Denis Altieri de Oliveira Moraes

Gráficos de Controle para o Monitoramento do Vetor de Médias em Processos Gaussianos Espaço-temporais

Belo Horizonte - MG, Brasil

2014

Denis Altieri de Oliveira Moraes

Gráficos de Controle para o Monitoramento do Vetor de Médias em Processos Gaussianos Espaço-temporais

Tese apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção de Título de Doutor em Estatística.

Área de concentração: Estatística e Probabilidade

Orientador: Prof. Dr. Luiz Henrique Duczmal

Co-orientador: Prof. Dr. Fernando Luiz Pereira de Oliveira
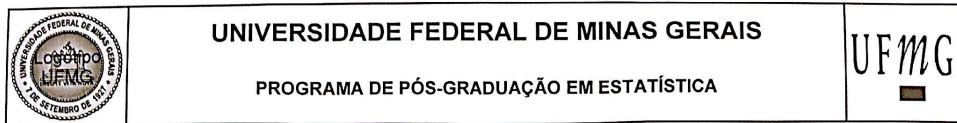
Belo Horizonte - MG, Brasil
2014

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UF*m*G

ATA DA DEFESA DE TESE DO ALUNO
DENIS ALTIERI DE OLIVEIRA MORAES

Realizou-se, no dia 17 de fevereiro de 2014, às 14:00 horas, 2076, da Universidade Federal de Minas Gerais, a 21ª defesa de tese, intitulada *"A self-oriented control chart for multivariate process location."*, apresentada por DENIS ALTIERI DE OLIVEIRA MORAES, número de registro 2010663386, graduado no curso de ESTATISTICA, como requisito parcial para a obtenção do grau de Doutor em ESTATÍSTICA, à seguinte Comissão Examinadora: Prof. Luiz Henrique Duczmal - Orientador (UFMG), Prof. Fernando Luiz Pereira de Oliveira – Co-orientador(UFOP), Prof. Frederico Rodrigues Borges da Cruz (UFMG), Prof. Roberto da Costa Quinino (UFMG), Prof.ª Linda Lee Ho (USP), Prof. Antonio Fernando Branco Costa (UNESP).

A Comissão considerou a tese:

(X) Aprovada

( ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
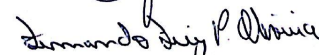Belo Horizonte, 17 de fevereiro de 2014.

Prof. Luiz Henrique Duczmal (Doutor)

Prof. Frederico Rodrigues Borges da Cruz (Doutor)

Prof. Roberto da Costa Quinino (Doutor)

Prof.ª Linda Lee Ho (Doutora)

Prof. Antonio Fernando Branco Costa (Doutor)

Prof. Fernando Luiz Pereira de Oliveira (Doutor)

Ao longo de minhas muitas vidas e até este momento, todas as virtudes que eu tenha criado, inclusive o mérito aqui gerado e todos o que vier a conseguir, dedico para o bem-estar de todos os seres. Que a doença, guerra, fome e sofrimento diminuam para todos os seres, enquanto sua sabedoria e compaixão aumentam nesta e em vidas futuras. Possa eu perceber todas as experiências como sendo tão insubstanciais quanto o tecido do sonho durante a noite e imediatamente despertar para a manifestação pura de todos os fenômenos. Possa eu rapidamente alcançar a iluminação para trabalhar sem cessar para o benefício de todos os seres sencientes. (Tara Vermelha - A Jóia Que Realiza Desejos)

Agradecimentos

Agradeço aos meu orientadores, Prof. Luiz Duczmal e Prof. Fernando de Oliveira, bem como aos demais valiosos colaboradores, pelo incentivo, sugestões e críticas positivas que culminaram na presente tese. Aos muitos colegas de estudos, os quais demonstraram grande generosidade durante o andamento das disciplinas mais difíceis, particularmente a Rodrigo Reis e Reinaldo Marques. Aos meus pais, Delci e Eneli (*in memorium*), por proverem amorosamente desde meu nascimento todas as condições para que esse trabalho se realizasse. Especialmente agradeço a Cláudia Cavalcante, por compartilhar meus dias e noites, ouvindo pacientemente todos os detalhes de minhas tempestades racionais. Finalmente, à Universidade Federal de Santa Maria e ao Programa de Pós-graduação CAPES-REUNI.

Resumo

Nessa tese são apresentadas análises comparativas de cartas de controle tradicionais e também novos métodos para o monitoramento dos vetores de médias em processos multivariados. O trabalho aborda os gráficos de controle multivariados para o monitoramento do vetor de médias de processos gaussianos com observações individuais. O controle estatístico do processo em que apenas uma observação está disponível a cada instante do tempo é um problema de difícil abordagem, já que não é possível detectar precisamente o deslocamento do vetor de médias por cartas do tipo Shewhart. Nesse caso é imprescindível o uso de cartas do tipo não-Shewhart, ou seja, considerar no instante atual a informação proveniente de observações passadas. Nesse sentido, diversos experimentos foram inicialmente realizados com o propósito de verificar a robustez dos métodos tradicionais baseados no parâmetro de não-centralidade. Foram investigadas alternativas ao método mais utilizado em aplicações práticas, o método MEWMA, com o uso de janelas deslizantes para a detecção de mudanças no vetor de médias do processo. Finalmente, foram propostos nesta tese novos gráficos de controle, também baseados no parâmetro de não-centralidade, contudo utilizando uma transformação linear mais eficiente que o método Análise de Componentes Principais. Verificou-se através de simulações de Monte Carlo que a estatística de controle proposta preenche uma lacuna existente quanto à aplicação dos métodos automáticos para o controle do vetor de médias de processos multivariados, sendo mais eficiente em termos de rapidez de detecção das mudanças do que os gráficos tradicionais em diversas situações.

Palavras-chave: Controle de qualidade, gráficos de controle, vetores de médias, processos gaussianos, observações individuais, projeções lineares.

Abstract

In this work we present comparative studies as well as new proposals on methods for statistical process control. Specifically, multivariate control charts with emphasis on monitoring the mean vector of Gaussian processes with individual observations. The statistical process control where only one observation is available at each instant of time is a difficult problem to approach, since it is not possible to accurately estimate the current process centre by means of Shewhart-type control charts, in which case it is essential to utilise non-Shewhart control charts, i.e., to consider at the current instant also information from past observations. Regard to this, several experiments were initially carried out in order to verify the robustness of the traditional methods based on the non-centrality parameter. Next, we investigated alternatives to the most common method used in practical applications, the MEWMA scheme, such as sliding window schemes for estimation of the current mean vector of the process. Finally, new control charts have been proposed, also based on the non-centrality parameter, but utilising a different criterion to obtain a linear transformation, more efficient than the known method Principal Component Analysis. It was found through experiments that the proposed statistics fills a gap regarding to the application of automata schemes for monitoring the centre of multivariate processes, being more efficient in terms of speed detection of shifts than the traditional quadratic approaches for a wide range of distances.

Keywords: Quality control, control charts, mean vectors, gaussian processes, individual observations, linear projection.

List of Figures

List of Tables

Nomenclature

α The confidence level.

$\bar{\mathbf{x}}_0$ The Phase I in-control estimated mean vector.

$\boldsymbol{\Sigma}$ The covariance matrix.

$\boldsymbol{\Sigma}_0$ The in-control variance-covariance matrix.

χ^2 The Chi-squared probability distribution.

Δ The difference for the variance of the process i.

δ The autocorrelation difference.

ϵ The Bayes error.

γ_i^2 The noncentrality parameter in the MCUSUM control chart.

λ The smoothing parameter in the MEWMA control chart.

$\mathbf{M}$ The mean vector difference from the in-control process.

ARL_0 The in-control average run length.

ARL_1 The out-of-control average run length.

$\mu\left(\frac{1}{2}\right)$ The Bhattacharyya distance.

$\mu\left(s\right)$ The Chernoff distance.

ω The statistical class or process.

Φ The autocorrelation level.

Ψ The mean vector difference in the MCUSUM control chart.

ρ The correlation coeficient.

σ_i The variance of the process i.

$\mathbf{A}$ The generic scatter matrix.

$\mathbf{e}$ The eigenvector of a scatter matrix.

$\mathbf{M}_0$ The in-control mean vector.

$\mathbf{M}_t$ The current mean vector at the instant t.

$\mathbf{S}_0$ The Phase I in-control variance-covariance matrix.

$\mathbf{X}$ The generic observed vector.

$\mathbf{x}_i$ The individual observation vector.

φ The smoothing constant in the exponential sliding window scheme.

ζ The noncentrality distance when the parameters are estimated.

a The linear coeficient of regression.

B The number of Monte Carlo simulations.

b The angular coeficient of regression.

C_t^+ The positive CUSUM-Lin statistic.

C_t^- The negative CUSUM-Lin statistic.

d^2 The noncentrality parameter.

f_i The density distribution of the i process.

h The in-control limit.

k The minimum difference in the MCUSUM control chart.

L_i The *i-th.* class.

m The number of sample vectors.

N The first occurrence of the alarm.

p The data dimensionality.

P_i The *a priori* probability of the class i

p_t The confidence control chart statistic at the instant t.

q_i The *a posteriori* probability of the class i.

$r\left(\mathbf{X}\right)$ The conditional error.

r^2 The coefficient of determination in the regression model.

s The weight of the class i in the Chernoff distance.

w The weight of the observed vector in the sliding window scheme.

X_t The Lin-MEWMA statistic at the instant t.

z_i^2 The noncentrality parameter in the MEWMA control chart.

Contents

Capítulo 1

Introdução

1.1 Apresentação

O início da revolução industrial, em torno dos anos 1800, foi marcado principalmente pela transição entre os lentos processos de produção artesanais para a produção em série com o uso de máquinas. Sendo um divisor de águas, a manufatura de itens de consumo em larga escala mudou de forma sem precedentes na história o panorama social e econômico da humanidade. Com o desenvolvimento tecnológico, surgiram próximo dos anos 1900 as rápidas máquinas a vapor, e com elas também a necessidade de evitar desperdícios controlando os processos automatizados de produção.

Sendo o principal nome associado aos estudos da área hoje denominada Controle Estatístico de Qualidade, Walter A. Shewhart, enquanto trabalhava na empresa Bell Telephones em 1925, foi pioneiro no uso de métodos estatísticos para decidir quando uma ação corretiva deveria ser aplicada a um processo. Seu principal trabalho nesse sentido foi publicado em 1931 no livro *Economic Control of Quality of Manufactured Product* (Shewhart, 1931).

Apoiado no alto poder computacional disponível atualmente para o rápido processamento e análise de volumes cada vez maiores de informações, novos métodos para

o controle estatístico de processos têm sido constantemente desenvolvidos. Da mesma forma, a presente tese tem como objetivo contribuir com para a melhoria dos métodos para o controle estatístico de qualidade, enriquecendo de forma global o grande universo do qual fazemos parte.

1.2   Principais contribuições

A abordagem das contribuições apresentadas nessa tese desenvolveu-se gradual e crescentemente. Inicialmente foi investigado os padrões de comportamento dos métodos tradicionalmente aplicados na indústria para o monitoramento do vetor de médias de um processo gaussiano. Nesse estudo introdutório foi analisado o comportamento dos limites de controle dos gráficos *multivariate cumulative sum* (MCUSUM), *multivariate exponentially weighted moving average* (MEWMA) e T^2 de Hotelling quando são utilizadas amostras de tamanho limitado proveniente da Fase I, e também considerando parâmetros conhecidos. Nesse sentido, foi observado que para lidar com a maior variação dos dados, os limites de controle tendem a aumentar à medida que períodos menores da Fase I são utilizados. A seguir, foi verificado que os gráficos de controle estudados são afetados de forma diferenciada quando diferentes fontes de variação ocorrem nos processos. Particularmente, o gráfico de controle de Hotelling apresenta maior sensibilidade para a detecção do aumento de variâncias no processo que os demais, enquanto os gráficos MCUSUM e MEWMA são mais sensíveis ao aumento de autocorrelação positiva nos processos. Apesar disso, os gráficos MCUSUM e MEWMA não apresentam nenhuma sensibilidade quanto a detecção da ocorrência de autocorrelação negativa em qualquer nível, efeito somente detectado pelo gráfico de Hotelling.

Em uma segunda etapa, foram propostos gráficos de controle de confiança, os quais derivam de uma relação probabilística das distâncias tradicionais, não limitadas superiormente, para o intervalo [0-1]. A utilização dos gráficos de confiança, além de

propor uma nova interpretação para as distâncias dos gráficos tradicionais, foi também útil para a análise e criação de uma correspondência o parâmetro de suavização λ, utilizado no gráfico MEWMA, e o tamanho de janelas deslizantes na estimação do vetor de médias. Embora útil, o tamanho da janela deslizante é um importante parâmetro até então de difícil escolha por parte do analista, que influencia diretamente a magnitude da mudança a ser detectada.

Também com a utilização dos gráficos de confiança, foi analisado o problema do efeito inercial que ocorre nos gráficos do tipo não-Shewhart. Tal efeito inercial pode ser compreendido como um retardamento na detecção de mudanças de grande magnitude, quando comparados à velocidade de detecção do método Shewhart. Nesse sentido, foi identificado que diferentemente do que podem suspeitar alguns pesquisadores, o uso de janelas deslizantes não reduz o efeito inercial em métodos não-Shewhart. Embora as janelas deslizantes sejam amplamente aplicadas em identificação de padrões comuns em certos processos, como sazonalidade e alternação de sinais, foi averiguado que seu uso amplifica o efeito inercial, retardando assim a detecção de grandes mudanças nos processos.

A principal contribuição desta tese recai sobre o problema da redução e seleção de variáveis do processo através de transformações lineares, geralmente realizado através da técnica conhecida como Análise de Componentes Principais (ACP). Destaca-se que com a aplicação da metodologia proposta no final desta tese para os gráficos de controle, foi possível obter uma redução significativa no efeito inercial. Nesse contexto, sabe-se que a ACP não fornece uma regra definitiva para o problema sobre quais componentes principais devem ser selecionadas, geralmente acarretando em perda de desempenho em termos de velocidade de detecção das mudanças. De acordo com a literatura em reconhecimento de padrões e processamento de sinais, a transformação linear realizada pelo método ACP é indicada para a representação de sinais, não sendo eficiente quando o propósito é classificação. Conforme exposto no trabalho, quando o objetivo é a classi-

ficação de dados, uma transformação linear utilizando um critério diferenciado produz uma redução de dimensionalidade mais efetiva. Como a implementação de tal critério na área de controle de processos era até o momento desconhecida, os experimentos com os novos gráficos de controle propostos evidenciaram que há de fato uma seleção ótima da direção a ser monitorada, tornando a redução da dimensão dos processos ótima no sentido de minimizar o tempo de detecção de mudanças no vetor de médias.

1.3   Organização da Tese

Como destacado anteriormente, o desenvolvimento desta tese deu-se de forma gradual. A partir do estudo e reconhecimento dos padrões de atuação dos métodos clássicos no monitoramento da média de processos multivariados, foram sendo estudadas as possibilidades de modificação desses métodos até o desenvolvimento bem sucedido de um novo método de controle.

Assim, o trabalho foi elaborado em três etapas, sendo que cada uma delas resultou em um artigo autônomo, aqui organizado sequencialmente como um único volume. Como os trabalhos foram elaborados independentemente, alguns conteúdos introdutórios aos métodos estão apresentados novamente nas três partes principais da tese.

No Capítulo 2 são apresentados os três métodos clássicos no monitoramento do vetor de médias de processos multivariados. Nesse capítulo é proposto um método de calibração comum, o estudo dos limites de controle e os padrões de desempenho dos gráficos de controle tradicionais frente a diferentes fontes de variabilidade.

O Capítulo 3 apresenta os gráficos de controle de confiança e também investiga a utilização das janelas deslizantes como abordagem alternativa para minimizar o efeito inercial dos gráficos de controle. Nesse contexto é apresentado de forma clara uma relação entre o parâmetro de suavização do método MEWMA e o tamanho das janelas deslizantes. As novas abordagens em gráficos de controle para o monitoramento do vetor

de médias de processos multivariados, denominadas de Lin-MEWMA e CUSUM-Lin, são apresentadas no Capítulo 4. Finalmente, o Capítulo 5 faz um resumo do trabalho realizado e expõe novos objetivos e planejamentos a ser implementados em futuros projetos.

Chapter 2

On the Hotelling's T^2, MCUSUM and MEWMA Control Charts' Performance with Different Variability Sources: A Simulation Study

Abstract

This work is a simulation study to investigate the sensitivity of multivariate control charts for monitoring mean vectors in a bivariate Gaussian process with individual observations. The multivariate cumulative sum (MCUSUM), the multivariate exponentially weighted moving average (MEWMA) and Hotelling's T^2 charts are selected for analysis due to their common dependency on the noncentrality parameter. The chart performance is evaluated through the average run length (ARL) or the average

time to signal. The impact of utilising thresholds computed from known parameters
or Phase I sample estimates is considered for mean vector shifts. Although designed
to monitor mean vectors, the sensibility of the control charts is additionally analysed
through different variability sources, including the mixing effect of mean vector shifts
with increasing variances or positive autocorrelation in the out-of-control process.

2.1 Introduction

Developed to study the influence of social castes in India in the early 20th century,
the Mahalanobis (1936) distance is an important example of a dissimilar metric in
various disciplines. Among many existing applications of this distance, in the statistical
process control (SPC) it is known as the noncentrality parameter. With the increase in
computational power over the last century and the growing number of applications, the
Monte Carlo method can help to understand its behaviour under different simulated
scenarios.

In SPC, the noncentrality parameter is frequently used in control charts to detect
process changes, triggering a signal as soon as the underlying process shifts from the
in-control state to the out-of-control state. To evaluate the control chart performance,
the metric typically adopted is the average run length (ARL) or average time to signal
(ATS). The ATS is the process ARL when the time interval between samples is fixed
at one time unit, as during this simulation study.

