Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Confidence Control Charts for Gaussian Mean Vectors with MEWMA and Sliding Window Schemes

D. A. O. Moraes, F. L. P. Oliveira,
L. Duczmal, & F. R. B. Cruz

Relatório Técnico
RTP-02/2013

Relatório Técnico
Série Pesquisa

Confidence Control Charts for Gaussian Mean Vectors with MEWMA and Sliding Window Schemes

D. A. O. Moraes[a,*], F. L. P. Oliveira[b], L. Duczmal[c], F. R. B. Cruz[c]

[a]*Departamento de Estatística, Universidade Federal de Santa Maria, 91105-900 – Santa Maria – RS, Brazil. E-mail:* daltieri@smail.ufsm.br
[b]*Departamento de Estatística, Universidade Federal de Ouro Preto, 35400-000 – Ouro Preto – MG, Brazil. E-mail:* fernandoluiz@iceb.ufop.br
[c]*Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901 – Belo Horizonte – MG, Brazil. E-mail:* duczmal@est.ufmg.br , fcruz@est.ufmg.br

[*]**Correspondind author:** Phone-fax: (+5531) 2531-6627

Abstract: *In this paper we show that for normal distributions, Hotelling´s T^2 and multivariate exponentially weighted moving average (MEWMA) distances are directly related to the Bhattacharyya distance. This relationship provides important information about an upper bound for the misclassification probability. In fact, this useful information indicates the degree of overlap between in- and out-of-control processes. Therefore, the first purpose of this research is to present a methodology to monitor the mean vector of a bivariate Gaussian process by means of an informative control chart based on the misclassification probability bounds. Additionally, a comparison study is carried out to measure the effects of estimating the actual mean vector through the MEWMA scheme and through sliding window schemes with uniform, linear, and exponential weights. The results show that the confidence MEWMA control chart is easier to calibrate and shows less inertia for large shifts in the mean vector than the sliding window approaches. Additionally, equivalences between the smoothing parameters and the window sizes are provided for a bivariate case.*

Keywords: Gaussian point processes, mean vectors, statistical process control, noncentrality parameter, Bhattacharyya distance, sliding window.

1. Introduction

For many industrial problems, the estimation of the misclassification probability is a subject of great interest. However, such a calculation may be a rather difficult task even when the observed data are normal. Therefore, the option of monitoring a process by means of its probability of being in or out of control is usually discarded. Recent advances in statistical techniques with applications to the $\bar{X}$ and S^2 control charts include both the univariate (Faraz and Saniga, 2013) and the multivariate (Niaki and Memar, 2009) cases. In fact, recent work by Niaki and Memar (2009) covers global process monitoring by controlling the mean vector and covariance matrix simultaneously.

Considering the process control of only mean vectors, the most utilised method to monitor large shifts is the Hotelling's T^2 control chart (Hotelling, 1947). In the case of smaller shifts, the multivariate exponentially weighted moving average (MEWMA) control chart is preferred (Lowry et al., 1992), mainly because of the simplicity of its implementation when compared to its more famous counterpart, the multivariate cumulative sums (MCUSUM) control chart (Crosier, 1998). Although the methodology described in this paper may be extended to multivariate global process monitored by probabilities, as an initial proposal, we only consider the process control of multivariate mean vectors.

If a closed-form expression is not available for the misclassification probability, one may seek either an approximate expression or an upper bound for the probability. A closed-form expression for the upper bound would be quite useful for many reasons. First, the computational effort would be reduced. Second, the evaluation of a simple formula would facilitate real-time insightful inferences about the actual process state. Furthermore, the misclassification error is known to increase significantly with the number of dimensions (Fukunaga, 1990), dramatically reducing the standard confidence

levels for the process actually being in control. Due to this fact, the evaluation of a probability measure instead of raw distances would provide more valuable information about the price to be paid for not knowing the alternative process state *a priori*. Focusing on this objective, this paper discusses the monitoring of Gaussian mean vectors by means of a simple distance transformation that leads to a control chart directly based on probabilities.

Additionally, when a process is monitored for small magnitude shifts in the mean vector with MEWMA-based control charts, another question that arises is the inertial phenomenon, which is known to delay change detection when such a change is of a large magnitude (Lowry et al., 1992). When avoiding the inertial phenomenon is essential, the analyst may seek alternative approaches to estimate the actual mean vector, which includes discarding old observations by means of some type of sliding window (SW) scheme. While the MEWMA method accumulates information about all the previous observations into the actual mean vector, the SW approach lowers the relative influence of old observations by giving heavy weights to only the most recent observations.

In fact, many authors (e.g., see Hwarng and Hubele, 1993a, 1993b; Guh and Shieu, 2005) have suggested a moving window approach as the essential tool for on-line pattern identification. However, two problems may be anticipated. The first problem is choosing the appropriate window size. The second problem is addressing unnatural patterns, i.e., when a misalignment of the pattern in time may occur. Additionally, the identified pattern could be different from the training pattern (Guh and Shieu, 2008; Hachicha and Ghorbel, 2012), and dynamic window sizes may be more appropriate. However, the use of dynamic sizes for the SW schemes is beyond the scope of this paper. As demonstrated in the computational experiments presented in this paper, the

use of fixed window sizes reflects directly on the magnitude of the shift to be detected. Some authors (e.g., Nikiforov, 2001) use SW schemes only with the significant observations from past data, but this procedure is excessively time consuming.

To provide an analysis of the effects of estimating the actual mean vector based on the MEWMA and SW approaches, three types of SWs are proposed here for comparison purposes. The possible weights for the SW observations may be uniform, linear, or exponential. As the SW size is an important parameter, the probability control charts using the SW schemes can also provide benchmark criteria for existing and future developments. Additionally, a probability-based control chart facilitates the comparative study through standardisation of the statistical distances into a 0-1 interval as upper bounds for the usual confidence levels.

Section 2 describes the main properties of the noncentrality parameter traditionally used to monitor the mean vector with Hotelling's T^2 and the MEWMA control charts. Also presented is the link between the noncentrality parameter and an upper bound for the misclassification probability. In Section 3, some computational experiments are presented to compare the performances of Hotelling's T^2 and the probability control charts. Further, the MEWMA and the SW approaches are compared by means of either their average run lengths (ARLs) or their average time to signal (ATS) because the intervals between observations are regular. Finally, Section 4 provides some final remarks and recommendations.

2. Methodology

2.1 A review of the noncentrality parameter

It is well known that the performance measured by the average run length (ARL) of traditional control charts such as Hotelling's T^2 and MEWMA depends only on the

noncentrality parameter and not on the shift's direction (Lowry et al., 1992). This distance is given by

$$d_t^2 = (\mathbf{X}_t - \mathbf{M}_0)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} (\mathbf{X}_t - \mathbf{M}_0), \quad (1)$$

where $\mathbf{X}_t$, $\mathbf{M}_0$ and $\boldsymbol{\Sigma}_0$ are the observed vector, the in-control mean vector, and the in-control covariance matrix, respectively. The decision rule gives an out-of-control signal when $d_t^2 > h_1$, where h_1 is a specified threshold that leads to a pre-specified false alarm rate, usually defined in terms of the in-control average run length (ARL_0).

In his original paper in 1947, Hotelling suggested the utilisation of d^2 instead of *d* to avoid the labour of extracting the square root, but with the massive increases in computational power in the last decades, this problem is no longer relevant. Thus, scale transformations on *d* do not modify the chart's performance. To maintain clarity in the effect on the in-control limits, in this paper, *d* is varied in the 0-4 range for ARL comparisons and in the 0-7 range for comparisons of the first and second order statistics at the transition phase. The phase transition can be understood as a limited run length on which the monitored process is moving from the in-control state to the out-of-control state. As it is expected that a shift from the in-control process should be investigated as soon as detected, the transition phase statistics are more important for analysis than the statistic after the stabilisation of the out-of-control process, which is seldom observed in practice. This subject will be further explored in this section.

While Hotelling's T^2 considers global process monitoring using outlying observations that are outside the in-control boundaries, the MEWMA statistic considers the entire process to be out-of-control as soon as

$$z_t^2 = (\mathbf{M}_t - \mathbf{M}_0)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{M}_t - \mathbf{M}_0) > h_2, \quad (2)$$

where $\mathbf{M}_t$ is the mean vector estimated with past and current information by a MEWMA scheme, such that

$$\mathbf{M}_t = (1-\lambda)\mathbf{M}_{t-1} + \lambda \mathbf{X}_t, 0 < \lambda \leq 1. \quad (3)$$

Observe that when $\lambda = 1$, the MEWMA distance reduces to the Hotelling's T^2 distance. This version of the MEWMA scheme considering equal weights for all variables is a reduction of the more general case in which different weights can be set to each variable of the vector of observations (Lowry et al., 1992). However, in this case the MEWMA chart becomes directionally oriented and the ARL may vary depending on the shift direction.

