. (2008). A process variability control chart. *Computational Statistics, 24*(2), 345–368.
Therrien, C. (1989). *Decision estimation and classificaton. An introduction to pattern recognition and related topics*. John Wiley & Sons.
Tou, J. T., & Gonzalez, R. C. (1974). *Pattern recognition principles*. Addison-Wesley Publishing Company.
Wang, S., & Reynolds, M. R. J. (2013). A glr control chart for monitoring the mean vector of a multivariate normal process. *Journal of Quality Technology, 45*(1), 18–33.
Yahav, I., & Galit, S. (2014). Directionally sensitive multivariate control charts in practice: Application to biosurveillance. *Quality and Reliability Engeneering International, 30*(2), 159–179.
Zhang, J., Li, Z., & Wang, Z. (2010). A multivariate control chart for simultaneously monitoring process mean and variability. *Computational Statistics and Data Analysis, 54*, 2244–2252.
Zou, C., & Tsung, F. (2008). Directional MEWMA schemes for multistage process monitoring and diagnosis. *Journal of Quality Technology, 40*(4), 407–427.