An important factor for rapid change detection is selecting the correct method, which
depends on the available data and the change to be monitored. Montgomery (2001)
elaboration of the decision schemes is a main reference to correctly choose a control
chart method. Lowry and Montgomery (1995) present an additional review. Although
the multivariate exponentially weighted moving average (MEWMA) of Lowry et al.
(1992) and the multivariate cumulative sum (MCUSUM) charts of Crosier (1988) are

popular and more suitable to detect small changes, the T^2 chart is suggested to monitor the mean vector for large-scale shifts. Mahmoud and Maravelakis (2010; 2011) use estimates of the parameters for evaluating the performance of two types of MCUSUM and MEWMA control charts. These charts are based on the noncentrality parameter and can be applied to a number p of variables, where $p \geq 1$. When $p = 1$, all methods are reduced to their respective univariate schemes, which are CUSUM, EWMA and X control charts. Recent work proposes several chart modifications, such as the double exponentially weighted moving average (dEWMA) method proposed by Alkahtani and Schaffer (2012).

To monitor the covariance matrix, Montgomery (2001) recommends the moving range and generalised variance tests, Riaz and Does (2008) suggest utilising supporting information and Costa and Machado (2008) postulate the VMAX procedure. Quinino et al. (2012) also propose a single statistic based on the mixture of variances (VMIX) to monitor the covariance matrix of bivariate processes. Yeh et al. (2006) proposed modifications of the EWMA method based on the generalised variance.

The approaches for simultaneously monitoring changes in the mean vector and the covariance matrix are numerous, and we highlight the integration of the exponentially weighted moving average procedure with the generalised likelihood ratio test of Zhang et al. (2010) and Khoo et al. (2010), whose statistics are based on the maximum of the absolute values of the two dEWMA statistics, one of which controls the mean vector and the other the covariance matrix. The numerous proposals to monitor changes in the mean vector of autocorrelated data (Montgomery, 2001) include traditional methods that fit the time series and subsequently implement control techniques on the model residuals produced by the fit.

Although designed to monitor mean vector shifts, the present study analyses the sensibility of the MCUSUM, MEWMA and Hotelling's T^2 statistics for process changes in the mean vector with different sources of variation. First, the chart performance

for mean vector shifts is compared considering known parameters or training on small samples sizes. The simulated mean vector shifts includes comparisons about the shift's direction in correlated and non-correlated processes with known parameters. Increasing variances and the influence of a vector autoregressive model are additionally measured and analysed using the ARL value, keeping the mean vector fixed.

The bivariate case is chosen to analyse the multivariate point process for their broad application in space-time problems and to provide a general example of multivariate data. Another positive aspect of studying only two variables is the visual potential to identify the in- and out-of-control observations in their original spaces in three-dimensional observations to make more direct conclusions and facilitate the understanding of what occurs in the higher-dimensional spaces. To emphasise this objective, Lowry and Montgomery (1995) recommends extending additional graphical approaches, including the Polyplot method (Blazek et al., 1987), beyond the Hotelling's T^2 statistic and including techniques, as the MEMWA chart, that effectively detect small changes. Although different methodologies for data visualization are employed together with the control charts, this work presents a tool for continuous data view in the scatter-plots, estimating confidence ellipses for the current process based on principal component analysis (PCA).

In Section 2, we present a review of the noncentrality parameter, which is the common distance used in the Hotelling's T^2, MCUSUM and MEWMA control charts is presented. Additionally we provide a brief explanation of PCA as a tool for data visualisation in the scatter-plots. Section 3 analyses results of experiments for different shifts in the mean vector with known parameters or sample estimates. As inertia problems may occur in non-Shewhart control charts, our recommendation is the use of simultaneous non-Shewhart and Shewhart-type control charts to avoid detection delay. Further, the ARLs measure the chart sensibility due to increasing variances, autocorrelation, and the resulting performance of mixing those effects with mean vector

shifts. The final section discusses the control charts performance and sensibility in the proposed scenarios and prospects for future work.

2.2   Control Charts Methodology

In the general case, suppose that vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, ... of dimension $p \times 1$ represent sequential observations of p characteristics over time. The observations $\mathbf{x}_i$, $i = 1, 2, \ldots$, are assumed to be independent random vectors of a multivariate normal distribution with a mean vector $\mathbf{M}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. Without loss of generality, consider that $\mathbf{M}_0 = (0,0,\ldots,0) = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}$.

The Hotelling's T^2, MCUSUM and MEWMA control charts analysed share the property that their performances, as measured by ARL, depend on $\mathbf{M}$ and $\boldsymbol{\Sigma}$ expressed as the noncentrality parameter d (Lowry et al., 1992), which is given by

$$d^2 = \mathbf{M}'\boldsymbol{\Sigma}^{-1}\mathbf{M} \tag{2.1}$$

When $\boldsymbol{\Sigma}_0$ is the identity matrix, d is reduced to a Euclidean distance. In his original formulation, Hotelling suggests the utilisation of d^2 to avoid the labor of extracting the square root. If the in-control process is not symmetrical around its centre of mass, as occurs for correlated variables, the Euclidean distance does not consider the process covariance, thus making it directionally dependent. To quantify the magnitude of shifts without directional dependence, the shifts should be correctly weighted by the covariance matrix. The statistical pattern recognition literature (Therrien, 1989) shows that the noncentrality parameter, also known as the Mahalanobis distance, is related to other statistical measures as the Divergence (D) and Bhattacharyya (B) distances. When the covariance matrices of two processes are equal, then $d^2 = D = 8B$.

In classification problems, those distances are present in single and composite tests.

For the composite problem, each observation is compared within two or more classes and attributed to the one with the shortest distance. The noncentrality parameter in a control chart evaluates the relative vector distance to the in-control process mean. Thus, problems related to the control charts often arise as single hypothesis tests, i.e., when the unique known class is the in-control state. The analysis of signals from radar devices is an example of single hypothesis tests applied in the pattern recognition field, when the aim is the simple recognition of potential targets.

Note the existence of two implicit assumptions in the performance comparisons based on the noncentrality parameter. First, any shift, regardless of size, must be detected as early as possible. Second, a shift from $\mathbf{M}_0$ to $\mathbf{M}_1$ is detected as quickly as a shift from $\mathbf{M}_0$ to $\mathbf{M}_2$ if $\mathbf{M}_1'\boldsymbol{\Sigma}_0^{-1}\mathbf{M}_1 = \mathbf{M}_2'\boldsymbol{\Sigma}_0^{-1}\mathbf{M}_2$. As the ARL value is a function of the noncentrality parameter d, the comparisons between the methods are simplified with analysis of the curve ARL vs. d. Alternatively, if the charts do not share this property, their relative performance may vary depending on $\boldsymbol{\Sigma}$, i.e., even for a given matrix $\boldsymbol{\Sigma}$, a chart may more effectively detect changes in some directions and less effectively in other directions.

2.2.1 Hotelling's T^2 Control Chart

The statistic proposed by Hotelling (1947) triggers a signal when there is a significant shift in the mean vector, such that

$$d_i^2 = (\mathbf{x}_i - \mathbf{M}_0)' \, \boldsymbol{\Sigma}_0^{-1} \, (\mathbf{x}_i - \mathbf{M}_0) > h_1 \tag{2.2}$$

where $h_1 > 0$ is the threshold specified to maintain a desired in-control average run length (ARL_0).

There are three asymptotic distributions in the literature to compute the thresholds

for the Hotelling's T^2 statistics. When we assume that the observations $\mathbf{x}_i$ are not time-dependent and that the process parameters are known, d_i^2 follows a Chi-squared distribution with p degrees of freedom and then $h_1 = \chi^2_{(p,1-\alpha)}$ (Seber, 1984). This is a a Phase II control chart called T^2-chart for individual observations with known parameters.

If m samples are used to compute $(\bar{\mathbf{x}}_0, \mathbf{S}_0)$, the estimates of $(\mathbf{M}_0, \boldsymbol{\Sigma}_0)$ and $\mathbf{x}_i$ is an individual observation that is not independent of the estimators, then the $d_i^2/d_0(m)$ statistic follows a Beta distribution with $p/2$ and $(m-p-1/2)$ degrees of freedom, where $d_0(m) = (m-1)^2\,m^{(-1)}$ and p is the data dimension. Thus, the upper control limit is given by $h_1 = d_0(m)\beta_{(1-\alpha,p/2,(m-p-1)/2)}$. This control chart is called a Phase I T^2-chart (Tracy et al., 1992).

If the estimators are utilised instead of the parameters and if $\mathbf{x}_f$ is a future individual observation that is independent of $(\bar{\mathbf{x}}_0, \mathbf{S}_0)$, then $d_f^2/d_1(m,p)$ follows an F-distribution with p and $(m-p)$ degrees of freedom, where $d_1(m,p) = p\,(m+1)\,(m-1)\,[m\,(m-p)]^{-1}$. Thus, the upper control limit of this multivariate Shewhart control chart is $d_1(m,p)F_{(1-\alpha,p,m-p)}$. This control chart is called a Phase II T^2-chart with unknown parameters.

Because the multivariate Shewhart control charts only consider the information given by the current observation, they are insensitive to small and moderate shifts in the mean vector. To overcome this problem, we concisely describe the multivariate CUSUM and EWMA schemes proposed in the literature.

2.2.2   Multivariate Cumulative Sum Chart

Among the multivariate CUSUM methods proposed by Crosier (1988), the method with the best properties in terms of performance triggers an alarm when the statistic

$$\gamma_i^2 = \left(\boldsymbol{\Psi}_i' \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Psi}_i\right) > h_2 \tag{2.3}$$

where $\boldsymbol{\Psi}_i = 0$, if $C_i \leq k$, $\boldsymbol{\Psi}_i = (\boldsymbol{\Psi}_{i-1} + \mathbf{x}_i - \mathbf{M}_0)(1 - k/C_i)$ if $C_i > k$, $\boldsymbol{\Psi}_0 = \mathbf{0}$, $k > 0$ and $C_i = \left[(\boldsymbol{\Psi}_{i-1} + \mathbf{x}_i - \mathbf{M}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\Psi}_{i-1} + \mathbf{x}_i - \mathbf{M}_0) \right]^{1/2}$, $i = 1,2,\dots$. The upper control limit h_2 is determined to provide a predefined in-control ARL by simulation. Because the ARL performance of this chart depends on the noncentrality parameter, Crosier recommends $k = d/2$ for a shift detection of d units.

2.2.3 Multivariate Exponentially Weighted Moving Average Chart

The MEWMA method proposed by Lowry et al. (1992) is a natural extension of the EWMA chart. Its multivariate formulation defines the EWMA vector as $\mathbf{z}_i = \boldsymbol{\lambda}\mathbf{x}_i + (\mathbf{I} - \boldsymbol{\lambda})\mathbf{z}_{i-1} = \sum_{j=1}^{i} \boldsymbol{\lambda}(\mathbf{I} - \boldsymbol{\lambda})^{i-j} \mathbf{x}_j$, $i = 1,2,\dots$, where $\mathbf{z}_0 = \mathbf{0}$, the initial in-control mean vector of the process, and $\boldsymbol{\lambda} = diag(\lambda_1,\lambda_2,\dots,\lambda_p)$, $0 \leq \lambda_j \leq 1$, $j = 1,2,\dots,p$. When $\boldsymbol{\lambda} = \mathbf{I}$, the MEWMA control chart is equivalent to the T^2-chart. Similar to other methods, this procedure triggers an out-of-control signal when

$$z_i^2 = \left(\mathbf{z}_i' \boldsymbol{\Sigma}_0^{-1} \mathbf{z}_i \right) > h_3 \tag{2.4}$$

where $h_3 > 0$ is chosen by simulation to obtain a predefined value of ARL_0 and $\boldsymbol{\Sigma}_{(\mathbf{z}_i)}$ is the covariance matrix of $\mathbf{z}_i$. If there is no reason to differentially weight the historical observations in the p characteristics, then $\lambda_1,\lambda_2,\dots,\lambda_p = \lambda$ is utilised, but when unequal weighting constants are considered, the ARL depends on the direction of the shift. The covariance matrix of $\mathbf{z}_i$ is calculated as $\boldsymbol{\Sigma}_{(\mathbf{z}_i)} = \sum_{j=1}^{i} \mathrm{Var}\left[\boldsymbol{\lambda}(\mathbf{I} - \boldsymbol{\lambda})^{i-j} \mathbf{x}_j \right] = \sum_{j=1}^{i} \boldsymbol{\lambda}(\mathbf{I} - \boldsymbol{\lambda})^{i-j'} \boldsymbol{\Sigma}(\mathbf{I} - \boldsymbol{\lambda})^{i-j} \boldsymbol{\lambda}$; when $\lambda_1 = \lambda_2 = \dots = \lambda$, $\boldsymbol{\Sigma}_{(\mathbf{z}_i)} = \left(1 - (1 - \lambda)^{2i} \right)(\lambda/(2 - \lambda))\boldsymbol{\Sigma}$. An approximation of the variance-covariance matrix $\boldsymbol{\Sigma}_{(\mathbf{z}_i)}$ as i approaches $+\infty$ is given as $\boldsymbol{\Sigma}_{(\mathbf{z}_i)} = \lambda/((2 - \lambda)\boldsymbol{\Sigma}_0)$; however, the appliance of exact variance-covariance matrix leads to a natural fast initial response (FIR) for the MEWMA

chart.

2.2.4   Confidence Ellipse Estimation by Principal Component Analysis

The principal components method is a common multivariate procedure for projecting the original variable space into an orthogonal space, so less transformed variables that represent different sources of variation can be monitored together with multivariate control charts or individually with univariate control charts (Bersemis et al., 2007). Among the applications of this very useful method in multivariate quality control, Jackson (1991) studied three types of control charts based on PCA. The first type is a T^2-control chart obtained from principal scores components, the second is a control chart for principal component residuals and the third is a control chart for each independent principal component's scores. Thus, further analysis could be made to monitor individual observations using their projections into the principal components. Bersemis et al. (2007) offer a detailed description of multivariate process control via PCA and other projection techniques. Making a distinction between signal classification and signal representation as exposed by Fukunaga (1990), the authors applies PCA as a descriptive tool, establishing in- and out-of-control regions for the current process and then visualization helps to an understand the process in conjunction with the control charts.

PCA aims to find a matrix $\boldsymbol{\Sigma}^*$ with a linear transformation of $\boldsymbol{\Sigma}$, which rotates the original axes in the directions of decreasing (or increasing) variability. In the bivariate case, the eigenvector $\mathbf{e}_1 = (e_{11}, e_{12})$ associated with the first principal component (PC_1) in the rotation matrix $\boldsymbol{\Sigma}^*$ indicates the direction of maximum process variability, and the first eigenvalue v_1 indicates the normalised size of variation in that direction. Similarly, the second principal component (PC_2) indicates the direction and magnitude of the axis with the second most significant variability, which is orthogonal to the first one.

To plot the in- and out-of-control ellipses in the scatter-plots showed in the present work, take the rotation angle of the estimated ellipse (in radians) with respect to the original coordinate system from the trigonometric rules, such as the arctangent of the first eigenvector, which points in the direction of greatest variability. Similarly, the second eigenvector provides the direction of the second most significant variation. The estimated axis sizes in the directions of the major and minor variability are normalised, and the eigenvalues are multiplied by the quantile of a multivariate normal distribution to establish the confidence region of $(1 - \alpha)\,100\%$, where α is the error probability. Assuming that the subjacent process is Gaussian, we adopt $z_{(1-\alpha)} = 3.023$, which corresponds to an error $\alpha = 0.005$ for the estimated axis size. The in-control ellipses are drawn in blue and the out-of-control ellipse in red.

2.3 Experiments and results

As described above, the calibration of the MCUSUM and MEWMA charts to obtain a predefined value for ARL_0 involves defining the k-factor in the MCUSUM and λ in the MEWMA. For the MCUSUM chart, Crosier (1988) notes that one should choose $k = d/2$ to detect a shift with a magnitude d corresponding to the noncentrality parameter. For the MEWMA method, Lowry et al. (1992) illustrate optimal schemes to choose the weighting factor λ, which generally must be in the [0.05; 0.25] range. According to the authors, the optimum suggested value of λ to detect a unitary change in the noncentrality parameter is 0.16. In the present work, $\lambda = 0.1$ is selected for the MEWMA chart as this value is a common choice in several papers to compare with the MCUSUM chart utilising $k = 0.5$ (which corresponds to a target shift detection of noncentrality value $d = 1$).

Set the k and λ values, the control limits h_i are estimated for the control charts to obtain the same ARL. To standardise the analysis with other studies, the charts are

calibrated to an ARL_0 of 200, which is an intermediate value for the mean time, after which a false alarm may be triggered. Defining in-control thresholds for each chart using the same ARL_0 ensures an equivalent type I error for the tests under the null hypothesis of no change in the process.

As the calibration procedure is complete, we compare the chart ARL_1 values, i.e., the chart performance with a shift in the process. When the underlying process is actually out-of-control with a mean vector shift, a smaller ARL_1 value corresponds to better chart performance. Conversely, the chart may trigger a signal for a different variation source, indicating that the chart is not robust for different causes of variation. Thus, the ARL_1 computed for other sources of variation and different mean vector shifts can be viewed as a disadvantage or lack of robustness of the chart and treated as a scale of sensibility. The ARL empirical computation is described in the next section. All routines for the experiments are elaborated using the R language (Development Core Team, 2008).

As previously stated, the Hotelling's T^2-control chart calibration is achieved using the process sample estimates or the known parameters to compute the thresholds by means of asymptotic distributions. An additional Hotelling's T^2-control chart calibration observes the false alarm rate and computes the probability of a signal, but we do not apply this procedure because the non-Shewhart control charts do not share this signal independence property. To compare the performance of non-Shewhart and Shewhart-type control chart, we choose the method described below in lieu of the traditional Markov Chain or integral equations approach for the MCUSUM and MEWMA threshold estimation.