The noncentrality parameter is very popular in the pattern recognition field (Therrien, 1989), also known as Mahalanobis distance, and has a close connection to the Bhattacharyya distance, which is derived from the most general case, the Chernoff bounds (Fukunaga, 1990). Those boundaries lead to a closed-form expression for computing an upper limit for the Bayes error in the case of normally distributed processes such as

$$\varepsilon = \sqrt{P_1 * P_2} \int \sqrt{p_1(X) * p_2(X)}\, dX = \sqrt{P_1 * P_2} e^{-\mu(1/2)}, \quad (4)$$

where

$$\mu(1/2) = \frac{1}{8}(\mathbf{M}_2 - \mathbf{M}_1)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\mathbf{M}_2 - \mathbf{M}_1) + \frac{1}{2} \ln \frac{\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}}. \quad (5)$$

The term $\mu(1/2)$ is known as the Bhattacharyya distance and is used as an important separability measure between two normal distributions, where $\mathbf{M}_i$ and $\mathbf{\Sigma}_i$, $i = 1,2$, are the mean vector and the covariance matrix of each class. This distance is composed of two terms. The first term carries the information about the process difference in the mean vectors, and the second corresponds to the difference in the covariance matrices.

Rao (1947) explained that this distance is an explicit function of the proportion of overlapping individuals in the two populations. Additionally, Rao (1947) mentioned

that Bhattacharyya had developed a perfectly general measure defined by the distance between two populations based on a metric of the Riemannian geometry, with the angular distance between points representing the populations in a unit sphere.

In the case of single-hypothesis tests, such as in statistical process control (SPC) problems, the out-of-control state is generally undetermined. In such cases, instead of utilising Equation (4), which assumes two known processes, it is more interesting to evaluate only the upper bound for the Type I error, which refers only to the known process, given by

$$\varepsilon_1 = \sqrt{P_2/P_1} \int \sqrt{p_1(X) * p_2(X)}\, dX = \sqrt{P_2/P_1} e^{-\mu(1/2)}. \quad (6)$$

Additionally, as this paper is focused only on the monitoring of mean vectors, the assumption of equal covariance matrices reduces the Bhattacharyya distance to the noncentrality parameter, except by a constant, assuming the form

$$\mu(1/2) = \frac{1}{8}(\mathbf{M}_t - \mathbf{M}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{M}_t - \mathbf{M}_0), \quad (7)$$

where $\mathbf{M}_t$ is the mean vector estimated at the instant *t*, $\mathbf{M}_0$ is the in-control mean vector, and $\boldsymbol{\Sigma}_0$ is the in-control covariance matrix. This simplified form preserves all the known properties of the Hotelling's T^2 and the MEWMA control chart with respect to the performance measured by the ARL.

To examine the main properties of this distance, let us consider the distribution of d^2 with the expected vector $\mathbf{M}$ and the covariance matrix $\boldsymbol{\Sigma}$ known for the in-control (IC) process. For the general problem, consider that the mean vector of the out-of-control state (OC), unknown in practice, is defined as $\mathbf{M}_1$. According to Fukunaga (1990) the standardised distance from individual observations to the process centre is

$$d^2 = (\mathbf{X} - \mathbf{M})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \mathbf{M}) = \mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^{n} z_i^2, \quad (8)$$

where $\mathbf{Z} = \mathbf{A}^{\mathrm{T}}(\mathbf{X} - \mathbf{M})$ and $\mathbf{A}$ is the whitening transformation. Because the expected vector and covariance matrix of $\mathbf{Z}$ are $\mathbf{0}$ and $\boldsymbol{I}$, respectively, the z_i's are uncorrelated, with $E(z_i) = 0$ and $Var(z_i) = 1$. Thus, the expected value and variance of d^2 for the IC process are

$$E(d^2|IC) = n\,E(z_i^2) = n, \tag{9}$$

$$Var(d^2|IC) = E((d^2)^2) - E^2(d^2),$$

$$Var(d^2|IC) = \sum_{i=1}^{n} E(z_i^4) + \sum_{i=1}^{n}\sum_{j=1_{i\neq j}}^{n} E(z_i^2 z_i^2) - n^2\,E^2(z_i^2). \tag{10}$$

When the z_i^2's are uncorrelated (this is satisfied when the z_i's are independent), and $E(z_i^4)$ is independent of i, the variance of d^2can be further simplified to

$$Var(d^2|IC) = n\gamma, \tag{11}$$

$$\gamma = E(z_i^4) - E^2(z_i^2) = E(z_i^4) - 1. \tag{12}$$

For normal distributions, when the z_i's are uncorrelated, they are also independent. Therefore, Equation (10) can be used to compute $Var(d^2|IC)$, and $\gamma = 2$. Note that in Equations (9) and (10), only the first and second order moments of d^2 are given. However, if the z_i's are normal, the density function of d^2 is the gamma density with $\beta = n/2 - 1$ and $\alpha = 1/2$. Because the z_i's are obtained by a linear transformation of $\mathbf{X}$, the z_i's are normal if $\mathbf{X}$ is normal. Note that the gamma distribution becomes an exponential distribution for $n = 2$. Indeed, the distribution of d^2 with the mean n and standart deviation $\sqrt{n\gamma}$ approximates the normal distribution when *n* is large (Fukunaga, 1990).

Considering the OC state with mean vector $\mathbf{M}_1$, the expected value of d^2 under the assumption of equal covariance matrices is given as

$$E(d^2|OC) = n + \mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1, \tag{13}$$

and the variance is given as

$$\mathrm{Var}(d^2|\mathrm{OC}) = 2n + 4\mathbf{M}_1^T\mathbf{M}_1. \quad (14)$$

These results may be extended to the case in which the sample mean and sample covariance matrix are used in the place of known parameters, as

$$\zeta = \frac{1}{N-1}\left(\mathbf{X} - \widehat{\mathbf{M}}\right)^T\widehat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{X} - \widehat{\mathbf{M}}\right). \quad (15)$$

When $\mathbf{X}$ is normal, ζ has the beta distribution (Fukunaga, 1990) with $\mathrm{E}(\zeta|\mathrm{IC}) = n/(N-1)$ and $\mathrm{Var}(\zeta|\mathrm{IC}) = 2n/(N-1)^2$.

The first and second order moments for Bhattacharyya distance for equal covariance matrices are easily deduced from the results above. The simulated experiments presented in the following section correspond with the presented theoretical values for the first and second moments of Bhattacharyya distance and for Hotelling's T^2 with high precision for the IC state. Due to the transition phase explained earlier, the statistics of the OC state do not converge to the expected values unless we consider an extended run length after the first alarm is signalled. Thus, we are not interested in confirming these asymptotic results for the stationary OC state but rather in inspecting statistical behaviour in the transition phase with fixed run lengths.

2.2 An alternative confidence control chart

Based on the theoretical results presented in the previous subsection, we propose a different look at process monitoring. Without actually modifying the control chart performances, one could transform the statistical raw distances and their respective in-control boundaries into probability values. First, if there is no special reason to weight the in- and out-of-control processes differently, the processes are equally weighted in Equation (6), thus reducing the upper bound on the type I error to $\exp(-\mu(1/2))$. If different weights for the processes are utilised, the result will be a scale modification on the statistic values while still preserving the 0-1 domain.

Observe that when the process is actually in control, either the estimated mean vector or the individual observations must not be significantly different from the in-control standard error levels. This fact leads to an upper bound of ε_I that is close to one because the in-control and current processes are completely overlapped. When the mean vector shifts from the in- to the out-of-control state, the upper bound on ε_I decreases, indicating less overlap among the processes. However, if the complementary probability is taken, it indicates an upper bound for the confidence level, which is closer to zero, meaning that the current process is not separate from the in-control state.

Based on such considerations, a confidence control chart utilising individual observations is taken as the standard level for the different ways of estimating the mean vector. This approach can be viewed as the MEWMA chart with $\lambda = 1$ or a sliding window chart with unitary window size. For this reason, this control chart is identified by the SW1 code (sliding window of size 1). This control chart is a simple scale transformation of Hotelling's T^2 with the use of Bhattacharyya distance, triggering a signal when

$$p_t = 1 - \exp\left[-\tfrac{1}{8}(\mathbf{X}_t - \mathbf{M}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\mathbf{X}_t - \mathbf{M}_0)\right] > h_1^*, \tag{16}$$

where h_1^* is the in-control upper limit to achieve a desired ARL_0.

If the individual observed vector is changed by a mean vector, it is possible to utilise the MEWMA or the SW schemes for estimation of the vector. Equation (3) is utilised to estimate $\mathbf{M}_t$ in the case of an MEWMA-based control chart. For all methods utilising mean vector estimates instead of individual observations, the probability control chart triggers an out-of-control signal as soon as

$$p_t = 1 - \exp\left[-\tfrac{1}{8}(\mathbf{M}_t - \mathbf{M}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\mathbf{M}_t - \mathbf{M}_0)\right] > h^*. \tag{17}$$

where h^* is chosen to achieve a desired ARL_0.

For all cases of SW schemes, only the observation vectors inside the current window are weighted and the mean vector is given by

$$\mathbf{M}_t = \sum_{i=t-k+1}^{t} w_i^* \mathbf{Y}_i, \tag{18}$$

with $\sum_{i=t-k+1}^{t} w_i^* = 1$ and k representing the window size. Let us now describe each type of weight considered.

In the *uniform* SW (USW) approach, the weights are equal for all the observations inside the window of size k, with w_i^U given by

$$w_i^U = \frac{1}{k}, i = t-k+1, \ldots, t. \tag{19}$$

The *linear* SW (LSW) approach gives more weight to the most recent observation and decreases linearly the weight as the observation gets older such that

$$w_i^L = \frac{\frac{j}{k}}{\sum_{j=1}^{k}\left(\frac{j}{k}\right)}. \tag{20}$$

Finally, in the *exponential* SW (ESW) scheme, the weights for the observation vectors inside the window are distributed by

$$w_i^E = \frac{j^{\varphi}}{\sum_{j=1}^{k}(j^{\varphi})}. \tag{21}$$

where φ is a smoothing factor between 0 and 1. When $\varphi = 1$, the exponentially weighted window converges to the uniform window. The smoothing factor φ utilised for the exponentially weighted window is fixed to 0.7 as it decays below 0.5 after two steps. The calculation of individual weights for the three proposed sliding window schemes of size 4 is illustrated in Table 1, while the weights for the windows with size 2 are shown in Table 2.

TABLE 1 IS AROUND HERE.

TABLE 2 IS AROUND HERE.

The control chart calibration procedure is performed in two steps to achieve an $ARL_0 = 200$ for all control charts. The first step adjusts the linear regression models in the form $d^2 = a + b * \ln(\text{ARL})$. This procedure gives an approximate first estimate of in-control thresholds for each chart. The second step in the calibration procedure iteratively adjusts the threshold by interpolation. The next section illustrates the functionality of the proposed control chart and presents analysis of the comparative experiments.

3. Results and discussion

The first part of the experiment compares Hotelling's T^2 and the SW1 control chart, which performs a scale transformation of Hotelling's distance. Figure 1-(a) shows the signal pattern for the case of no change in the mean vector (that is, $d = 0$), while Figures 1-(b) and 1-(c) shift the mean vector process at time $t = 201$ to the distances d = 3 and d = 6, respectively. In the respective scatter plots illustrated in Figures 1-(d), 1-(e) and 1-(f), the out-of-control observation vectors are marked with light red dots, while the in-control vectors are marked with dark black dots. In Figure 1, the vertical lines in the middle of the chart delimit the change point. The horizontal dashed lines are the in-control thresholds for the pre-defined $ARL_0 = 200$. Given a probability value, the in-control upper limit for the SW1 chart is $h^*_{SW1} = 0.7362$ (73,62%). The corresponding in-control noncentral distance that holds for an $ARL_0 = 200$ in the Hotelling's T^2 control chart is $d = 3.265$, which is a scale transformation of h^*_{SW1}.

FIGURE 1 IS AROUND HERE.

Noteworthy in Figure 1-(c), is the fact that most of the out-of-control observation vectors do not overlap with the in-control region, resulting in probability values converging to 1. This result indicates that when considering individual observation vectors, the confidence level converges to 1 when the processes do not

overlap. This pattern does not hold for Hotelling's T^2 statistic because no bound exists there for the maximum values, which makes out-of-control signals difficult to interpret.

A more detailed summary of the raw distances and their equivalent confidence levels are given in Table 3, where $\overline{d^2}$ and $\bar{p}$ are the average values and Sd(*) are the standard deviations computed over 100,000 sample replications of size 10. Notice that the simulated experiments confirm with high precision the parameters of the Hotelling's T^2 statistic for the IC state (i.e., $d = 0$). As expected, the ARL for both charts is the same, which indicates that the transformation of Hotelling's T^2 into probabilities using the Bhattacharyya distance does not actually modify ARL performance. As indicated earlier, for $d > 0$ the estimates may not converge to the expected values due to the transition phase affecting the first 10 observations after the process has changed. As the ARL is a function of *d*, a fixed run length is affected differently for distinct values of *d*.

TABLE 3 IS AROUND HERE.

Figure 2-(a) is composed of four sets of control charts, and their respective two-dimensional scatter plots are shown in Figure 2-(b). The MEWMA-based control charts utilise $\lambda = 1$, such that they perform identically to the SW1 control chart, at a maximum standard confidence level to protect the global in-control process region.

FIGURE 2 IS AROUND HERE.

As evident in Figure 2, both USW and ESW control charts with sliding windows of size 2 (SW2) performs identically because $\varphi = 1$. Also noticeable is the reduction in the in-control limits of the USW and ESW charts in Figure 2, which is $h^*_{\mathrm{USW}} = h^*_{\mathrm{ESW}} = 0.4811$ (d = 2.2908). This reduction indicates that the chart becomes sensitive to small changes in the mean vector, no matter the individual distances. The in-control limit for the LSW chart with SW2 scheme is $h^*_{\mathrm{LSW}} = 0.5166$ (d = 2.4115).

Figure 3 shows the reducing effect on the confidence levels for all control charts. The MEWMA-based control chart with λ = 0.7 is called MEWMA.7 and the transformed in-control limit is $h^*_{\text{MEWMA.7}} = 0.5086$ (d = 2.3842). For the SW2 chart with the ESW scheme in Figure 3, the estimated threshold is $h^*_{\text{ESW}} = 0.4901$ (d = 2.3212).

FIGURE 3 IS AROUND HERE.

Despite the fact that the control charts become more sensitive to small shifts in the mean vector, a drawback of the USW, LSW, and ESW schemes with SW2 is noteworthy, a drawback that allows some extreme, clearly out-of-control values to be considered in-control. In the same manner, many vectors that could be considered in-control are marked with out-of-control dots. This happens because the observation vector receives, at the instant t–1, too much weight in the SW approach for the current observation vector to compensate (see Table 2). The MEWMA-based control chart seems to avoid this problem, providing a better differentiation between the in- and out-of-control vectors. Such behaviour is because the MEWMA scheme accumulates all the past information in the current mean vector, while the SW scheme does not.

More detailed information concerning the mean and standard deviation of the transformed statistics for all control charts are given in Table 4 and 5. Notably, there is a reduction in the in-control limits for small distances. That reduction provides insight into the optimum distance that can be efficiently detected for each chart configuration, which is below d = 3 for the MEWMA.7 and SW2 charts.

TABLE 4 IS AROUND HERE.

TABLE 5 IS AROUND HERE.

The ARL comparisons between the MEWMA.7, MEWMA.4, SW2 and SW4 control charts are given in Tables 6. Observing Table 6, although the SW2 control

charts perform better than the SW1 chart (Table 1) and similarly to the MEWMA.7 chart for small shifts, an inertial effect is visible for distances larger than $d = 3.0$ in the case of SW schemes. Ordering the schemes from the least to the most sensitive with respect to the inertial effect, the MEWMA chart performs better, followed by the USW, ESW and LSW charts.

TABLE 6 IS AROUND HERE.

When the window size increases to 4 (SW4 control charts), the MEWMA-based control chart has the λ parameter decreased from 0.7 to 0.4, and for comparison purposes, is called MEWMA.4. Figure 4 illustrates the standard patterns for the four confidence control charts for a shift of magnitude $d = 3$. The respective in-control limits are very close to each other, and all of them lead to completely separable processes, which are $h^*_{MEWMA} = 0.2747$ (d = 1.6028), $h^*_{USW} = 0.2677$ (d = 1.5789), $h^*_{LSW} = 0.3190$ (d = 1.7530) and $h^*_{ESW} = 0.3082$ (d = 1.7168). Table 5 provides the summary statistics for the MEWMA.4 and SW4 schemes.