2.3.1   Control chart calibration

In this work, the control chart calibration is computed by specifying a sequence of approximate thresholds for each control chart and recording the resulting in-control

average run length (ARL_0). Then, a linear regression model in the form $\ln(ARL) = a + b \times h_i$ is fitted to estimate the target threshold for a predefined $\text{ARL}_0 = 200$.
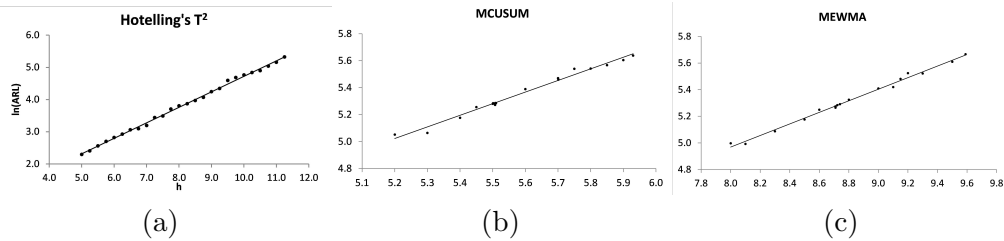
Figure 2.1: Fitted linear regressions for control chart's calibration utilising known parameters

To access the ARL_0 estimation for each threshold value, the number of samples m is set equals 2,000. This quantity of observations show high probability ($> 99.9\%$ for $\text{ARL}_0 = 200$) of triggering a false alarm when the process is actually in-control. When observed, the position N of the first alarm occurrence is recorded as the run length, and the mean value of N, computed from B Monte Carlo simulations, is the ARL_0 for that threshold. In the experiments performed with known parameters and 20,000 Monte Carlo simulations, the maximum observed run length is 1,844. To avoid missing values when the signal does not occur, m is set to a maximum of 2,000. The experiments are performed with $B = 2,000$ Monte Carlo simulations to speed up the regression adjustment step, and $B = 20,000$ are performed with the estimated in-control limit for the final ARL_0 and ARL_1 computation.

To illustrate the threshold sensitivity in the Hotelling's T^2-control chart, Table 2.1 contains the asymptotic values for a Phase II control chart, based on the F distribution, and respective standard errors obtained in 100,000 Monte Carlo simulations, where $m = 100$ observations of the in-control process were simulated each step for parameter estimation. The highest threshold ($h_{0.999} = 15.13$) corresponds to an ARL_0 value of 1,000 observations. Similarly, the confidence level of 99.5% ($h_{0.995} = 11.42$) shows an average run length that is five times less (200.7) than the first one.

Table 2.1: Confidence level, estimated thresholds, ARL_0 and standard errors for a
Phase II T^2-control chart

$(1 - \alpha)\,100\%$	h_1	ARL_0	SE
99.9	15.13	1000.1	3.127
99.5	11.42	200.7	0.625
99.0	9.85	99.7	0.316
95.0	6.30	19.9	0.063

With the asymptotic values of Table 2.2 for the Hotelling's T^2-chart and the refer-
ence values in the original papers for the MEWMA and MCUSUM charts, the linear
regressions displayed in the Figure 2.1 are fitted with known parameters for all charts'
calibrations. The minimum number of threshold values for the regression estimation is
10, and the maximum is 18. The values along the vertical axes are ARL_0 values in a
logarithmic scale, and the values along the horizontal axes are the in-control limits. The
parameter estimation is carried out with $m = 25$, 50 and 100 for the number of Phase I
samples of individual observations. The thresholds value sequences for estimating the
regression model of each control chart are set to approximate the ARL_0 between 100
and 300. All fitted regressions, goodness-of-fit and estimated thresholds for a target
$\text{ARL}_0 = 200$ are shown in Table 2.3.

Table 2.2: Asymptotic thresholds of Hotelling's T^2 Phases I and II

	Asymptotic distribution		
m	χ^2-distribution	$Beta$-distribution	F-distribution
25	-	8.81	14.61
50	-	9.69	12.34
100	-	10.14	11.42
$(\mathbf{M}_0, \boldsymbol{\Sigma}_0)$	10.60	-	-

We first compare the estimated thresholds with the asymptotic values of Hotelling's
T^2 control charts presented in Table 2.2. In Table 2.3, we notice that the linear
regression approach shows a tendency of larger thresholds for smaller sample sizes, as

expected, and the estimate for known parameters (10.61) is very close to the asymptotic value (10.60). Second, for sample estimates, the regression adjusted thresholds are close to the asymptotic average between the two distributions. For $m = 25$, the asymptotic threshold average is 11.71, while the regression estimate is 11.26. For $m = 50$ and 100, the asymptotic threshold averages are 11.02 and 10.78, respectively, while the regression estimates are 10.93 and 10.60. Thus, the linear regression approach is effective for calibrating Hotelling's T^2-control chart.

Table 2.3: Adjusted linear regression models with sample estimates and known parameters

Control Chart	m	Fitted model	r^2	$\hat{h}_i = (ARL_0)$
T^2	25	$ln\,(\text{ARL}_0) = 0.3710 * h_1 + 1.1206$	0.956	11.26
	50	$ln\,(\text{ARL}_0) = 0.4628 * h_1 + 0.2414$	0.971	10.93
	100	$ln\,(\text{ARL}_0) = 0.4954 * h_1 + 0.0463$	0.985	10.60
	$(\mathbf{M}_0, \mathbf{\Sigma}_0)$	$ln\,(\text{ARL}_0) = 0.4773 * h_1 + 0.2351$	0.992	10.61
MCUSUM	25	$ln\,(\text{ARL}_0) = 0.5138 * h_2 + 1.7467$	0.974	6.91
	50	$ln\,(\text{ARL}_0) = 0.6589 * h_2 + 1.1829$	0.993	6.25
	100	$ln\,(\text{ARL}_0) = 0.7343 * h_2 + 0.9318$	0.997	5.95
	$(\mathbf{M}_0, \mathbf{\Sigma}_0)$	$ln\,(\text{ARL}_0) = 0.8611 * h_2 + 0.5451$	0.986	5.52
MEWMA	25	$ln\,(\text{ARL}_0) = 0.2811 * h_3 + 2.0399$	0.963	11.59
	50	$ln\,(\text{ARL}_0) = 0.3628 * h_3 + 1.5648$	0.974	10.29
	100	$ln\,(\text{ARL}_0) = 0.3892 * h_3 + 1.5511$	0.994	9.63
	$(\mathbf{M}_0, \mathbf{\Sigma}_0)$	$ln\,(\text{ARL}_0) = 0.4348 * h_3 + 1.4904$	0.993	8.76

2.3.2   In-control and out-of-control run length distributions

To compare the run length distributions calibrated with known parameters, Figure 2.2 presents a sequence of density histograms based on the resulting run length values for 20,000 Monte Carlo simulations. Figure 2.2(a) shows the run length distributions for the three control charts and the first moment of each distribution are the ARL_0 values of the last group in Table 2.4, which is based on the known parameters $(\mathbf{M}_0, \mathbf{\Sigma}_0)$.

Some large observations are not displayed in the histograms because the horizontal axes were set to not lose scale definition. When the shift is large as in Figure 2.3(c),

(a) $d = 0$

(b) $d = 0.71$

(c) $d = 1.41$
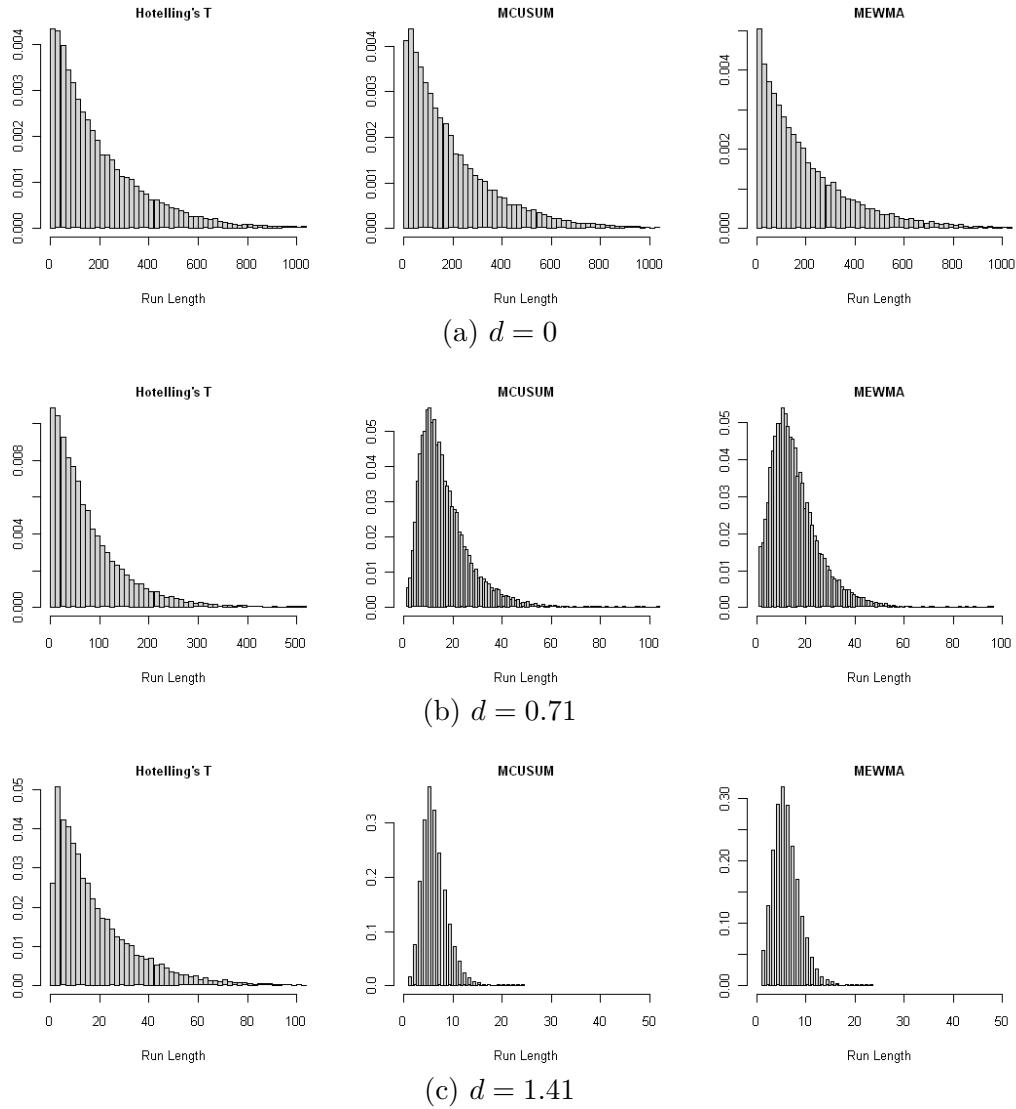
Figure 2.2: Empirical in- and out-of-control run length distributions

the Hotelling's T^2 is clearly the most favourable to change detection. The run length distribution of the MEWMA control chart as seen Figure 2.2(b)-(c) confirm that this control chart with $\lambda = 0.1$ tends to perform better than the MCUSUM with $k = 0.5$. This result is also confirmed by the ARL_1 values produced in the experiments with known parameters of Table 2.4.

It is observed that simulating long runs is not necessary for defining the out-of-control

(a) $d = 2.12$

(b) $d = 2.82$

(c) $d = 3.54$
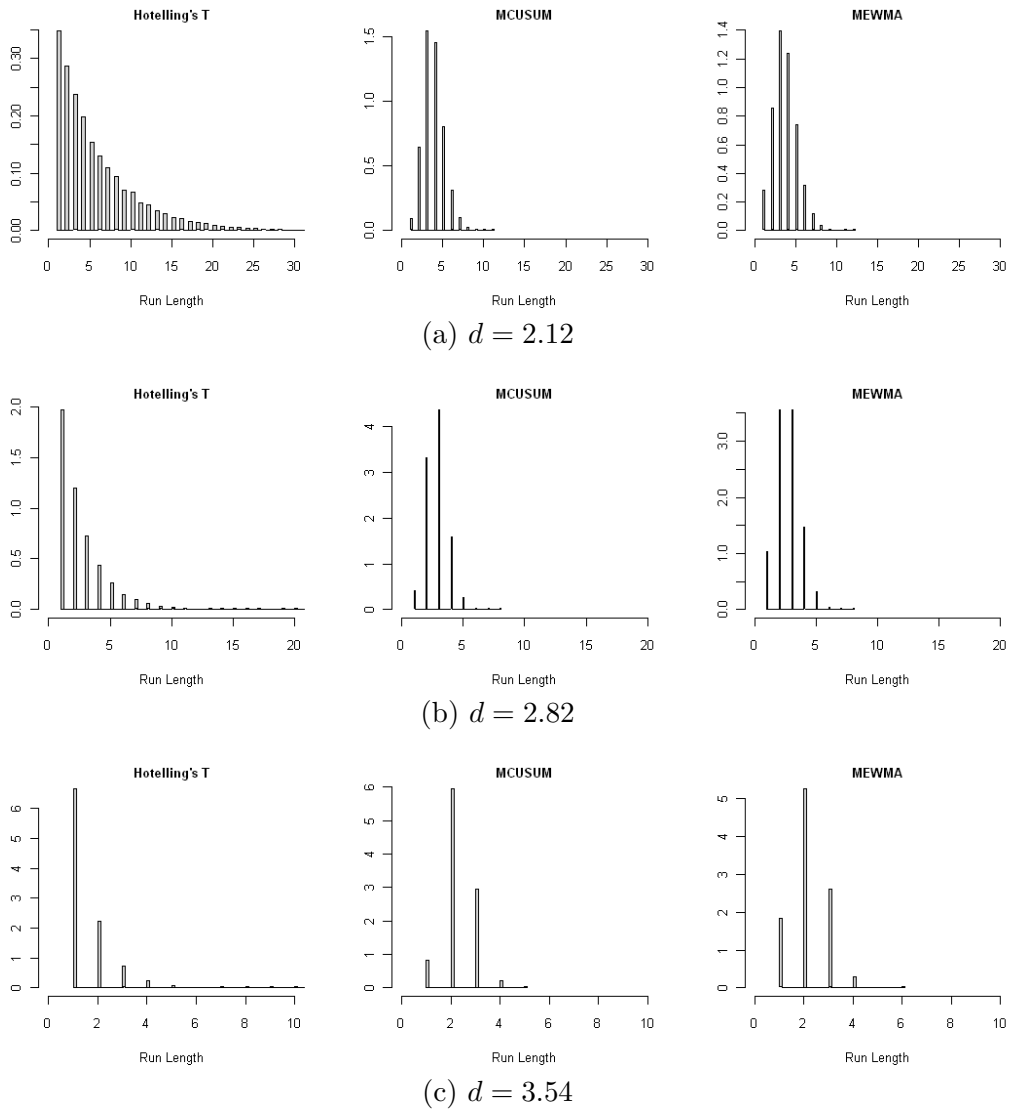
Figure 2.3: Empirical in- and out-of-control run length distributions

run length distributions. For the ARL_1 estimation, the total number of simulated sample
observations (m) utilised for each mean vector shift was 1,500, 1,000, 500, 200 and 100,
which considerably reduced the computing time costs.

2.3.3   Performance comparison for mean vector shifts

Excluding the specific mean vector shifts to test the MCUSUM method, Crosier (1988) evaluates d unitary increments in the 0 to 5 range for several dimensions. Lowry and Montgomery (1995) in the MEWMA method evaluate shifts through d increments in the 0 to 3 range. The present work simulates changes in equal increments for both dimensions, and the centre of the process is shifted diagonally from (0.0, 0.0) to (2.5, 2.5) by increments of 0.5 in both dimensions, representing d values in the 0 to 3.54 range. The probability inside the area defined by three standard deviations in a non-correlated process is 0.995.

Figure 2.4 illustrates the proposed mean vector shifts for correlated and non-correlated processes. The non-correlated scheme in Figure 2.4 is calibrated with estimated and known parameters. The correlated variable schemes of Figure 2.4 are compared only for the case of known parameters. The experienced reader should argue why to do comparisons with correlated variables since the charts performance does not depends on the shifts direction. For that reasoning, the authors intend to emphasize that a shift of magnitude d is different when it occurs in the process directions of major and minor variability.
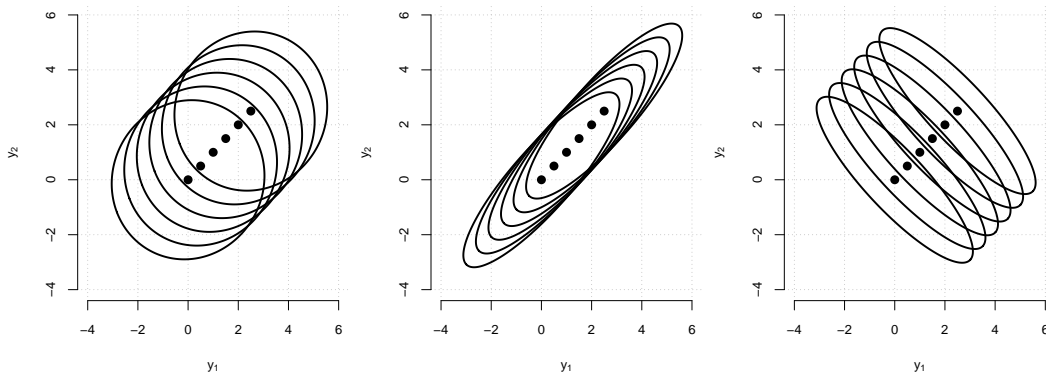
Figure 2.4: Diagram of the proposed mean vector shifts for non-correlated and correlated processes

For non-correlated processes shown in Figure 2.4, the in-control limits are delimited with known parameters and sample estimates of three different sizes ($m = 25; 50; 100$) in Phase I. Table 2.4 presents the resulting in- and out-of-control ARLs. Although the standard error (SE) of the estimate is relatively large for the ARL_0 due to the regular number of Monte Carlo simulations ($B = 20,000$), the ARL_1 values show little variation.