FIGURE 4 IS AROUND HERE.

Figure 5-(a) shows an ARL comparison of all control charts, while Figure 5-(b) uses the natural logarithm to amplify the differences. Splitting the comparison into two groups, Figure 6-(a) compares the SW1, MEWMA.7 and SW2 control charts, while Figure 6-(b) compares the SW1, MEWMA.4 and SW4 control charts. Noticeable in the figures is the high degree of inertia effect produced by the SW schemes.

FIGURE 5 IS AROUND HERE.

The second set of control charts in Figure 6-(b) compares the SW1, MEWMA.4 and SW4 charts. Although these control charts perform better for shifts below $d = 2.0$, they have a higher degree of inertial effect than the SW2 charts for shifts in which $d = 4$. Again, the USW4 has the best performance, which is comparable to the MEWMA.4

chart. With respect to robustness against the inertia impact of large shifts, the USW approach seems to be the most effective scheme. The LSW and ESW schemes perform worse in both cases when compared to the USW scheme for large shifts. While the differences between the SW schemes for small shifts are not evident in the SW2 charts, the LSW4 and ESW4 charts perform worse than the USW4 chart for small shifts as well.

FIGURE 6 IS AROUND HERE.

Figure 7 presents comparisons between the mean values produced by the MEWMA chart varying the λ parameter from 1 to 0.1 by 0.1 intervals and the mean values for the sliding windows chart with window sizes 1, 2, 4, 6, 8, 10, 12, 14, 16 and 20. This experiment makes it possible to choose the appropriate window size that would present expected performance similar to a specific λ value for the MEWMA chart. As shown in the above experiments, this expected equivalent performance is limited to an ideal range of noncentral values, as the sliding windows tend to present considerably more inertia from large shifts than the MEWMA method.

FIGURE 7 IS AROUND HERE.

For example, the mean value on the MEWMA chart with $\lambda = 0.7$ is close to 15%, which is similar to the sliding window of size 2. For $\lambda = 0.4$, the window size that presents the approximate mean value is 4. Thus, if a specific magnitude of mean shift in the process requires $\lambda = 0.2$ on the MEWMA chart, to achieve similar performance with sliding window schemes one should select a window of size 10 or 12, depending on the sliding window scheme. Additionally, note that the φ parameter in the exponentially weighted window was fixed at 0.7 for all window sizes, but it can be reduced as the window size increases to compensate for differences.

A detailed comparison between the MEWMA-based confidence control chart baselines (mean values) and standard deviations for the in-control process, with λ varying from 1 to 0.1 by 0.1 units, is illustrated in Figure 8. These values agree completely with the expected ones. To analyse the out-of-control behaviour of the proposed statistic, the mean vector is shifted, with d varying in the 0.5-7 range by 0.5 units. This information on the first and second order moments of the proposed statistic also provides valid informative support for the decision makers.

FIGURE 8 IS AROUND HERE.

From the results presented in Figures 9 and 10, an interesting out-of-control statistics pattern is noticeable for the confidence chart in the transition period that is fixed to 20 observations. In Figure 9 (a), the λ's are positioned above the MEWMA bars and the window sizes for the SW schemes are specified in the horizontal axis. Note that, as the baseline decreases with the smoothing parameter λ, the standard deviation has a maximum point in Figure 9 (b) that is highly affected by the inertial period of 20 observations and that does not converge to the expected value.

FIGURE 9 IS AROUND HERE.

FIGURE 10 IS AROUND HERE.

We observe that the mean value and standard deviation of the stationary out-of-control state only makes sense if the researcher waits for the convergence after the inertial period. That waiting generally does not take place for the problems found in SPC because the monitoring stage is stopped after the first signal occurrence, and the out-of-control process stabilisation is not verified in practice.

To illustrate a common decision problem that occurs in many applications, take the example given in Figure 10. A researcher is monitoring a bivariate Gaussian process without any prior information about the direction of change. Thus, the non-directional

MEWMA control can be selected with sliding window schemes to also perform on-line pattern identification. Additionally, as the researcher has no prior information about the magnitude of the shift, a control chart to monitor large shifts can be configured at the cost of not detecting the change if it is a small change. Otherwise, the control chart can be configured to detect a small shift, but at the cost of an inertial delay if the actual shift occurring in the process is large.

In a simulated scenario such as the example in Figure 10, the threshold of all proposed control charts are nominally specified to detect a small shift utilising d = 1, which given in probability is $h^* = 1 - \exp\left(-\left(\frac{1}{8}\right) * 1^2\right) = 0.1175$. In Figure 10 (a), showing an extended run length of size 400 for the in-control process, the charts are not exactly calibrated to the same ARL, but as shown in Figure 7, they are expected to show similar performance for small shifts in the mean vector. In fact, the ESW scheme is clearly configured to detect smaller changes than all other concurrent schemes, given the excessive number of false alarms shown in Figure 10 (a). In Figure 10 (b) and (c), a short run length of size 40 is monitored when two types of shifts occur at position t = 21.

As all the control charts are configured to detect small shifts, they performed similarly in detecting a shift of size 1. However, when a large shift of size d = 7 occurred, as shown in Figure 10 (c), the ESW scheme clearly performed the worst. Such behaviour in the SW schemes can be related to the number of observations that are needed to fulfil the actual estimate of the mean vector to compensate for the information history accumulated by the MEWMA scheme.

It is interesting to note that while the expected value for the proposed statistic when d = 7.0 and λ = 0.1 can be given by

$$E\left(p|\mathbf{M}_1=(7,0)\right)=E\left(1-\exp\left(-\frac{1}{8}(d^2)\right)|\mathbf{M}_1=(7,0)\right)$$
$$=1-\exp\left(-E\left(\frac{1}{8}d^2|\mathbf{M}_1=(7,0)\right)\right)$$
$$=1-\exp\left(-\frac{1}{8}E\left(d^2|\mathbf{M}_1=(7,0)\right)\right)=1-\exp\left(-\frac{1}{8}(n+\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1)\right)$$
$$=1-\exp\left(-\frac{1}{8}\left(2+(7\quad 0)\binom{7}{0}\right)\right)=1-\exp\left(-\frac{1}{8}(2+49)\right)$$
$$=1-\exp\left(-\frac{51}{8}\right)=0.9983,$$

the observed value of the first 20 vectors in 50,000 Monte Carlo simulations converges to 0.7886. This reflects the high inertial effect suffered in the control charts configured for small change detection when the actual change happening in the process is large.

4. Conclusion

In this paper, we discuss an alternative way of monitoring Gaussian mean vectors through the use of an upper bound for the confidence that the process is in control. Instead of monitoring the noncentrality parameter, we suggest the use of the Bhattacharyya distance and its relationship with the upper bound of the misclassification error. While the traditional distance of Hotelling's T^2 has no maximum values, the proposed confidence control chart based on probabilities for individual observation vectors manifests a useful distinction between processes in the 0-1 range. In this case, when the out-of-control process becomes completely separable from (not overlapped with) the in-control process, the proposed statistic converges to 1, not going to infinity.

Additionally, we show that the probability control chart for individual observation vectors can be extended to more general cases, the monitoring of small shifts through the use of MEWMA-based control charts and control charts with sliding window schemes. In the same manner as the MEWMA method, instead of using

individual observation vectors, the sliding window approaches are commonly used to estimate the actual mean vector for different purposes, including on-line pattern recognition. We show the equivalence in performance measured by the ARL among the MEWMA-based control charts and sliding window schemes for specific parameters.

While this equivalence holds for small shifts in the mean vector, the sliding window approach proves to be more susceptible to the inertial effect for large shifts than the MEWMA-based scheme. Indeed, in the same manner that a decrease in the weighting factor λ in the MEWMA chart helps in identifying small shifts, an increase in the sliding window size corresponds to more effective detection of smaller shifts but with a greater inertial effect for large shifts than the MEWMA-based chart.

Future work on this topic includes the monitoring of the covariance matrix of a Gaussian process through the use of probability-based control charts, as well global process monitoring, i.e., the joint monitoring of the mean vector and covariance matrix of a multivariate Gaussian process.

Acknowledgments

The authors would like to thank the Brazilian agencies, CAPES, CNPq, and FAPEMIG, for the financial support.

References

Crosier, R. B. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, 1998; **30**(3): 291-303.

Faraz, A. & Saniga, E. Multiobjective Genetic Algorithm Approach to the Economic Statistical Design of Control Charts with an Application to Xbar and S^2 Charts. *Quality and Reliability Engineering International*, 2013; **29**: 407-415.