The main observed effect due to the utilisation of small sample sizes to train the control chart in Table 2.4 is an increasing delay in the change detection as the sample size decreases. The MCUSUM and MEWMA charts perform very similarly for all shifts, with faster performance than the Hotelling's T^2-chart for the shifts (2.0, 2.0) that are situated around the target noncentrality value ($d = 1$). To this point, the inertia effect begins to delay the change detection for the non-Shewhart charts, and the shift for the point (2.5, 2.5) is more likely to be detected with the Shewhart-type control chart.

Table 2.4: Average run length for mean vector shifts with sample estimates and known parameters

	d	0	0.71	1.41	2.12	2.82	3.54
m	Chart	ARL_0	ARL_1	ARL_1	ARL_1	ARL_1	ARL_1
	MCUSUM	199.4	26.3	7.3	4.5	3.3	2.6
25	MEWMA	201.1	28.1	7.8	4.7	3.4	2.7
	T^2	199.3	91.7	26.5	7.7	3.1	1.7
	MCUSUM	197.3	20.2	6.8	4.2	3.1	2.5
50	MEWMA	199.1	20.9	7.1	4.4	3.3	2.6
	T^2	199.9	87.3	21.9	6.6	2.8	1.6
	MCUSUM	203.1	18.2	6.4	4.0	3.0	2.4
100	MEWMA	200.3	18.4	6.8	4.2	3.1	2.5
	T^2	200.6	84.2	20.5	6.3	2.7	1.5
	MCUSUM	198.6	16.0	6.0	3.7	2.8	2.3
$(\mathbf{M}_0, \mathbf{\Sigma}_0)$	MEWMA	197.2	15.2	5.8	3.6	2.7	2.1
	T^2	201.0	76.9	18.6	5.7	2.5	1.5

The simulations with modified correlation structures shown in Figure 2.4 determine

the changes only in the first quadrant, thus allowing changes to represent both large
and small shifts respective to the noncentrality parameter. Figure 2.5 demonstrates how
much those different structures allow the same shift in the mean vector to represent
a relatively different shift in the noncentrality parameter. Compared with the non-
correlated process represented by the solid line, the mean vector shifts measured by the
noncentrality parameter are smaller in the positively correlated process (gray line) and
larger in the negatively correlated process (dashed line).
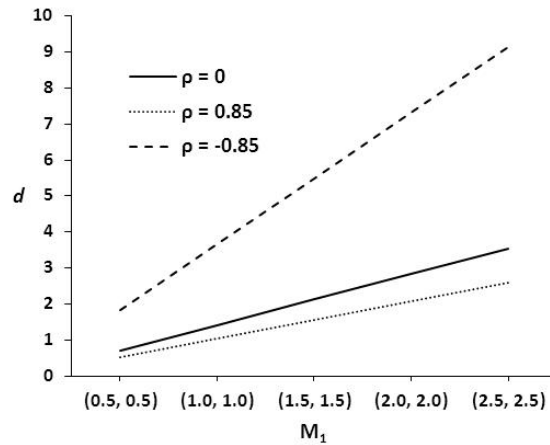
Figure 2.5: Shift size with respect to the noncentrality parameter in the correlated and
non-correlated processes

Throughout the rest of the experiments, the in-control limits are obtained by cal-
ibrating the control chart with the process known parameters. The results for ARL
comparison in detecting the process shift schemes of Figure 2.4 are shown in Table 2.5
and represented in d units in Figure 2.6. Although the MCUSUM and MEWMA charts
have equivalent performance in the three situations, the performance of the Hotelling's
T^2-chart exceeds the others after the first shift in Figure 2.6(c), when the shift occurs in
the minor axis of the negatively correlated process. This result does not contradict the
finding that the Hotelling's T^2-chart is a better monitor of large shifts. This situation
occurs because a large distance between the processes tends to result in inertia in the

MCUSUM and MEWMA methods (Lowry et al., 1992), delay change detection and make the Hotelling's T^2-chart more efficient in the third shifting scheme. We observe that larger shifts often occur in the direction of larger variability. A more realistic suggestion in this case is to observe relatively less movement in the direction of minor variability. The large values in the direction of lower variability are usually related to typing errors in the data acquisition procedure (Montgomery, 2001). An observation of the point where the charts intersect indicates that the relative magnitude of the changes in all three cases is the same when measured using the noncentrality parameter ($d \cong 2.6$).

Table 2.5: Average run length for mean vector shifts with known parameters and correlated processes

	d	0	0.71	1.41	2.12	2.82	3.54
ρ	Chart	ARL_0	ARL_1	ARL_1	ARL_1	ARL_1	ARL
	MCUSUM	200.8	26.8	9.0	5.3	3.8	3.0
0.85	MEWMA	202.0	25.0	8.7	5.1	3.7	2.9
	T^2	203.9	112.8	38.9	14.3	6.2	3.2
	MCUSUM	199.7	4.4	2.2	1.6	1.2	1.0
-0.85	MEWMA	201.1	4.3	2.1	1.5	1.2	1.0
	T^2	201.5	9.0	1.4	1.0	1.0	1.0

(a) $\rho = 0$ (b) $\rho = +0.85$ (c) $\rho = -0.85$
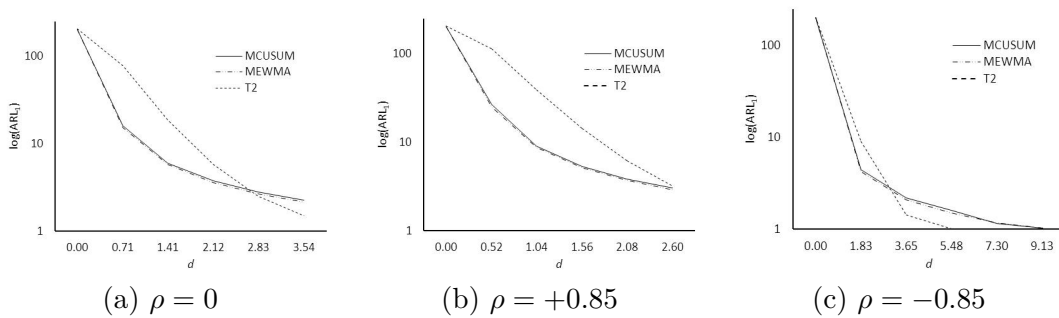
Figure 2.6: The control charts' performance for the three simulated processes and mean vector shifts as noncentrality values

Figure 2.7 illustrates three simulated runs for mean vector shifts in the process.

The changes in the processes are moderate and large shifts. The inertia effect can be viewed in Figure 2.7(b) with the non-Shewhart control charts. As the Hotelling's T^2-chart triggers a signal at the very first out-of-control observation, the MCUSUM and MEWMA charts are affected by a delay in triggering the signal.

(a) $d = 1.41$, $\mathbf{M}_1 = (1.0, 1.0)$ (b) $d = 3.54$, $\mathbf{M}_1 = (2.5, 2.5)$
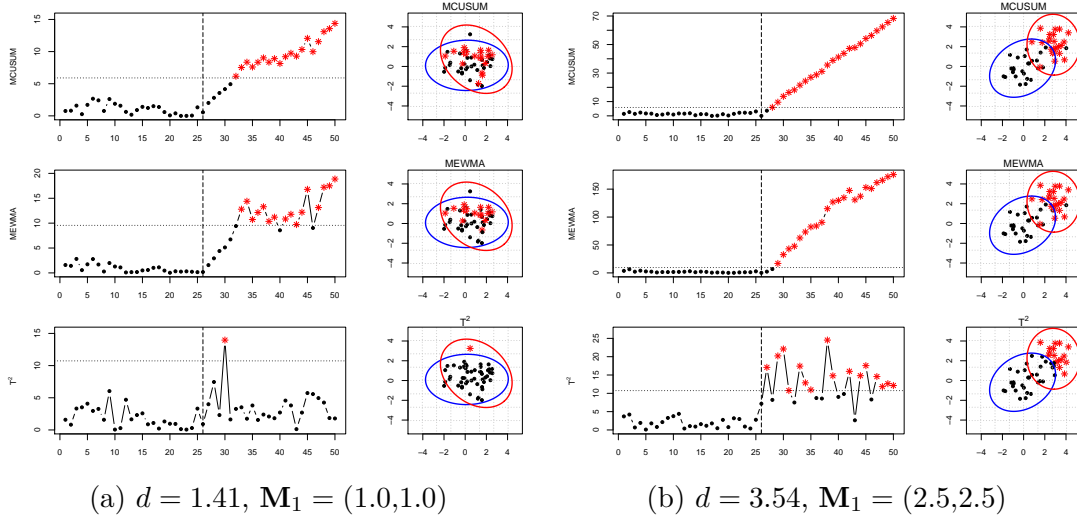
Figure 2.7: Control chart patterns for moderate and large mean vector shifts

2.3.4   The influence of increasing variances

If the variances of the underlying process changes from (σ_1^2, σ_2^2) to $(\sigma_1^2 + \Delta, \sigma_2^2 + \Delta) = (\sigma_1^2, \sigma_2^2)^*$, control chart may be affected because this overdispersion increases the individual observation distances with respect to the center of the in-control process. The results in Table 2.6 reflect the effect of process variance increases, which demonstrates a larger influence on the Hotelling's T^2-control chart.

The patterns that result from simultaneously altering the mean vector and covariance matrix generate a greater tendency for signal growth than the pattern shown in Figure 2.8(a). When the mean vector changes, the subsequent process changes stemming from the modification of the covariance matrix greatly influence the charts' performance. Whereas the MCUSUM and MEWMA methods consider the entire pro-

Table 2.6: The influence of increasing the process variances on the average run length

$(\sigma_1^2, \sigma_2^2)^*$	(1.0, 1.0)	(1.5, 1.5)	(2.0, 2.0)	(2.5, 2.5)
Control chart	ARL_0	ARL_1	ARL_1	ARL_1
MCUSUM	198.1	52.3	26.8	17.9
MEWMA	202.8	56.3	29.0	18.9
T^2	202.3	35.0	14.2	8.4

cess to be out-of-control, beginning with the characterisation of a new centre of gravity,
the Hotelling's T^2-chart discriminates all of the observations that exceed the coverage
area, as defined by the in-control process.

(a) $\mathbf{M}_1 = (1.0, 1.0)$, $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (b) $\mathbf{M}_1 = (1.0, 1.0)$, $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
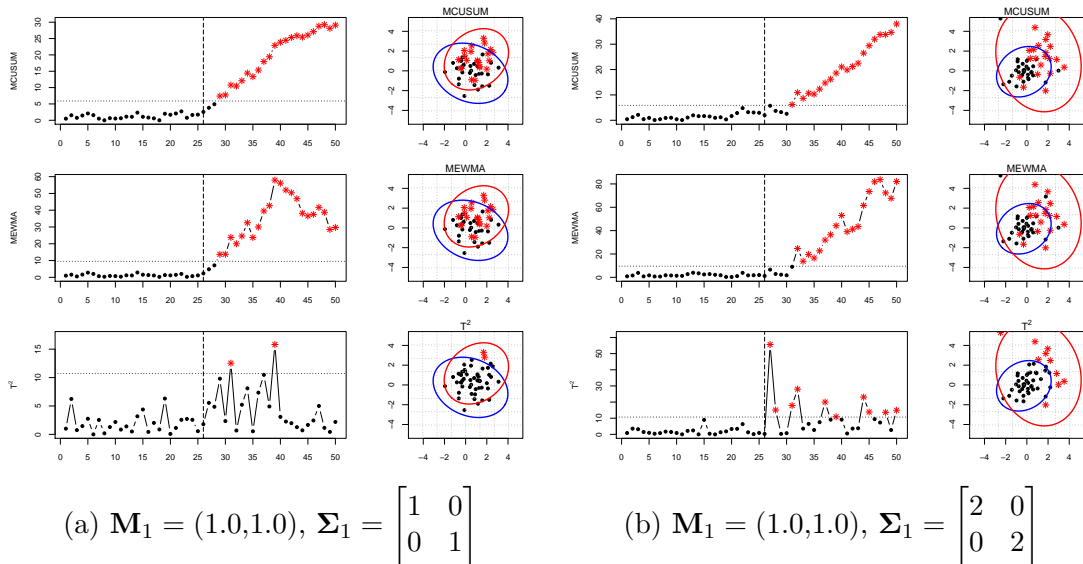
Figure 2.8: Control chart patterns for mean vector shifts and increased variances

The difference in the charts' performance can be observed by comparing Table 2.6
and 2.7. Table 2.7 shows the influence of variance increasing over the mean vector
shifts. This experiment performs only a slight increase in the variances, which changes
from $(\sigma_1^2, \sigma_2^2) = (1.0, 1.0)$ to $(\sigma_1^2, \sigma_2^2)^* = (1.5, 1.5)$. The combined effect of the increased
variances and mean vector shifts is more evident for the Hotelling's T^2 control chart,
where the reductions in the ARL_1 are expressive. The MCUSUM and MEWMA control
charts demonstrate a significant difference in the ARL_1 only for the first shift. The

ARL_1 reduction for the non-Shewhart charts in the second shift is less significant for bigger shifts.

Table 2.7: The ARL influence of simultaneously increasing the variances and shifting the mean vector

	Chart	MCUSUM	MEWMA	T^2
$(\sigma_1^2, \sigma_2^2)^*$	d		ARL_1	
(1.0,1.0)	0.71	16.0	15.2	76.9
	1.41	6.0	5.8	18.6
	2.12	3.7	3.6	5.7
(1.5,1.5)	0.71	13.5	13.1	21.4
	1.41	5.8	5.6	8.9
	2.12	3.7	3.6	4.1
(2.0,2.0)	0.71	11.5	11.3	10.7
	1.41	5.6	5.4	5.9
	2.12	3.7	3.5	3.3

2.3.5   The influence of autocorrelation

To verify the control charts' performance when the process is perturbed with autocorrelation, the out-of-control processes are simulated based on a first-order autoregressive model, VAR(1), on a scale of increasing intensity. The autocorrelations in both dimensions are $\Phi = (\phi_1, \phi_2) = (0.0,0.0)$ for the in-control process and $\Phi = (\phi_1 + \delta, \phi_1 + \delta) = (\delta, \delta)$ for the out-of-control process. The autocorrelation levels are $\delta = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$. The degree of influence on the processes is indirectly evaluated by comparing the ARL_1 values produced in the experiments described in Table 2.8 and 2.9 to the values presented in Table 2.4, where only the mean vector was modified.

The experiment in Table 2.8 demonstrates that pure autocorrelation in the out-of-control process results in small mean vector shifts, being less noticed by the Hotelling's T^2-chart. The experiment in Table 2.9 inserts low autocorrelation levels in the process

Table 2.8: The ARL influence of purely increasing autocorrelation levels

Φ	(0.0, 0.0)	(0.1, 0.1)	(0.2, 0.2)	(0.3, 0.3)	(0.4, 0.4)
Chart	ARL_0	ARL_1	ARL_1	ARL_1	ARL_1
MCUSUM	198.5	105.3	60.1	36.1	23.9
MEWMA	198.2	111.1	64.7	39.8	25.9
T^2	200.8	194.0	163.6	130.4	92.2
Φ	(0.5, 0.5)	(0.6, 0.6)	(0.7, 0.7)	(0.8, 0.8)	(0.9, 0.9)
Chart	ARL_1	ARL_1	ARL_1	ARL_1	ARL_1
MCUSUM	16.4	11.7	8.5	6.2	4.6
MEWMA	17.4	12.2	8.7	6.2	4.5
T^2	60.5	37.7	21.8	12.8	7.1

Table 2.9: The ARL influence of simultaneously increasing variances and shifting the
mean vector

Φ	(0.0,0.0)			(0.1,0.1)		
$\mathbf{M}_1$	(0.0,0.0)	(0.5,0.5)	(1.0,1.0)	(1.5,1.5)	(2.0,2.0)	(2.5,2.5)
Chart	ARL_0	ARL_1	ARL_1	ARL_1	ARL_1	ARL_1
MCUSUM	201.0	13.2	5.3	3.4	2.5	2.1
MEWMA	201.1	12.8	5.1	3.2	2.4	2.0
T^2	203.1	64.2	14.2	4.4	2.0	1.3
Φ	(0.0,0.0)			(0.2,0.2)		
$\mathbf{M}_1$	(0.0,0.0)	(0.5,0.5)	(1.0,1.0)	(1.5,1.5)	(2.0,2.0)	(2.5,2.5)
Chart	ARL_0	ARL_1	ARL_1	ARL_1	ARL_1	ARL_1
MCUSUM	201.0	10.8	4.6	3.0	2.8	2.3
MEWMA	201.1	10.6	4.5	2.9	2.7	2.2
T^2	202.1	48.4	10.2	3.2	2.2	1.5
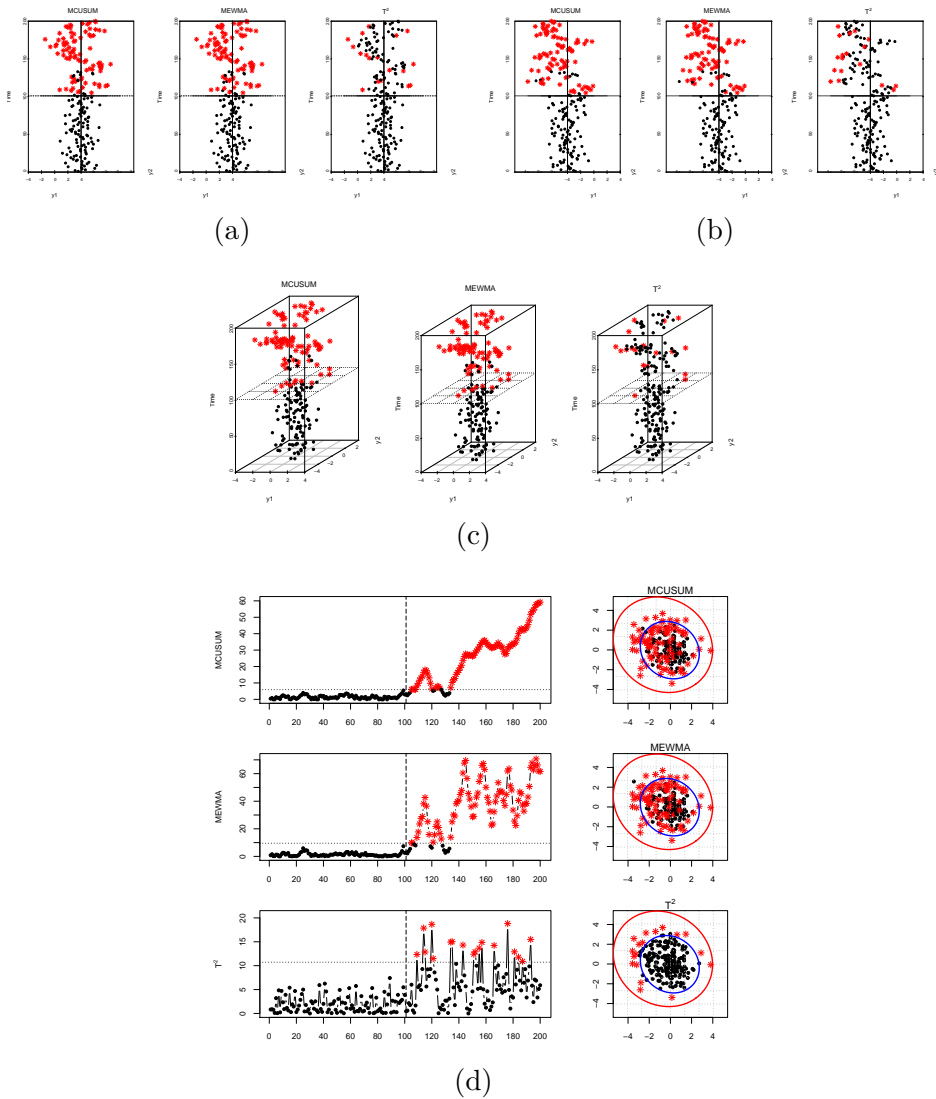
Figure 2.9: Three-dimensional scatter plots and control charts with confidence ellipses for a purely autocorrelated out-of-control process $\Phi = (0.8, 0.8)$

and simultaneously shifts the mean vector. The combined effect of mean vector shifts and autocorrelation shows a reduction in the ARL_1 primarily for small shifts with all control charts and an acceleration of the change detection as the autocorrelation level increases. Further, the increase in the autocorrelation with mean vector shifts demonstrates a bigger influence on the Hotelling's T^2-chart. Comparing with simply

(a)

(b)

(c)

(d)
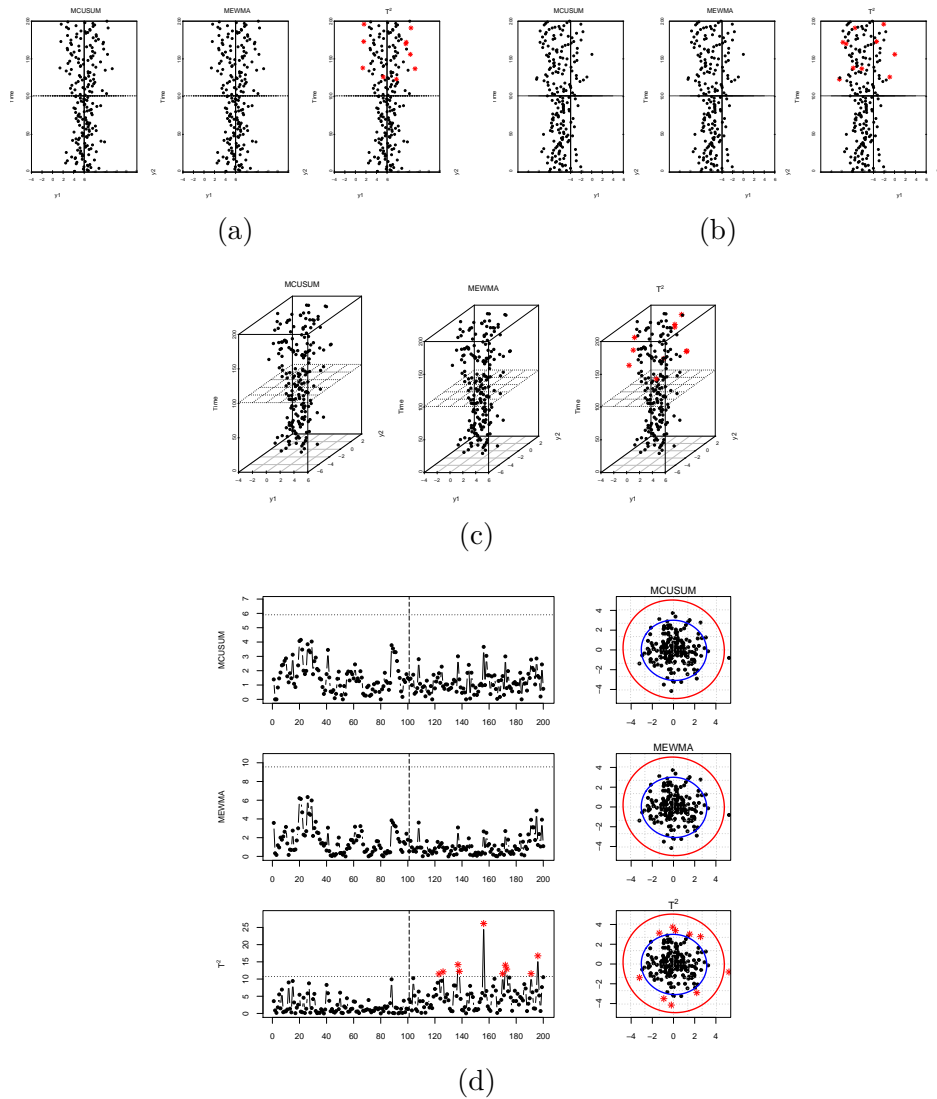
Figure 2.10: Three-dimensional scatter plots and control charts with confidence ellipses
for the negative autocorrelated process $\Phi = (-0.8, -0.8)$

shifting the mean vector to $\mathbf{M}_1 = (0.5,0.5)$, the experiments utilising $\Phi = (0.1,0.1)$
and $\mathbf{M}_1 = (0.5,0.5)$ accelerate the Hotelling's T^2-chart to identify the first two smaller
shifts by 17.7% and 23%, while the MCUSUM and MEWMA charts are accelerated by
18% and 16.3%, respectively. When $\Phi = (0.2,0.2)$, the acceleration in change detection
are also more evident for the small shifts in the mean vector.

To visualise the effect of pure autocorrelation in the out-of-control process, Figure 2.9 illustrates the action pattern of the control charts when $\Phi = (0.8,0.8)$. The non-Shewhart control charts are more likely to detect the change of an oscillatory movement in the process, which is characteristic of positive autocorrelation. Observing the three-dimensional scatter plots of Figure 2.9, this variation source initially perform a small shift in the process mean vector.

In all cases, the distances measured using the noncentrality parameter suggest that the effects resulting from purely autocorrelated processes favour detection by the MCUSUM and MEWMA charts. The individual observation distances do not represent significant shifts relative to the in-control process. In situations when a shift in the mean vector occurs with a regular autocorrelation level, the Hotelling's T^2-chart can be more effective to change detection.

The MCUSUM and MEWMA charts are generally considered effective for detecting shifts that are caused by the occurrence of purely positive autocorrelation at regular and high levels; for negative autocorrelation, the Hotelling's T^2-chart is most suitable to perceive the changes as outliers in the process.

Figure 2.10 shows the negative autocorrelated out-of-control process pattern results, which are only detected using the Hotelling's T^2-chart, especially if the magnitude of the autocorrelation is large and the outliers appear beyond the in-control region. Conversely, the MCUSUM and MEWMA statistics result in no sensitivity to the negative autocorrelated process because the alternating individual observations cancel each other.

2.4   Discussion

The present study investigates the behavioural patterns and performance of the control charts widely applied in SPC to monitor the mean vector, i.e., the MCUSUM, MEWMA

and Hotelling's T^2-charts, with individual observations and combined sources of variability. The performance comparison of the charts with controlled simulated changes is carried out by the estimated ARLs. A procedure for the confidence ellipse estimation based on principal component analysis is briefly described as a descriptive tool for visualisation purposes.

As a unifying approach to estimate the in-control limits of Shewhart- and non-Shewhart-type control charts, the linear regressions of the thresholds as a function of the average run length are adjusted for known parameters and sample estimates of different sizes. The estimated thresholds for the three selected control charts agree with the values described in the literature. As noted above, an estimation of the control charts thresholds with a small number of individual Phase I observations increases the in-control limits and delays change detection for all studied control charts.

Different scenarios for the correlated process are established to compare the mean vector shifts in the directions of larger and smaller variations. These experiments demonstrate how the performance of selected control charts is dependent only on the shift magnitude, as measured by the noncentrality parameter.

The estimated in-control limits for known parameters are applied to study the effects of combining mean vector shifts and increasing variances or serial autocorrelations in the out-of-control process. We demonstrate that the resulting effect on the charts' performance following an increase only in the variances in the out-of-control process is more evident in the Hotelling's T^2-chart, even for relatively small variance increments. The combined effect of jointly shifting the mean vector and increasing the variances in low levels makes the Hotelling's T^2-control chart as competitive as the non-Shewhart charts, even for moderate shift sizes in the mean vector because the influence of increasing the process dispersion additionally increases the individual noncentrality values.

The effect of serial autocorrelation simulated in the out-of-control process follows a first-order autoregressive model at different intensity levels. Among all proposed

increments in the autocorrelation, the experiments in this work showed that the effects of purely autocorrelated processes can be identified as small shifts in the mean vector, with correspondingly better change detection by the MCUSUM and MEWMA charts. When combined with the mean vector shifts, low levels of positive autocorrelation demonstrate that the Hotelling's T^2-control chart is as competitive as or better than the non-Shewhart charts for moderate and large shifts in the mean vector.

The control charts' performance in the presence of negative autocorrelation shows that only the Hotelling's T^2-chart recognises the effects as outliers due to their continuous process dispersion. Whereas the MCUSUM method voids the information about the shift when the observations produce statistics with alternating signs in the cumulative sums, the MEWMA chart becomes ineffective by weighting the current value with the historical average, which is also affected by continuously alternating observations.

The noncentrality parameter is widely applied in single-hypothesis problems to control the mean vectors in multivariate processes because the analysis of all possible alternative distributions prior to decision-making is impossible. This technique works well when the observation size is small, e.g., 1 or 2. When p increases, however, mapping the p-dimensional space for a one-dimensional distance may destroy valuable information for the correct classifications that exist in the original spaces, if the direction of the shift is unknown a priori. A careful search of the conditional discriminant functions that determine an ideal intersection between the processes can reveal an optimal procedure.

As previously noted, the literature on statistical pattern recognition proposes several dissimilarity measures to improve class separability. Measures specifically delineated for multivariate normal distributions, such as appropriate linear transformations of the Mahalanobis distance, show superior performance in ARL when compared with the charts in the present study, on which topic work will soon be published.

Chapter 3

Confidence Control Charts with MEWMA and Sliding Window Schemes

Abstract

In this paper we show that for normal distributions, Hotelling's T^2 and multivariate exponentially weighted moving average (MEWMA) statistics are directly related to the Chernoff distance. This relationship provides important information about an upper bound for the misclassification probability, indicating the degree of overlap between in- and out-of-control processes. Therefore, the purpose of this research is to present a methodology to monitor mean vectors of Gaussian process by means of an informative control chart based on the 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 MEWMA control chart is easier to calibrate and shows less inertia for large shifts than the sliding window approach. Equivalences

between the smoothing parameter and the window size are provided for a bivariate case.

3.1 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 cases (Niaki and Memar, 2009) covering the 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). With respect to the Hotelling's statistics, Machado and Costa (2009) comments that its drawback is the difficulty to interpret the obtained values and proposed a method to monitor bivariate process based on the ZMAX and VMAX statistics. Quinino et al. (2012) also propose a single statistic based on the mixture of variances (VMIX) to monitor the covariance matrix of bivariate processes. In the case of smaller shifts, the multivariate exponentially weighted moving average (MEWMA) control chart is preferred chart (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, 1988). 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 (Hwarng and Hubele, 1993b,a; Guh and Shiue, 2005) have suggested a sliding 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 Shiue, 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 (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 3.2 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 3.3, 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 3.4 provides some final remarks and recommendations.

3.2 Methodology

3.2.1 A review on the Hotelling's T^2 and MEWMA control charts

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

$$d_t^2 = (\mathbf{X}_t - \mathbf{M}_0)' \, \boldsymbol{\Sigma}_0^{-1} \, (\mathbf{X}_t - \mathbf{M}_0) \tag{3.1}$$

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$_0$).

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 $z_t^2 > h_2$, with

$$z_t^2 = (\mathbf{M}_t - \mathbf{M}_0)' \, \boldsymbol{\Sigma}_0^{-1} \, (\mathbf{M}_t - \mathbf{M}_0). \tag{3.2}$$

In Equation (3.2), $\mathbf{M}_t$ is the mean vector estimated with past and current information by a MEWMA scheme, such that

$$\mathbf{M}_t = (1 - \lambda) \, \mathbf{M}_{t-1} + \lambda \mathbf{Y}_t. \tag{3.3}$$

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.

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

$$d^2 = (\mathbf{X} - \mathbf{M})' \, \boldsymbol{\Sigma}^{-1} \, (\mathbf{X} - \mathbf{M}) = \mathbf{Z}'\mathbf{Z} = \sum_{i=1}^{p} z_i^2 \qquad (3.4)$$

where $\mathbf{Z} = \mathbf{A}' (\mathbf{X} - \mathbf{M})$ 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 $\mathbf{z}_i$'s are uncorrelated, with $E(z_i) = 0$ and $\mathrm{Var}(z_i) = 1$. Thus, the expected value and variance of d^2 for the IC process are

$$E\left(d^2 | \mathrm{IC}\right) = pE\left(z_i^2\right) = p, \qquad (3.5)$$

and

$$\mathrm{Var}\left(d^2 | \mathrm{IC}\right) = E\left(\left(d^2\right)^2\right) - E^2\left(d^2\right) = \sum_{i=1}^{p} E\left(z_i^4\right) + \sum_{i=1}^{p}\sum_{j \neq i}^{p} E\left(z_i^2 z_j^2\right) - p^2 E^2\left(z_i^2\right). \qquad (3.6)$$

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

$$Var\left(d^2 | \mathrm{IC}\right) = p\gamma \qquad (3.7)$$

$$\gamma = E\left(z_i^4\right) - E^2\left(z_i^2\right) = E\left(z_i^4\right) - 1 \tag{3.8}$$

For normal distributions, when the z_i's are uncorrelated, they are also independent. Therefore, Equation (3.8) can be used to compute $\mathrm{Var}\left(d^2|\mathrm{IC}\right)$, and $\gamma = 2$. Note that in Equations (3.5) and (3.7), 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 $\alpha = 1/2$ and $\beta = p/2 - 1$. 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 $p = 2$.

Indeed, the distribution of d^2 with the mean n and standard deviation $\sqrt{2p}$ approximates the normal distribution when p 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

$$E\left(d^2|\mathrm{OC}\right) = p + \mathbf{M}_1'\mathbf{M}_1, \tag{3.9}$$

and the variance is given as

$$\mathrm{Var}\left(d^2|\mathrm{OC}\right) = p + 4\mathbf{M}_1'\mathbf{M}_1. \tag{3.10}$$

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

$$\zeta = \frac{1}{m-1}\left(\mathbf{X} - \hat{\mathbf{M}}\right)' \hat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{X} - \hat{\mathbf{M}}\right). \tag{3.11}$$

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

3.2.2   Upper bounds on the error probability

The noncentrality parameter, also known as Mahalanobis distance in the pattern recognition field (Therrien, 1989), has a close connection to the Bhattacharyya (1943) distance, which is derived from the most general case, the Chernoff (1952) distance. Those boundaries lead to a closed-form expression for computing an upper limit for the Bayes error in the case of normally distributed processes. The following is a brief discussion on the upper bounds of the error probability.

Let $\mathbf{X}$ be an observation vector, and let be our purpose to determine whether $\mathbf{X}$ belongs to ω_1 or ω_2. A decision rule based simply on probabilities may be written as follows:

$$q_1\left(\mathbf{X}\right) > q_2\left(\mathbf{X}\right) \to \mathbf{X} \in \omega_1, \tag{3.12}$$

where $q_i\left(X\right)$ is a posteriori probability of ω_i given $\mathbf{X}, i = 1,2$.