Fukunaga, K. *Introduction to Statistical Pattern Recognition*. (2nd ed.). Boston: Academic Press, 1990.

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

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

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

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

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

Hwarng, H. B. & Hubele, N. F. X control chart pattern identification through efficient off-line neural network training. *IIE transactions*, 1993a; **25**(3): 27–40.

Lowry, C. A., Woodall, W.H. & Rigdon, S. E. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 1992; **34**(1): 46-53.

Niaki, S.T.A. & Memar, A.O. A new statistical process control method to monitor and diagnose bivariate normal mean vectors and covariance matrices simultaneously. *The International Journal of Advanced Manufacturing Technology*, 2009; **43**: 964-981.

Nikiforov, I. V. A simple change detection scheme. *Signal Processing*, 2001; **81**: 149-172.

Rao, C. R. On the distance between two populations. Shankya: *The Indian Journal of Statistics*, 1949; **9**: 246-24.

Rao, C.R. The problem of classification and distance between two populations. *Nature*, 1947; 159, 30.

Therrien, C. W. *Decision Estimation and Classificaton, an Introduction to Pattern Recognition and Related Topics*. John Wiley & Sons, 1989.

Table 1: Weights computation for sliding window schemes with size 4 (SW4)

Window scheme \ *Window position*→	*t*-3	*t*-2	*t*-1	*t*	Σ
Uniform scheme	1/4	1/4	1/4	1/4	4/4
Uniform weights	0.250	0.250	0.250	0.250	1
Linear scheme	1/4	2/4	3/4	4/4	10/4
Linear weights	0.100	0.200	0.300	0.400	1
Exponential scheme	0.7^4	0.7^3	0.7^2	0.7^1	1.77
Exponential weights	0.135	0.193	0.276	0.395	1

Table 2: Weights for sliding window schemes with size 2 (SW2)

Window scheme \ *Window position*→	*t*-1	*t*
Uniform	0.500	0.500
Linear	0.333	0.667
Exponential	0.412	0.588

Table 3: Summary of Hotelling's T2 and SW1 statistics with ARL comparison

d	$\overline{d^2}$	$\mathrm{Sd}(d^2)$	ARL	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	ARL
0.0	2.000	1.850	200.6	20.00	15.74	200.6
	0.006	0.006	0.634	0.001	0.000	0.634
0.5	2.251	2.070	118.8	21.97	16.93	117.7
	0.007	0.007	0.376	0.001	0.001	0.372
1.0	3.001	2.642	43.1	27.62	19.57	43.1
	0.009	0.008	0.136	0.001	0.001	0.136
1.5	4.252	3.407	16.0	36.13	21.96	16.0
	0.013	0.011	0.051	0.001	0.001	0.051
2.0	6.003	4.263	7.0	46.39	22.98	7.0
	0.019	0.013	0.022	0.001	0.001	0.022
2.5	8.253	5.163	3.6	57.19	22.30	3.6
	0.026	0.016	0.011	0.002	0.001	0.011
3.0	11.004	6.086	2.2	67.49	20.13	2.2
	0.035	0.019	0.007	0.002	0.001	0.007
3.5	14.270	7.039	1.5	76.51	16.97	1.5
	0.143	0.070	0.005	0.002	0.001	0.005
4.0	18.021	7.987	1.2	83.86	13.39	1.2
	0.180	0.080	0.004	0.003	0.000	0.004
$h(\mathrm{ARL}_0=200)$	10.66			73.62		

Table 4: Summary statistics for the MEWMA.7 and SW2 control charts

	MEWMA.7		USW2		LSW2		ESW2	
d	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$
0.0	11.84	9.85	11.10	9.05	12.18	10.00	11.40	9.32
	0.08	0.07	0.08	0.06	0.09	0.07	0.08	0.07
0.5	14.07	11.25	13.36	10.48	14.35	11.39	13.63	10.74
	0.10	0.08	0.09	0.07	0.10	0.08	0.10	0.08
1.0	20.39	14.22	19.78	13.54	20.52	14.44	19.94	13.81
	0.14	0.10	0.14	0.10	0.15	0.10	0.14	0.10
1.5	29.87	16.92	29.36	16.44	29.74	17.45	29.38	16.78
	0.21	0.12	0.21	0.12	0.21	0.12	0.21	0.12
2.0	41.22	18.36	40.77	18.29	40.71	19.60	40.61	18.82
	0.29	0.13	0.29	0.13	0.29	0.14	0.29	0.13
2.5	53.09	18.33	52.61	18.95	52.09	20.75	52.24	19.79
	0.38	0.13	0.37	0.13	0.37	0.15	0.37	0.14
3.0	64.31	17.05	63.67	18.63	62.72	21.12	63.11	19.90
	0.45	0.12	0.45	0.13	0.44	0.15	0.45	0.14
3.5	74.06	14.94	73.19	17.65	71.86	20.99	72.45	19.42
	0.52	0.11	0.52	0.12	0.51	0.15	0.51	0.14
4.0	81.94	12.44	80.80	16.29	79.14	20.59	79.90	18.61
	0.58	0.09	0.57	0.12	0.56	0.15	0.56	0.13
$h(\mathrm{ARL}_0=200)$	50.86		48.11		51.66		49.01	

Table 5: Summary statistics for the MEWMA.4 and SW4 control charts

	MEWMA.4		USW4		LSW4		ESW4	
d	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$	$\bar{p}(\%)$	$\mathrm{Sd}(p)(\%)$
0.0	5.76	4.72	5.45	4.28	6.58	5.29	6.32	5.09
	0.04	0.03	0.04	0.03	0.05	0.04	0.04	0.04
0.5	7.84	5.94	7.61	5.65	8.64	6.64	8.39	6.44
	0.06	0.04	0.05	0.04	0.06	0.05	0.06	0.05
1.0	13.77	8.68	13.73	8.84	14.49	9.82	14.24	9.64
	0.10	0.06	0.10	0.06	0.10	0.07	0.10	0.07
1.5	22.72	11.73	22.88	12.67	23.25	13.62	22.99	13.51
	0.16	0.08	0.16	0.09	0.16	0.10	0.16	0.10
2.0	33.51	14.53	33.76	16.53	33.68	17.36	33.42	17.40
	0.24	0.10	0.24	0.12	0.24	0.12	0.24	0.12
2.5	44.92	16.78	45.04	20.01	44.55	20.64	44.26	20.88
	0.32	0.12	0.32	0.14	0.32	0.15	0.31	0.15
3.0	55.88	18.34	55.62	22.80	54.81	23.16	54.46	23.62
	0.40	0.13	0.39	0.16	0.39	0.16	0.39	0.17
3.5	65.63	19.13	64.78	24.73	63.76	24.76	63.35	25.45
	0.46	0.14	0.46	0.17	0.45	0.18	0.45	0.18
4.0	73.77	19.19	72.21	25.77	71.11	25.43	70.63	26.33
	0.52	0.14	0.51	0.18	0.50	0.18	0.50	0.19
$h(\mathrm{ARL}_0=200)$	27.47		26.77		31.90		30.82	

Table 6: ARL comparison between MEWMA and SW control charts

d	EWMA.7	USW2	LSW2	ESW2	EWMA.4	USW4	LSW4	ESW4
0.0	198.9	202.7	201.8	201.9	199.3	199.0	199.1	200.7
	4.45	4.53	4.51	4.51	4.46	4.45	4.45	4.49
0.5	83.2	82.0	84.6	81.0	52.6	54.4	61.1	59.2
	0.83	0.82	0.85	0.81	0.53	0.54	0.61	0.59
1.0	22.7	22.2	23.7	22.4	12.8	13.2	15.4	14.8
	0.23	0.22	0.24	0.22	0.13	0.13	0.15	0.15
1.5	8.4	8.0	8.6	8.1	5.6	5.7	6.7	6.5
	0.08	0.08	0.09	0.08	0.06	0.06	0.07	0.06
2.0	4.1	3.9	4.2	4.0	3.4	3.7	4.3	4.2
	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04
2.5	2.6	2.5	2.8	2.6	2.5	3.0	3.5	3.5
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
3.0	1.9	2.0	2.2	2.1	2.0	2.6	3.0	3.0
	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03
3.5	1.5	1.8	2.0	1.9	1.7	2.3	2.6	2.6
	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03
4.0	1.3	1.6	1.9	1.8	1.5	2.1	2.2	2.3
	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02