In general, a decision rule does not lead to perfect classification and is convenient to evaluate the performance of a decision rule. The conditional error $r\left(X\right)$, given $\mathbf{X}$, due to the decision rule (3.12) is either $q_1\left(\mathbf{X}\right)$ or $q_2\left(\mathbf{X}\right)$ which ever smaller. That is,

$$r\left(\mathbf{X}\right) = \min\left[q_1\left(\mathbf{X}\right), q_2\left(\mathbf{X}\right)\right]. \tag{3.13}$$

The total error, which is called the Bayes error (ϵ), is computed by

$$\epsilon = E\left(r\left(\mathbf{X}\right)\right) = \int r\left(\mathbf{X}\right) f\left(\mathbf{X}\right) d\mathbf{X} = \int min\left[P_1 f_1\left(\mathbf{X}\right), P_2 f_2\left(\mathbf{X}\right)\right] d\mathbf{X} =$$
$$P_1 \int_{L_2} f_1\left(\mathbf{X}\right) d\mathbf{X} + P_1 \int_{L_1} f_2\left(\mathbf{X}\right) d\mathbf{X} = P_1 \epsilon_1 + P_2 \epsilon_2, \tag{3.14}$$

where P_i and $f_i, i = 1,2$ are the weights and density functions of each class, respec-

tively.

If the classes are well separable the error is near to zero, and increases when they are more overlapped. It is evident that the calculations of the error probability in high-dimensional spaces are a difficult task. Even when the observation vectors have a normal distribution, we must resort to numerical techniques. If we cannot obtain a closed-form expression for the error probability, we may take some other approach. We may seek either an approximate expression for the error probability, or an upper bound on the error probability. In this section, we will discuss some upper bounds of error probability.

An upper bound of the integrand may be by making use of

$$\min [a,b] \leq a^s b^{1-s}, 0 \leq s \leq 1. \tag{3.15}$$

Using the inequality of Equation (3.15), ϵ can be bounded by

$$\epsilon_u = P_1^s P_2^{1-s} \int f_1^s (\mathbf{X}) f_2^{1-s} (\mathbf{X}) \, dX, 0 \leq s \leq 1, \tag{3.16}$$

where ϵ_u indicates an upper bound of ϵ. This ϵ_u is called Chernoff (1952) bound. The optimum s can be easely obtained minimizing ϵ_u. When two density functions are normal as $MVN (\mathbf{M}_1, \mathbf{\Sigma}_1)$ and $MVN (\mathbf{M}_2, \mathbf{\Sigma}_2)$ the integration can be carried out to obtain a closed expression,

$$\epsilon_u = \int f_1^s (\mathbf{X}) f_2^{1-s} (\mathbf{X}) \, dX = e^{-\mu(s)}, \tag{3.17}$$

where $\mu (s)$ is the Chernoff distance given by,

$$\mu\left(s\right) = \frac{s\left(1-s\right)}{2}\left(\mathbf{M}_2 - \mathbf{M}_1\right)'\left[s\boldsymbol{\Sigma}_1 + \left(1-s\right)\boldsymbol{\Sigma}_2\right]^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right) + $$
$$\frac{1}{2}ln\left[\frac{\left|s\boldsymbol{\Sigma}_1 + \left(1-s\right)\boldsymbol{\Sigma}_2\right|}{\left(|\boldsymbol{\Sigma}_1|^s|\boldsymbol{\Sigma}_2|^{1-s}\right)}\right]. \quad (3.18)$$

For this case, the optimum s can be obtained by plotting $\mu\left(s\right)$ for various s given $\mathbf{M}_i$ and $\boldsymbol{\Sigma}_i$. The optimum s is the one which gives the maximum value for $\mu\left(s\right)$. If we do not insist on the optimum selection of s, we may obtain a less complicated bound. One of the possibilities is to select $s = 1/2$. Thus, the upper bound is given by

$$\epsilon = \sqrt{P_1 P_2}\int\sqrt{f_1\left(X\right)f_1\left(X\right)}dX = \sqrt{P_1 P_2}e^{-\mu\left(\frac{1}{2}\right)} \quad (3.19)$$

where,

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8}\left(\mathbf{M}_2 - \mathbf{M}_1\right)'\left[\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right]^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right) + \frac{1}{2}ln\frac{\left|\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right|}{\sqrt{|\boldsymbol{\Sigma}_1|\,|\boldsymbol{\Sigma}_2|}}. \quad (3.20)$$

The term $\mu\left(\frac{1}{2}\right)$ is known as the Bhattacharyya distance and is used as an important separability measure between two normal distributions. 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 (1949) 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 consider the upper bound on the Bayes error of Equation (3.21), 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

$$\epsilon_I = \sqrt{P_1/P_2} \int \sqrt{f_1\left(X\right)f_1\left(X\right)}dX = \sqrt{P_1/P_2}e^{-\mu\left(\frac{1}{2}\right)}. \tag{3.21}$$

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

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8}\left(\mathbf{M}_t - \mathbf{M}_0\right)'\mathbf{\Sigma}_0^{-1}\left(\mathbf{M}_t - \mathbf{M}_0\right) \tag{3.22}$$

where $\mathbf{M}_t$ is the mean vector estimated at the instant t, $\mathbf{M}_0$ is the in-control mean vector, and $\mathbf{\Sigma}_0$ 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.

In his original paper, 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.

In the transition phase, d is varied 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 the next section.

The first and second order moments for Bhattacharyya distance for equal covariance matrices are easily deduced from the results of Section 3.2.1. 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.

3.2.3   The confidence control charts

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 3.21, thus reducing the upper bound on the type I error to $\exp\left(-\mu\left(\frac{1}{2}\right)\right)$. 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), triggering a signal when

$$p_t = 1 - \exp\left[-\frac{1}{8}\left(\mathbf{X}_t - \mathbf{M}_0\right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{X}_t - \mathbf{M}_0\right)\right] > h_1^*, \tag{3.23}$$

where h_1^* is the threshold to achieve a desired ARL_0.

3.2.4   Confidence control chart with MEWMA scheme

If the individual observed vector is changed by a mean vector utilising past data information, it is possible to utilise the MEWMA or the SW schemes for estimation of the actual mean vector. Equation (3.3) is utilised to estimate $\mathbf{M}_t$ in the case of an MEWMA-based control chart. For all methods utilising the current mean vector estimates instead of individual observations, the probability control chart triggers an out-of-control signal as soon as

$$p_t = 1 - \exp\left[-\frac{1}{8}\left(\mathbf{M}_t - \mathbf{M}_0\right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{M}_t - \mathbf{M}_0\right)\right] > h_2^*, \tag{3.24}$$

with h_2^* as the upper limit to achieve an ARL_0 value.

3.2.5   Confidence control chart with sliding window schemes

For all cases of sliding window (SW) schemes, the same form of Equation (3.24) is utilised but only the observation vectors inside the current window are weighted. Thus, the current mean vector $\mathbf{M}_t$ is given by

$$\mathbf{M}_t = \sum_{i=t-k+1}^{t} w_i^* \mathbf{Y}_i, \tag{3.25}$$

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

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

$$w_i^U = \frac{1}{k}, i = t - k + 1, \ldots, t. \tag{3.26}$$

Second, the linear sliding window (LSW) approach gives more weight to the most recent observation and decreases linearly the weight as the observation gets older such that

$$w_i^L = \frac{\frac{j}{k}}{\sum_{j=1}^{k} \frac{j}{k}}. \tag{3.27}$$

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

$$w_i^E = \frac{j^\varphi}{\sum_{j=1}^{k} j^\varphi}, \tag{3.28}$$

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 3.1, while the weights for the windows with size 2 are shown in Table 3.2.

The control chart calibration procedure is performed in two steps to achieve an

Table 3.1: Weights computation for sliding window schemes with size 4 (SW4)

		Position			
Scheme	$t-3$	$t-2$	$t-1$	t	*Sum*
Uniform	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{4}{4}$
Uniform weights	0.25	0.25	0.25	0.25	1
Linear	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{10}{4}$
Linear weights	0.1	0.2	0.3	0.4	1
Exponential	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 3.2: Weights for sliding window schemes with size 2 (SW2)

	Position		
Scheme	$t-1$	t	*Sum*
Uniform	0.50	0.50	1
Linear	0.33	0.67	1
Exponential	0.41	0.59	1

$ARL_0 = 200$ for all control charts. The first step adjusts the linear regression models in the form $d^2 = a + b \cdot ln\,(ARL)$. This procedure gives an approximate first estimate of 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.

3.3 Experiments and results

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. In Figure 3.1, 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 3.1(a) shows the signal pattern for the case of no change in the mean vector (that is, $d = 0$), while Figures 3.1(b) and (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 3.1(d), (e) and (f), the out-of-control observation vectors are marked with light red dots, while the in-control vectors are marked with dark black dots.
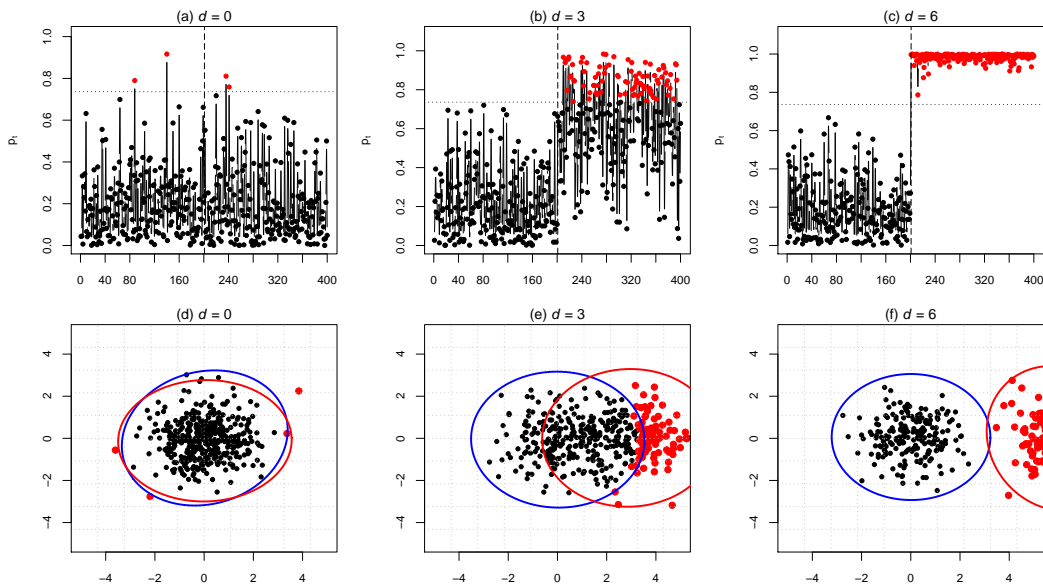
Figure 3.1: Confidence control chart for individual vectors (SW1) with scatter plots

Noteworthy in Figure 3.1(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 3.3, 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 3.3: Summary of Hotelling's T^2 and SW1 statistics with ARL comparison

d	$\overline{d^2}$	$Sd\left(d^2\right)$	ARL	$\overline{p}$(x100%)	$Sd\left(p\right)$	ARL
0.0	2.0	1.9	200.6	20.0	15.7	200.6
0.5	2.3	2.1	118.8	22.0	16.9	117.7
1.0	3.0	2.6	43.1	27.6	19.6	43.1
1.5	4.3	3.4	16.0	36.1	22.0	16.0
2.0	6.0	4.3	7.0	46.4	23.0	7.0
2.5	8.3	5.2	3.6	57.2	22.3	3.6
3.0	11.0	6.1	2.2	67.5	20.1	2.2
3.5	14.3	7.0	1.5	76.5	17.0	1.5
4.0	18.0	8.0	1.2	83.9	13.4	1.2
h	10.7			73.6		

Figure 3.2 is composed of the four sets of control charts and theirs respective two-dimensional scatter plots. The Confidence MEWMA control chart utilises $\lambda = 1$, such that it performs identically to the SW1 control chart, at a maximum standard confidence level to protect the global in-control process region. As evident in this example, 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 3.2 , which is $h^*_{\text{USW}} = h^*_{\text{ESW}} = 0.4811$ ($d = 2.2908$). This reduction indicates that the chart becomes sensitive to small changes in the mean vector,

Figure 3.2: Confidence control charts with $\lambda = 1$, SW2, $\varphi = 1$ and the respective scatter plots ($\mathbf{M}_1 = (3,0)$)

no matter the individual distances. The in-control limit for the LSW chart with SW2 scheme is $h^*_{\text{LSW}} = 0.5166$ ($d = 2.4115$).

Figure 3.3 shows the reducing effect on the confidence levels for all control charts. The Confidence MEWMA control chart with $\lambda = 0.7$ is called MEWMA.7 and the transformed in-control limit is $h_{MEWMA.7} = 0.5086$ ($d = 2.3842$). For the SW2 chart with the ESW scheme in Figure 3.2, the estimated threshold is $h_{ESW} = 0.4901$ ($d = 2.3212$).

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 3.2). 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
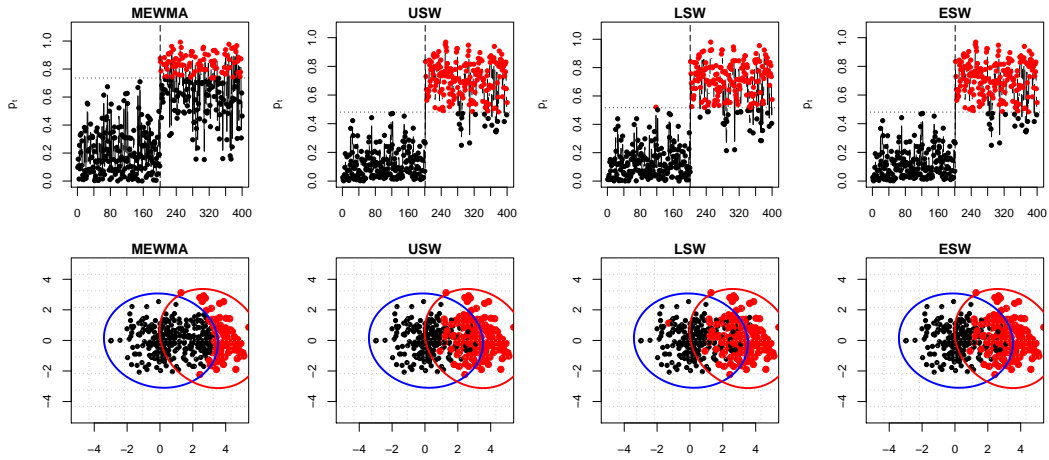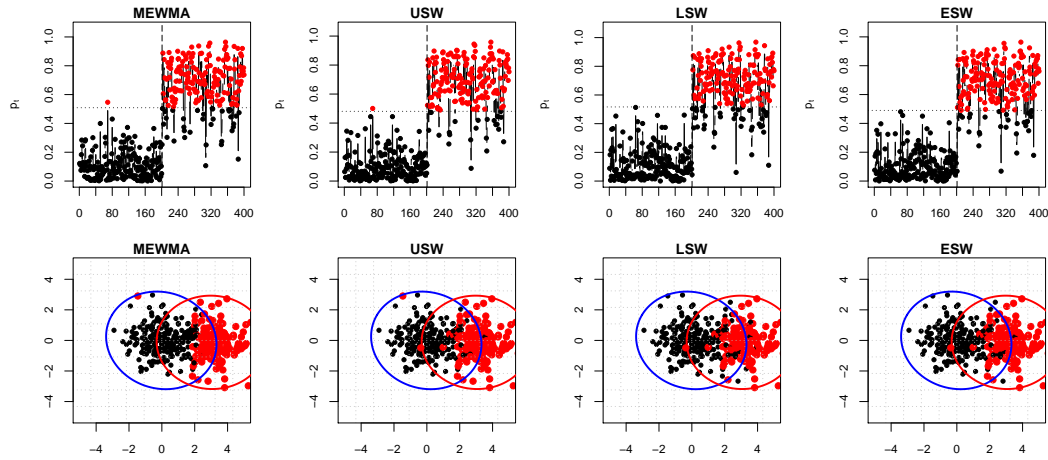
Figure 3.3: Confidence control charts with $\lambda = 0.7$, SW2, $\varphi = 0.7$ and the respective scatter plots ($\mathbf{M}_1 = (3,0)$)

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 3.4 and 3.5. 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 3.4: Summary statistics for the MEWMA.7 and SW2 control charts

	MEWMA.7		USW2		LSW2		ESW2	
d	$\overline{p}$(x100%)	$Sd\,(p)$	$\overline{p}$(x100%)	$Sd\,(p)$	$\overline{p}$(x100%)	$Sd\,(p)$	$\overline{p}$(x100%)	$Sd\,(p)$
0.0	11.8	9.9	11.1	9.1	12.2	10.0	11.4	9.3
0.5	14.1	11.3	13.4	10.5	14.4	11.4	13.6	10.7
1.0	20.4	14.2	19.8	13.5	20.5	14.4	19.9	13.8
1.5	29.9	16.9	29.4	16.4	29.7	17.5	29.4	16.8
2.0	41.2	18.4	40.8	18.3	40.7	19.6	40.6	18.8
2.5	53.1	18.3	52.6	19.0	52.1	20.8	52.2	19.8
3.0	64.3	17.1	63.7	18.6	62.7	21.1	63.1	19.9
3.5	74.1	14.9	73.2	17.7	71.9	21.0	72.5	19.4
4.0	81.9	12.4	80.8	16.3	79.1	20.6	79.9	18.6
h	50.9		48.1		51.7		49.0	

Table 3.5: Summary statistics for the MEWMA.4 and SW4 control charts

	MEWMA.4		USW4		LSW4		ESW4	
d	$\overline{p}$(x100%)	$Sd(p)$	$\overline{p}$(x100%)	$Sd(p)$	$\overline{p}$(x100%)	$Sd(p)$	$\overline{p}$(x100%)	$Sd(p)$
0.0	5.8	4.7	5.5	4.3	6.6	5.3	6.3	5.1
0.5	7.8	5.9	7.6	5.7	8.6	6.6	8.4	6.4
1.0	13.8	8.7	13.7	8.8	14.5	9.8	14.2	9.6
1.5	22.7	11.7	22.9	12.7	23.3	13.6	23.0	13.5
2.0	33.5	14.5	33.8	16.5	33.7	17.4	33.4	17.4
2.5	44.9	16.8	45.0	20.0	44.6	20.6	44.3	20.9
3.0	55.9	18.3	55.6	22.8	54.8	23.2	54.5	23.6
3.5	65.6	19.1	64.8	24.7	63.8	24.8	63.4	25.5
4.0	73.8	19.2	72.2	25.8	71.1	25.4	70.6	26.3
h	27.5		26.8		31.9		30.8	

The ARL comparisons between the MEWMA.7, MEWMA.4, SW2 and SW4 control charts are given in Table 3.6. Although the SW2 control charts perform better than the SW1 chart (Table 3.1) and similarly to the MEWMA.7 chart for small shifts, an inertial effect is visible for distances larger than $d = 3$ 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 3.6: 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
0.5	83.2	82.0	84.6	81.0	52.6	54.4	61.1	59.2
1.0	22.7	22.2	23.7	22.4	12.8	13.2	15.4	14.8
1.5	8.4	8.0	8.6	8.1	5.6	5.7	6.7	6.5
2.0	4.1	3.9	4.2	4.0	3.4	3.7	4.3	4.2
2.5	2.6	2.5	2.8	2.6	2.5	3.0	3.5	3.5
3.0	1.9	2.0	2.2	2.1	2.0	2.6	3.0	3.0
3.5	1.5	1.8	2.0	1.9	1.7	2.3	2.6	2.6
4.0	1.3	1.6	1.9	1.8	1.5	2.1	2.2	2.3

When the sliding window size increases to 4 (SW4), the Confidence MEWMA control chart has the λ parameter decreased from 0.7 to 0.4, and for comparison purposes, is

called MEWMA.4. Figure 3.4 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 $h_{\text{MEWMA.4}} = 0.2747$ ($d = 1.6028$), $h_{\text{USW}} = 0.2677$ ($d = 1.5789$), $h_{\text{LSW}} = 0.4901$ ($d = 1.7530$) and $h_{\text{ESW}} = 0.3082$ ($d = 1.7168$). Table 3.5 provides the summary statistics for the MEWMA.4 and SW4 schemes.
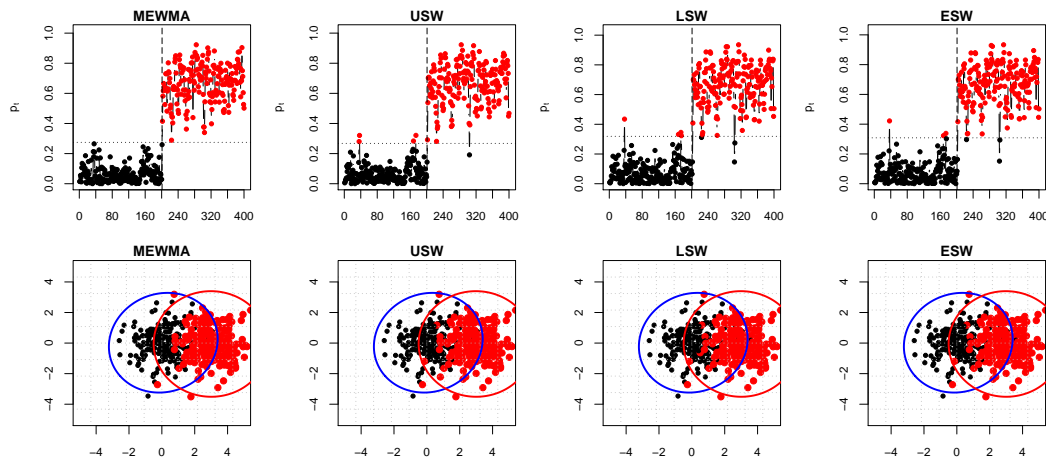
Figure 3.4: Confidence control charts with $\lambda = 0.4$, SW4, $\varphi = 0.7$ and the respective scatter plots ($\mathbf{M}_1 = (3,0)$)

Figure 3.5(a) shows an ARL comparison of all control charts, while Figure 3.5(b) uses the natural logarithm to amplify the differences in the tail. Splitting the comparison into two groups, Figure 3.6(a) compares the SW1, MEWMA.7 and SW2 control charts, while Figure 3.6(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.

The second set of control charts in Figure 3.6(b) compares the SW1, MEWMA.4 and SW4 charts. Although these sliding window control charts perform better for shifts below $d = 2$, 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

(a) ARL (b) ln(ARL)
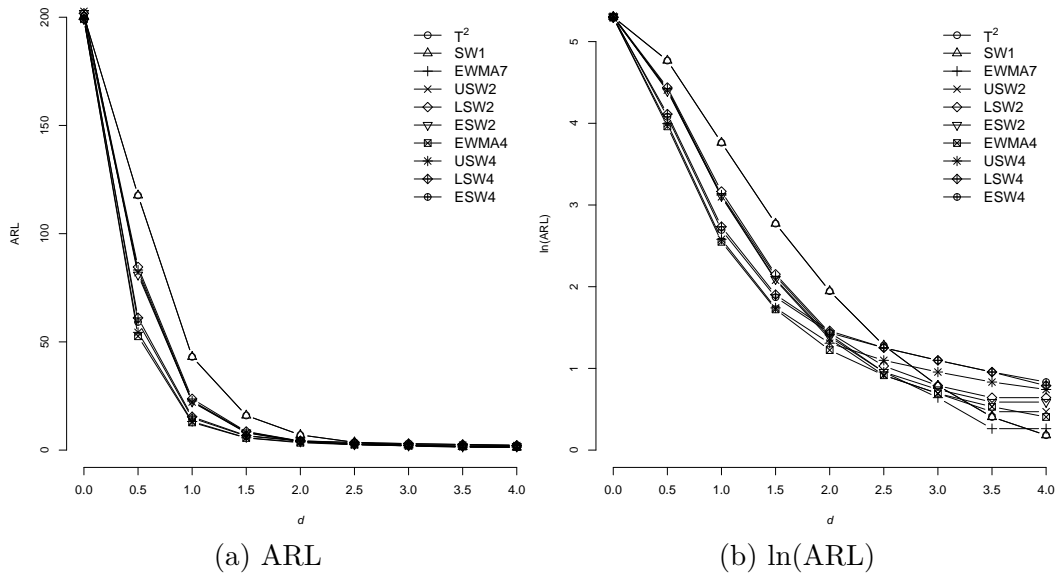
Figure 3.5: ARL and ln(ARL) comparison for all control charts

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 3.7 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. The λ's are positioned above the MEWMA bars and the window sizes for the SW schemes are specified in the horizontal axis. 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.

(a) SW1, MEWMA.7 and SW2   (b) SW1, MEWMA.4 and SW4
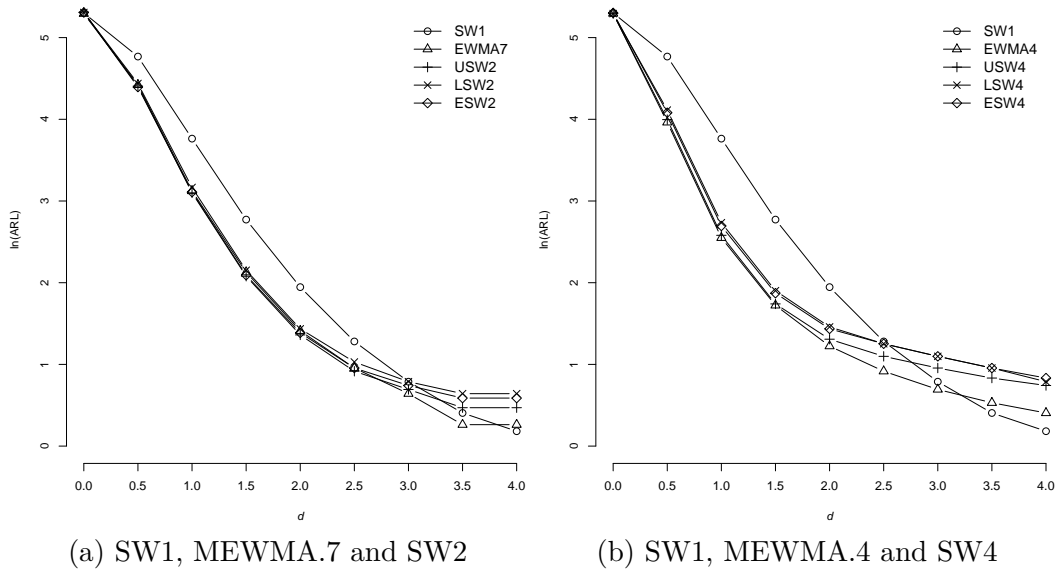
Figure 3.6: Comparison for SW1, MEWMA.7, SW2, MEWMA.4 and SW4 schemes
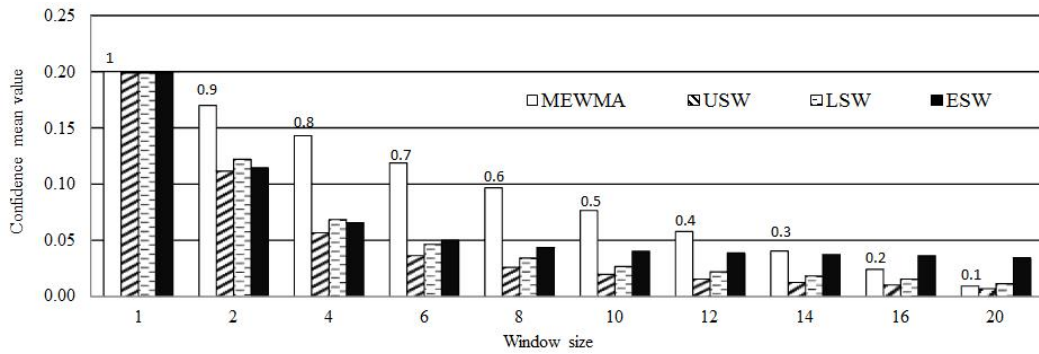
Figure 3.7: Comparison of mean values of the MEWMA and SW control charts

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 3.8. These values agree completely with the expected ones.

Figure 3.8: Mean value and standard deviation of the Confidence MEWMA control chart for the in-control process with various λ's

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. From the results presented in Figure 3.9, an interesting out-of-control statistics pattern is noticeable for the confidence chart in the transition period that is fixed to 20 observations. Note that, as the mean values in Figure 3.9(a) decreases with the smoothing parameter λ, the standard deviation has a maximum point in Figure 3.9(b) that is highly affected by the inertial period of 20 observations and that does not converge to the expected value.

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
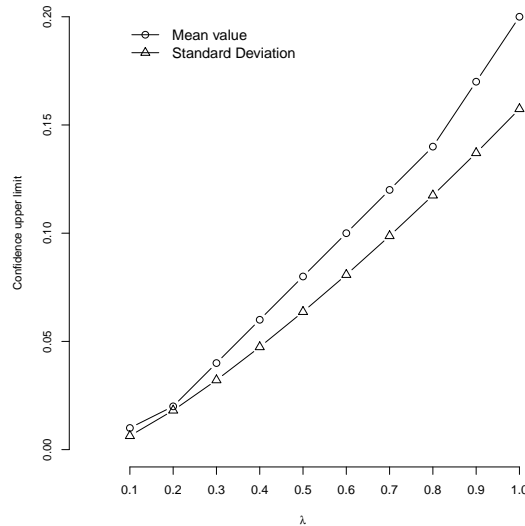
(a) Mean   (b) Standard deviation
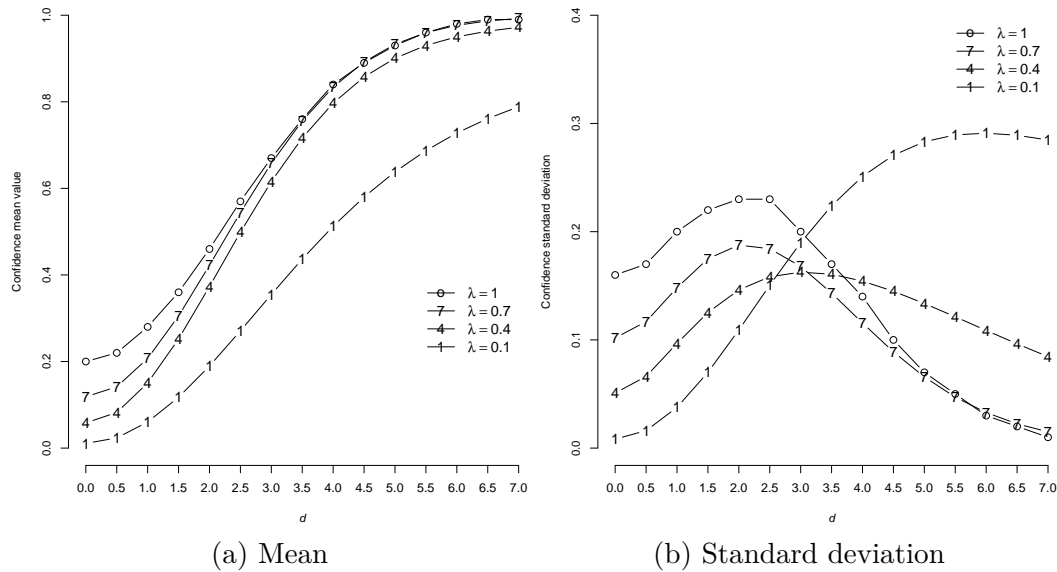
Figure 3.9: Mean value and standard deviation of the Confidence MEWMA control chart for the out-of-control process with various λ's

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 Figures 3.10- 3.12. 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 3.10, the threshold of all proposed control charts are nominally specified to detect a small shift utilising $d = 1$,
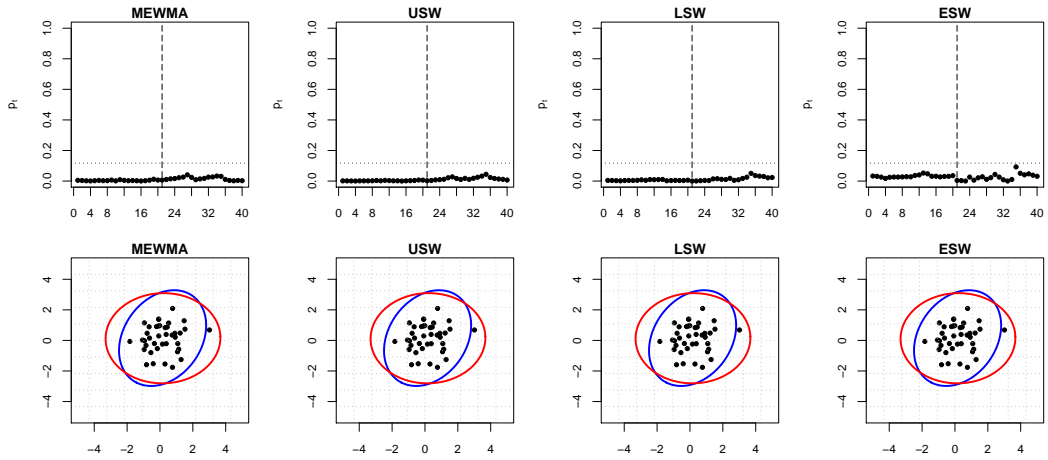
Figure 3.10: Transitional phase comparison for $d = 0$ with MEWMA.1 and SW20 schemes
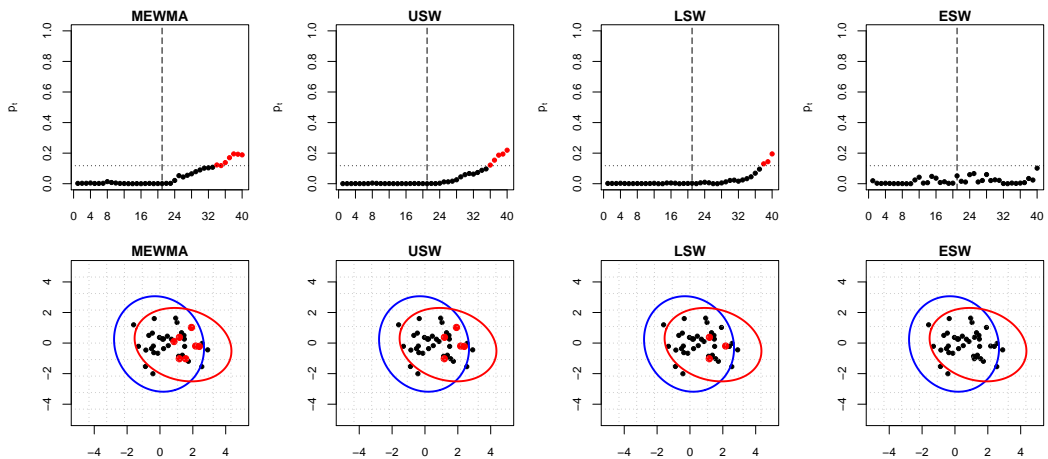
Figure 3.11: Transitional phase comparison for $d = 1$ with MEWMA.1 and SW20 schemes

which given in probability is $h^* = 1 - \exp\left[-\frac{1}{8}1^2\right] = 0.1175$. Observe that all control charts are not exactly calibrated to the same ARL, but as shown in Figure 3.7, 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 high confidence mean value shown in Figure 3.7. In Figure 3.11 and 3.12, a short run length of size 40 is monitored when two types of shifts occur at
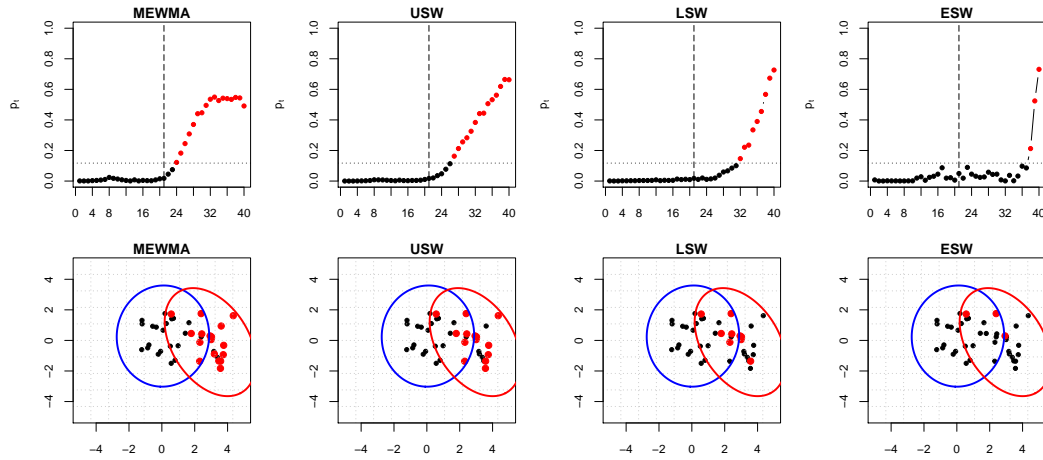
Figure 3.12: Transitional phase comparison for $d = 3$ with MEWMA.1 and SW20 schemes

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 $d = 3$ occurred, as shown in Figure 3.12, 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.