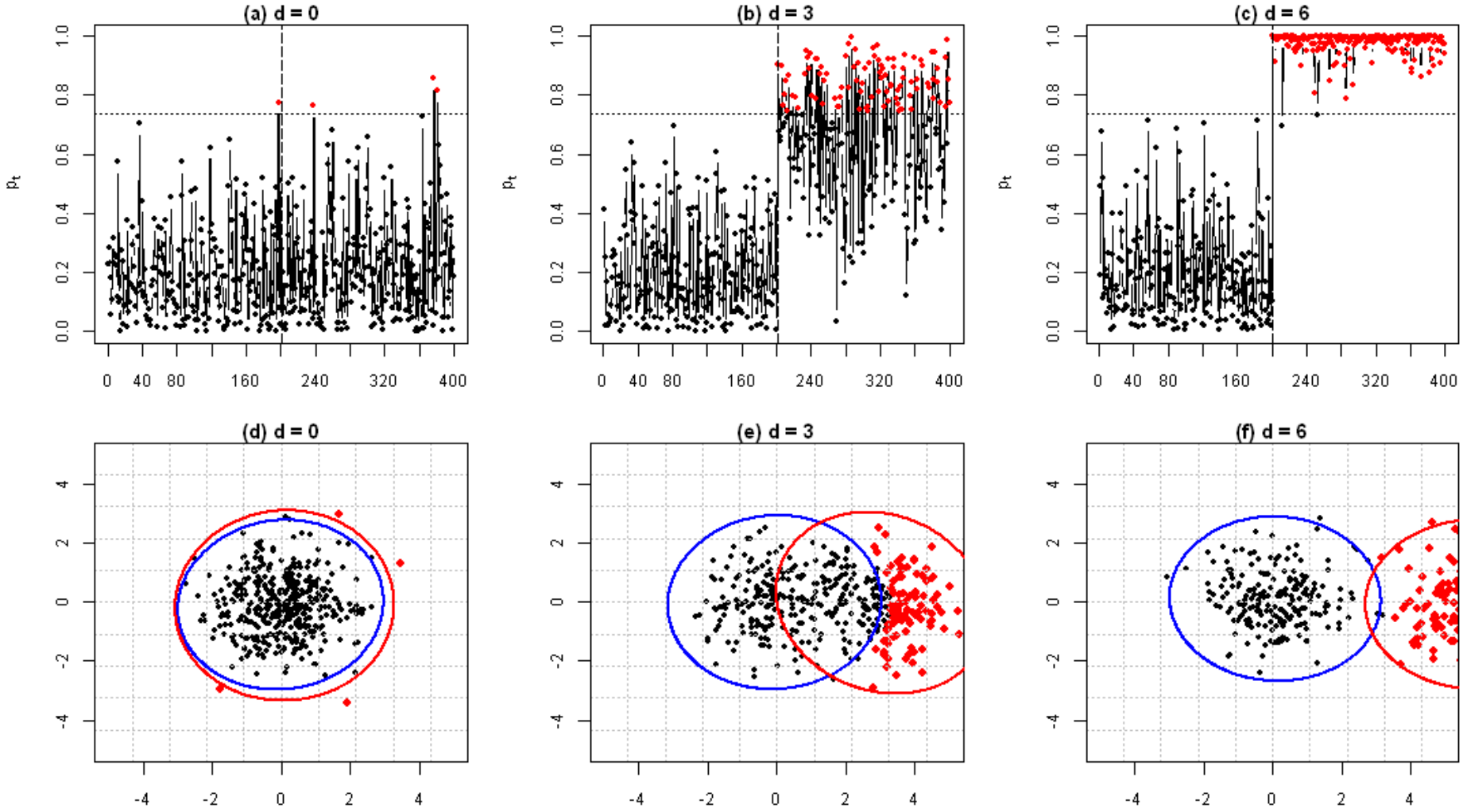
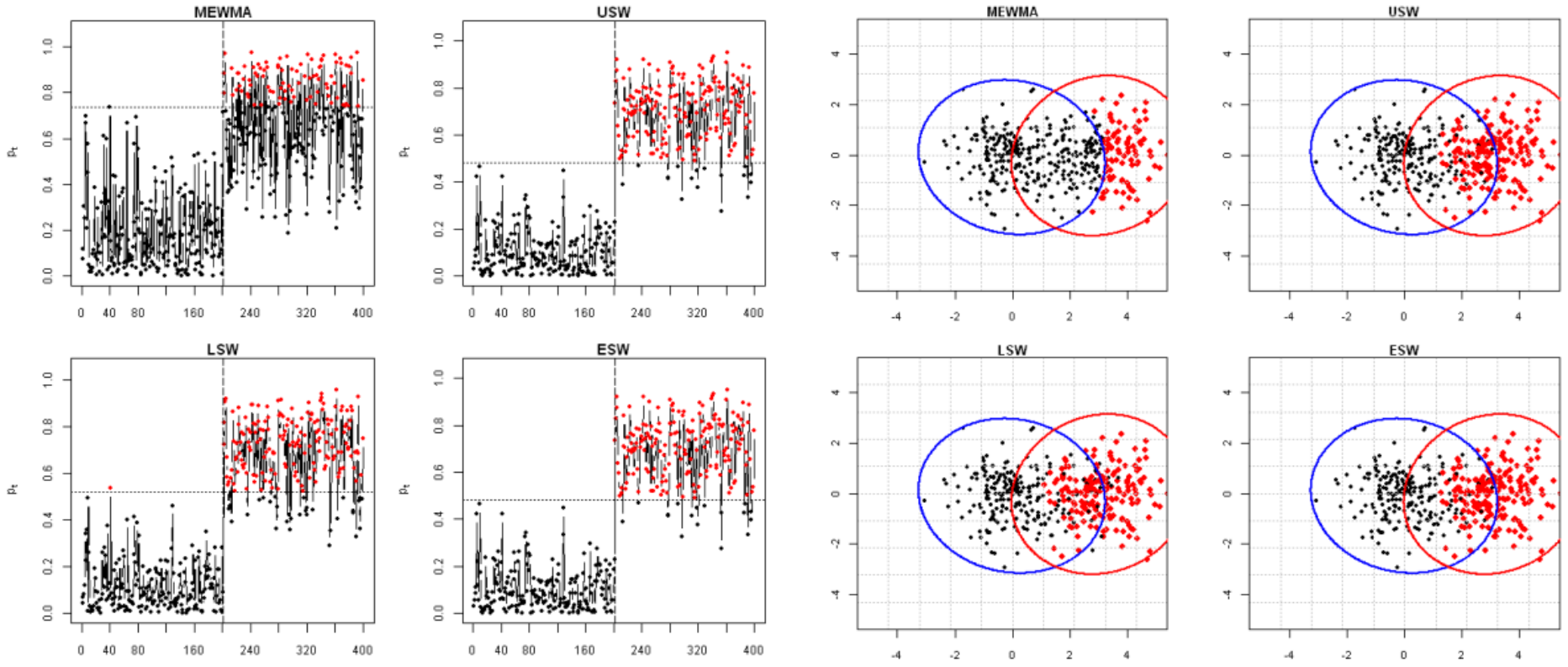
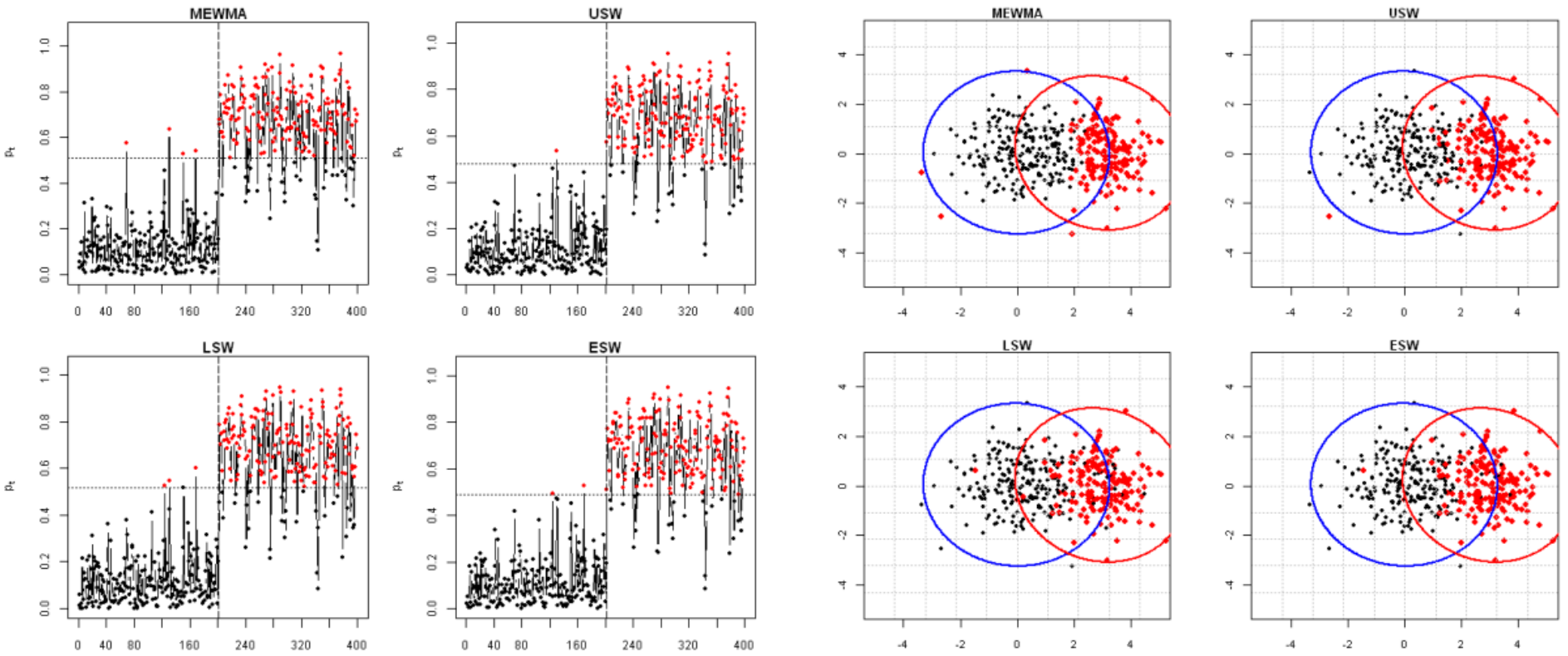


Figure 1 – Confidence control chart for individual vectors (SW1) with scatter plots



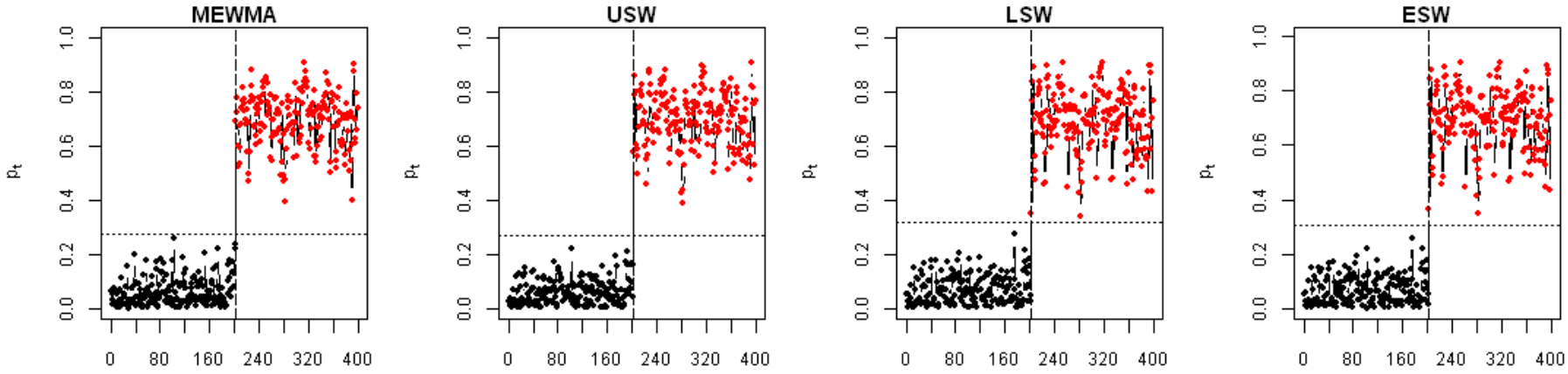
(a) Control Charts (b) Scatter plots

Figure 2: Confidence control charts (a) with λ = 1, SW2, φ = 1 and the respective scatter plots (b)



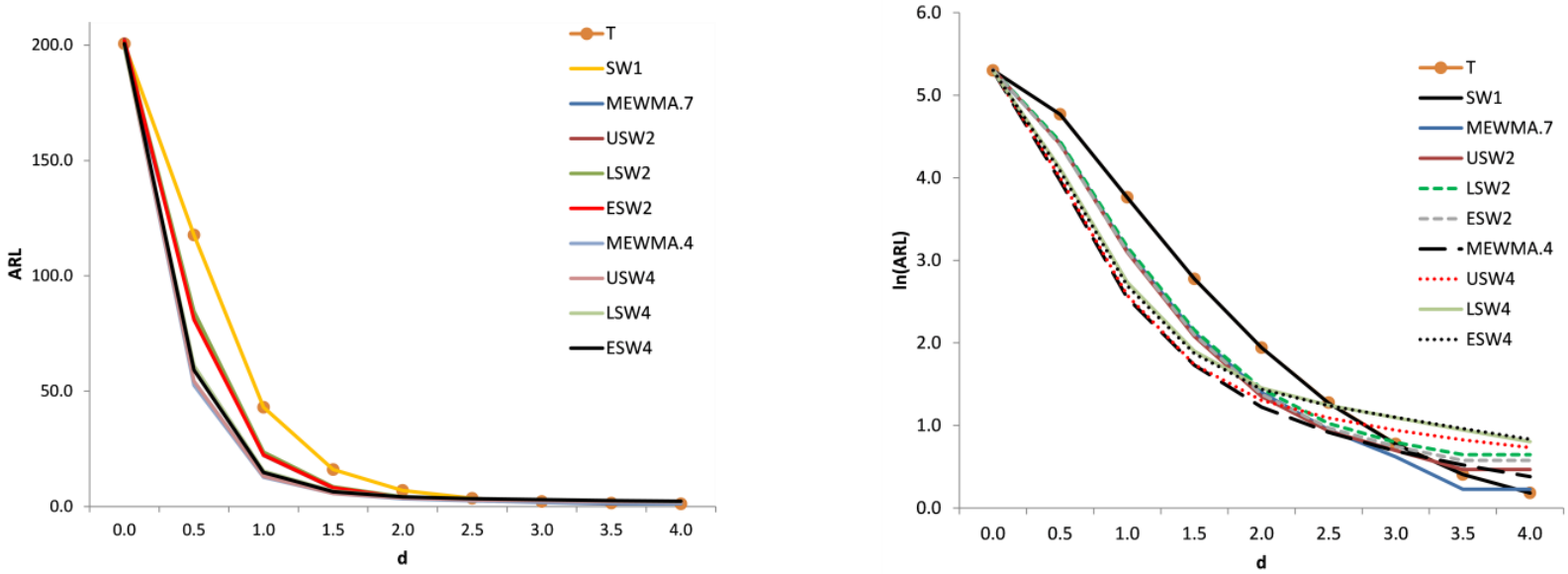
(a) Control charts (b) Scatter plots

Figure 3: Confidence control charts with λ = 0.7, SW2, φ = 0.7 and the respective scatter plots



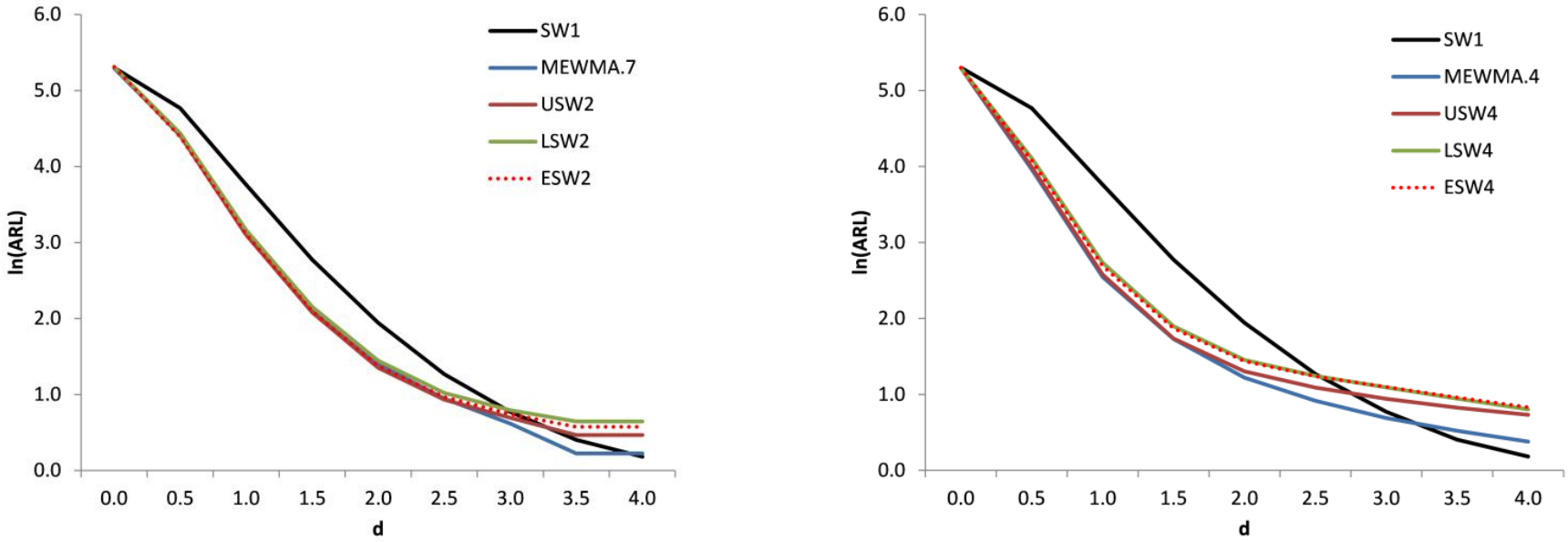
(a) (b) (c) (d)

Figure 4: Confidence control charts with λ = 0.4, SW4 and φ = 0.7



(a) ARL (b) ln(ARL)

Figure 5: (a) ARL and (b) ln(ARL) comparison for all control charts



(a) (b)

Figure 6: ln(ARL) comparison for (a) SW1, MEWMA.7 and SW2 and (b) SW1, MEWMA.4 and SW4 control charts

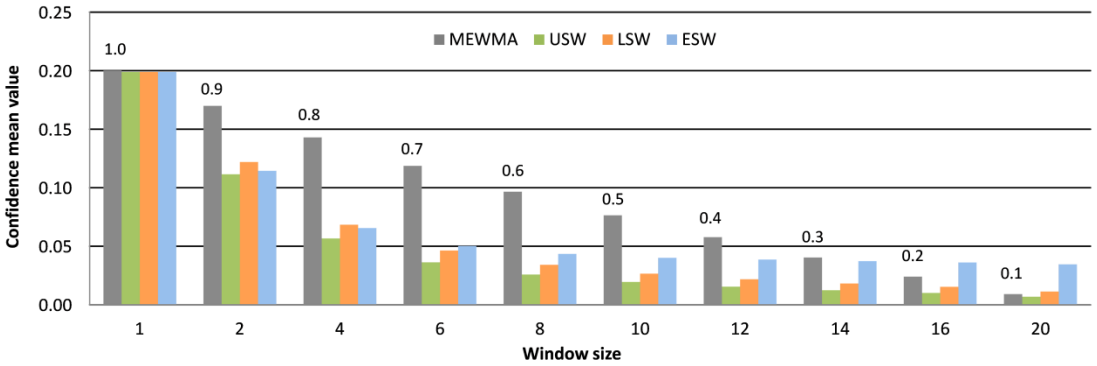


Figure 7: Comparison of mean values of the MEWMA and SW control charts

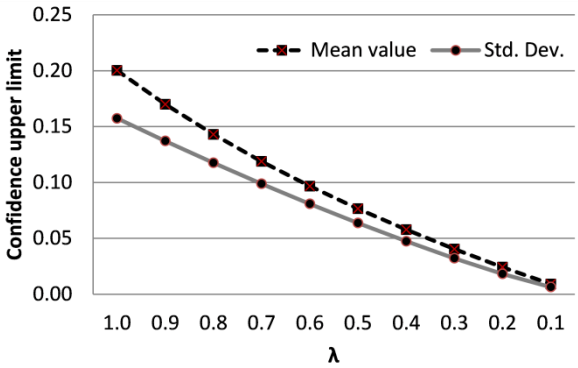
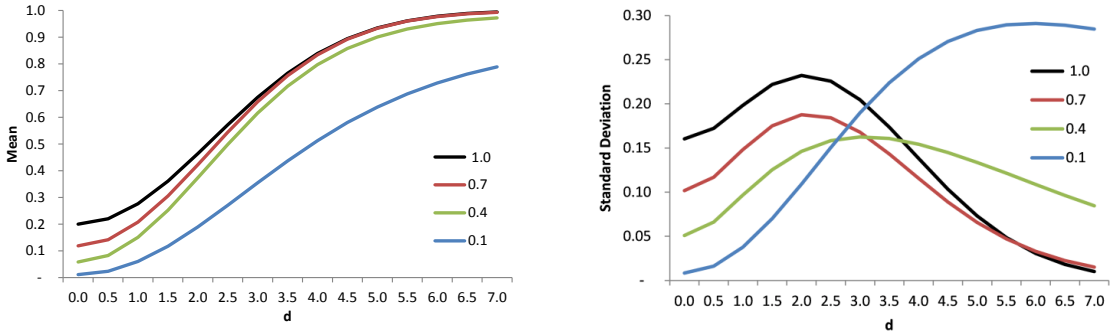
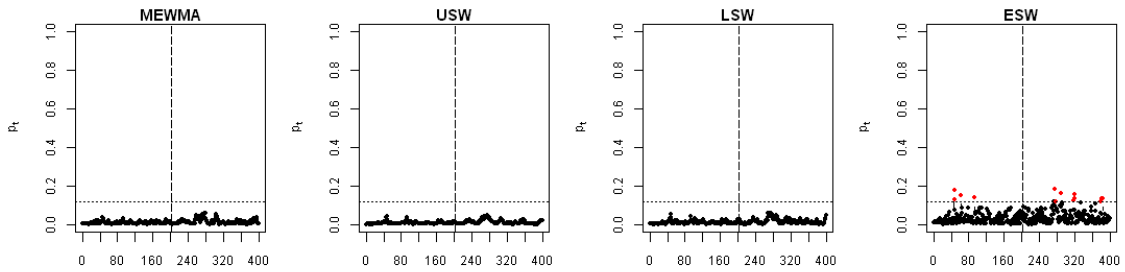


Figure 8: Mean value and standard deviation of the Confidence MEWMA CC for the in-control process with various λ's

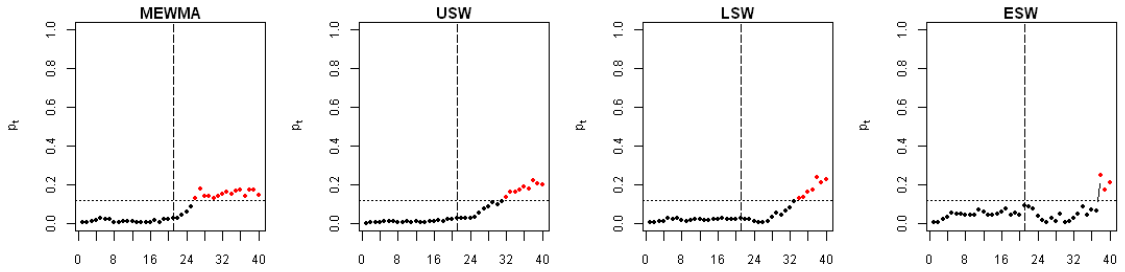


(a) (b)

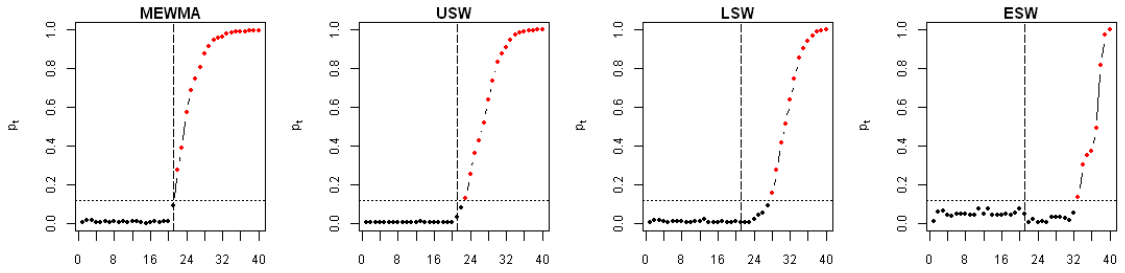
Figure 9: Mean value (a) and standard deviation (b) of the MEWMA-based control chart for the out-of-control process with various λ's



(a) $d = 0$



(b) $d = 1$



(c) $d = 7$

Figure 10: Transitional phase comparison for (a) $d = 0$, (b) $d = 1$, (c) $d = 7$ with MEWMA.1 and SW20 schemes