Additionally, to emphasize the context of Figure 3.9, it is interesting to note that while the expected value for the proposed statistic when $d = 7$ and $\lambda = 0.1$ can be

given by

$$
\begin{aligned}
E\left(p|\mathbf{M}_1 = (7,0)\right) &= E\left[1 - \exp\left(-\frac{1}{8}d^2\right)|\mathbf{M}_1 = (7,0)\right] \\
&= 1 - \exp\left[-E\left(\frac{1}{8}d^2|\mathbf{M}_1 = (7,0)\right)\right] \\
&= 1 - \exp\left[-\frac{1}{8}E\left(d^2|\mathbf{M}_1 = (7,0)\right)\right] \\
&= 1 - \exp\left[-\frac{1}{8}\left(n + \mathbf{M}_1'\mathbf{M}_1\right)\right] \\
&= 1 - \exp\left[-\frac{1}{8}\left(2 + (7,0)'(7,0)\right)\right] \\
&= 1 - \exp\left[-\frac{1}{8}(2 + 49)\right] \\
&= 1 - \exp\left[-\frac{51}{8}\right] = 0.9983 \quad (3.29)
\end{aligned}
$$

The observed value of the first 20 individual 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.

3.4   Discussion

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.

Chapter 4

Self-oriented Control Charts for Efficient Monitoring of Mean Vectors

Abstract

This work presents a procedure for monitoring the centre of multivariate processes by optimizing 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 criteria 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, maximizing efficiency with respect to the average run length.

4.1 Introduction

Quality assessment should be considered in several respects, as it is a deciding factor when one intends to acquire a product or service. Among the modern techniques for quality improvement and control, we highlight the statistical methods used initially by Shewhart (1931) that led to the formal statistical process control (SPC) field. As the field of automatic process control grew, many issues concerning machine learning and risk prevention arose. In the specific case of monitoring multivariate processes when few information sources about the possible changes are available, machines must be calibrated to notify problems promptly as they arise. Thus, 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. After an alarm warning, the units used are to be inspected to compute losses. In some cases, the cost involved in the production of a few defective units may be of a serious consequence.

For a single characteristic, control charts can be considered as the definition of the thresholds that the values of the quality characteristic must satisfy such that the process is considered stable. However, in multivariate processes, defining the control limits is a more complex task. In addition to the computational costs due to high correlations among the original variables, an increase in the number of parameters also requires an increase in the number of samples for the accurate estimation of the current state of the process.

In terms of controlling the mean vector, a proeminent but often neglected feature is the emptiness property of the multivariate space centre illustrated in Figure 4.1, which contradicts the common concepts of one-dimensional spaces. As can be observed in

(a) One-dimensional density   (b) Two-dimensional density
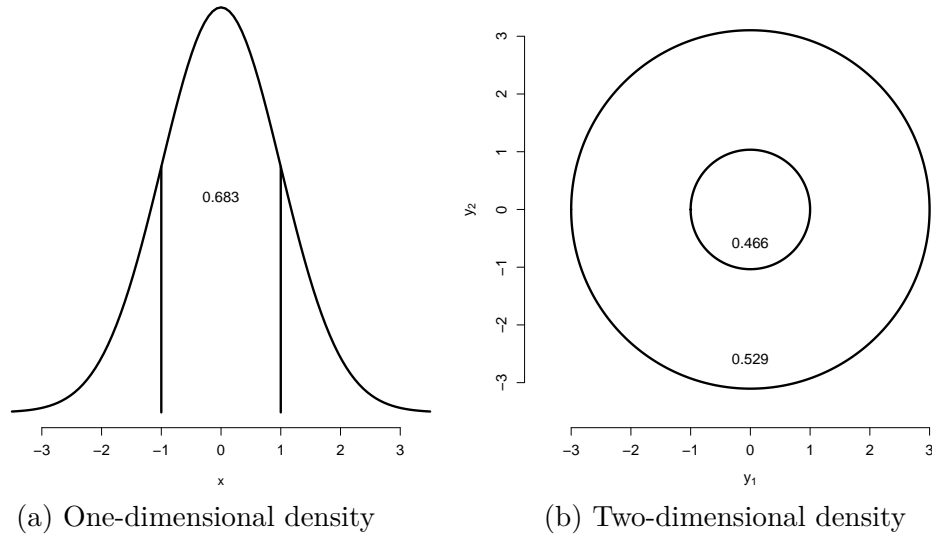
Figure 4.1: Emptiness property of the centre of multivariate spaces

Figure 4.1 a), the probability density between the -1 to +1 interval sums to 68.3%, while the bi-dimensional density inside the circle of radius 1 sums to 46.6%. This effect increases continuously with the dimensions of the data. The effect of this property on Gaussian data implies that when the data size increases, the likelihood that observations are present near the centre of the distribution decreases. In the case of high-dimensional spaces, the majority of the data concentrate in the tails of the distribution, and the presence of observations near the process centre becomes a very rare event. In other words, in high-dimensional spaces, Gaussian data and many other heavy-tailed distributions tend to concentrate in the tails, with highly sparse data in between.

Analysing this aspect of multivariate data, Jimenez and Landgrebe (1998) 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. Although some feature reduction techniques are already well known in the pattern recognition and digital image analysis, this methodology remains unexplored in time-varying single-hypothesis problems, which are extensively

used in multivariate process control and monitoring.

Many authors, such as Jackson (1991), Kourti and MacGregor (1996) and Choi et al. (2005), suggest monitoring schemes 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 in terms of ARL. This phenomenon possibly occurs because the linear and orthogonal transformations of the PCA method do not 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 the classification accuracy in terms of the Bayes error. In this work, we propose a linear control chart for mean vectors that can additionally provide optimum performance in terms of the average run length for a wide range of changes in the process.

Assuming that the observation is a random sample with a conditional density function that depends on its class, the purpose of this work is to find an efficient feature that classifies samples in time-varying processes with a minimum probability of error. Although the same type of linear transformation utilised in PCA is applied in the present work, a different criterion for optimization is adopted.

4.1.1 Control charts for Gaussian processes

The 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 in-control (IC) state or the out-of-control (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.

For the SPC of covariance matrices, traditional approaches include the use of moving ranges or generalised variance tests (Montgomery, 2001), while new approaches include the use of auxiliary information proposed by Riaz and Does (2008) and the VMAX procedure of Costa and Machado (2008). Quinino et al. (2012) also propose a single statistic based on the mixture of variances (VMIX) to monitor the covariance matrix of bivariate processes. Recent proposals for simultaneous monitoring mean vectors and covariance matrices are developed by Zhang et al. (2010) and Khoo et al. (2010), with great improvements in the monitoring process. In the case of autocorrelated data, the known proposals for monitoring mean vectors include time series modelling and controlling the residuals of the adjusted model.

The most applied methods for monitoring mean vectors of a stable process employ Hotelling's T^2 (1947), the Crosier's (1988) Multivariate Cumulative Sum (MCUSUM),

and the Multivariate Exponentially Weighted Moving Average (MEWMA) chart of Lowry et al. (1992). These three charts are directly comparable due to the common dependency of the ARL on the noncentrality parameter. 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 observation vectors in terms of the ARL.

4.1.2 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 and Tsung, 2008) that incorporate the generalized likelihood ratio test is a successful approach.

The noncentrality parameter is a squared distance with known distribution and properties (Fukunaga, 1990). The distance between the mean vectors of two populations is also known as the Mahalanobis distance (Mahalanobis, 1936). 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 decrease.

The Mahalanobis distance can be directly linked with other dissimilarity measures used in classification problems (Tou and 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 paid for unknowing the alternative class.

In this paper a procedure is presented for monitoring the mean vectors of multivariate Gaussian processes by optimising the Mahalanobis distance with respect to the maximum separability between the mean vectors. The single and time-continuous hypothesis test performed using the control charts can be viewed as a binary pattern recognition problem with the aim of classifying a given sample into one of two possible classes, IC or OC. As the parameters of the second class (the OC state) can be estimated *a posteriori*, the single-hypothesis problem can be extended to a two-class problem. This procedure allows for the optimisation of the method in terms of maximum separability between classes in the IC and the OC states. In this optimal space, the proposed 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. The structure of the paper is the following. Section 4.2 describes the methodology. Section 4.3 shows numerical simulations, and Section 4.4 presents the final remarks.

4.2 Methodology

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 noncentrality value is exactly the Hotelling's distance $(T^2 = d^2)$ of individual observation vectors to the process centre, given by

$$d^2 = (\mathbf{x}_i - \mathbf{M}_0)' \, \mathbf{\Sigma}_0^{-1} \, (\mathbf{x}_i - \mathbf{M}_0) > h_1 \tag{4.1}$$

where $\mathbf{x}_i, \mathbf{M}_0$ and $\mathbf{\Sigma}_0$ 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

$$d_M^2 = (\mathbf{M}_2 - \mathbf{M}_1)' \left[\frac{\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2}{2} \right]^{-1} (\mathbf{M}_2 - \mathbf{M}_1) \tag{4.2}$$

where $\mathbf{M}_i, \mathbf{\Sigma}_i, i = 1,2$, are respectively the mean vectors and the covariance matrices of the process i. 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.

4.2.1 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 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 the 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 ($\mathbf{S}_w$) shows the scatter samples around their mean vectors, given by $\mathbf{S}_w = \sum_{i=1}^{L} P_i \mathbf{\Sigma}_i$, where P_i is the *a priori* probability of each of the L classes. A between-scatter matrix ($\mathbf{S}_b$) is the scatter of the expected vectors around the mixture means, $\mathbf{S}_b = \sum_{i=1}^{L} P_i \left(\mathbf{M}_i - \mathbf{M}_0 \right) \left(\mathbf{M}_i - \mathbf{M}_0 \right)'$. The mixture scatter matrix is the covariance of all samples regardless of their class assignments, defined as $\mathbf{S}_w + \mathbf{S}_b$.

To be an effective class separability criterion, these matrices need to be converted to a number that should increase when $\mathbf{S}_b$ increases or $\mathbf{S}_w$ decreases. A typical criterion for preserving class separability is $J = tr \left(\mathbf{S}_w^{-1} \mathbf{S}_b \right)$. 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., $\mathbf{Y}|\omega_1 \ N_p \left(\mathbf{M}_1, \mathbf{\Sigma}_1 \right)$. Thus, the decision boundaries are adjusted according to the averaged covariance matrix instead of using only the prior information

for $\boldsymbol{\Sigma}_0$. 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

$$\mathbf{X} = \mathbf{A}'\mathbf{Y} \tag{4.3}$$

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 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, the Mahalanobis distance defined in Equation (4.2) becomes the noncentrality parameter comparing two population mean vectors,

$$d^2 = (\mathbf{M}_2 - \mathbf{M}_1)' \, \boldsymbol{\Sigma}^{-1} (\mathbf{M}_2 - \mathbf{M}_1) \tag{4.4}$$

Note that the product in the right-hand side of Equation (4.4) has dimension 1×1 and therefore becomes

$$d^2 = tr\left((\mathbf{M}_2 - \mathbf{M}_1)' \, \boldsymbol{\Sigma}^{-1} (\mathbf{M}_2 - \mathbf{M}_1)\right) \tag{4.5}$$

For two matrices $\mathbf{A}$ and $\mathbf{B}$, it is known that $tr\,(\mathbf{AB}) = tr\,(\mathbf{BA})$, such that

$$d^2 = tr\left(\boldsymbol{\Sigma}^{-1} (\mathbf{M}_2 - \mathbf{M}_1) (\mathbf{M}_2 - \mathbf{M}_1)'\right) \tag{4.6}$$

In the right-hand side of Equation (4.6), observe that $(\mathbf{M}_2 - \mathbf{M}_1)(\mathbf{M}_2 - \mathbf{M}_1)'$ is a matrix with dimensions $p \times p$ generated by the product of a vector by itself; therefore, its rank equals 1. In this case, $\left(\boldsymbol{\Sigma}^{-1}(\mathbf{M}_2 - \mathbf{M}_1)(\mathbf{M}_2 - \mathbf{M}_1)'\right)$ is also a $p \times p$ matrix with a rank equal to 1 and has only one eigenvalue other than zero, i.e.,

$$\lambda_1 \neq 0, \lambda_2 = \lambda_3 = \ldots = \lambda_n = 0 \tag{4.7}$$

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, Equation (4.6) becomes

$$d^2 = \lambda_1 \tag{4.8}$$

with

$$\lambda_1 = (\mathbf{M}_2 - \mathbf{M}_1)' \boldsymbol{\Sigma}^{-1}(\mathbf{M}_2 - \mathbf{M}_1) \tag{4.9}$$

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 λ_1 is exactly the noncentrality value, this fact guarantee the direct comparison of the proposed control charts with the selected ones.

For a vector $\mathbf{Y}$ with dimensions $p \times 1$, the extracted feature (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, X can be obtained by projecting $\mathbf{Y}$ on e_1 according to Equation (4.3).

$$X = \mathbf{e}_1'\mathbf{Y} \tag{4.10}$$

The direction of $\mathbf{e}_1$ can be obtained from the general problem of finding eigenvalues

and eigenvectors in a generic matrix $\boldsymbol{\eta}$ such that

$$\boldsymbol{\eta}\mathbf{e}_1 = \mathbf{e}_1\lambda_1 \tag{4.11}$$

In this case,

$$\boldsymbol{\eta} = \boldsymbol{\Sigma}^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right)\left(\mathbf{M}_2 - \mathbf{M}_1\right)' \tag{4.12}$$

Then,

$$\boldsymbol{\Sigma}^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right)\left(\mathbf{M}_2 - \mathbf{M}_1\right)'\mathbf{e}_1 = \mathbf{e}_1\left(\mathbf{M}_2 - \mathbf{M}_1\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right) \tag{4.13}$$

Therefore,

$$\mathbf{e}_1 = \boldsymbol{\Sigma}^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right). \tag{4.14}$$

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

$$X = \left(\boldsymbol{\Sigma}^{-1}\left(\mathbf{M}_2 - \mathbf{M}_1\right)\right)'\mathbf{Y}. \tag{4.15}$$

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.

4.2.2   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. 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. This 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 current mean vector $\mathbf{M}_t$ is estimated at the instant t with the inclusion of $\mathbf{Y}_t$ such that

$$\mathbf{M}_t = (1 - \lambda)\,\mathbf{M}_{t-1} + \lambda\mathbf{Y}_t \tag{4.16}$$

where $\mathbf{M}_0 = \mathbf{0}$ and $0 > \lambda \geq 1$. 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, they are called the Linear MEWMA (Lin-MEWMA) and Cumulative Sum of X (CUSUM-Lin) control charts.

Because the resultant variable approximates a normal distribution, 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 $\mathbf{Y}|\omega_0 \; MVN\,(\mathbf{M}_0, \mathbf{\Sigma}_0)$. The transformed variable triggers an out-of-control signal as soon as

$$X_t = \left(\mathbf{\Sigma}_0^{-1}\,(\mathbf{M}_t - \mathbf{M}_0)\right)' \mathbf{Y}_t \geq \pm h_1 \tag{4.17}$$

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

4.2.3   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. This procedure was less effective than the vectorial CUSUM (MCUSUM) chart. Similarly, this work employs the CUSUM scheme over a one-dimensional statistic of a multivariate process.

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

$$y_t = \frac{X_t - \mu_X}{\sigma_X} \tag{4.18}$$

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 Monte Carlo simulation. Then, the positive and negative standardised CUSUMs are defined as

$$C_t^+ = max\left[0, y_t - k + C_{t-1}^+\right] \tag{4.19}$$

$$C_t^- = max\left[0, -k - y_t + C_{t-1}^-\right] \tag{4.20}$$

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.

4.3   Experiments and results

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 Hotelling's 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 $k = 0.5$ (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.

4.3.1   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 $B = 5,000$ Monte Carlo simulations were performed to obtain the first ARL_0 estimates. For the final control chart comparisons, $B = 50,000$ 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 p-dimensional multivariate processes are simulated using $p = 2$ and $p = 4$.

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 (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.

4.3.2 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 4.1. The first five observations of the data are generated from a bivariate normal process with $\mathbf{M}_0 = (0, 0)$, unitary variances and correlation $\rho = 0.5$. 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 an average distance value for the bivariate case.

First of all, notice that both Lin-MEWMA and CUSUM-Lin charts were calibrated for a process with unknown parameters (Phase I 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 and 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 Lowry et al. (1992) paper, λ is set to

0.1 for comparison purposes.

Table 4.1: 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

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 $d \leq 2$. 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.

4.3.3 Control chart pattern analysis

The illustrations that follow compare the common patterns of the control chart for mean vector shifts. Figure 4.2 presents the simulation of an extended run length of $m = 400$ 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 4.2), and the vertical dashed line in the middle of the control chart indicates a change point for the

process.
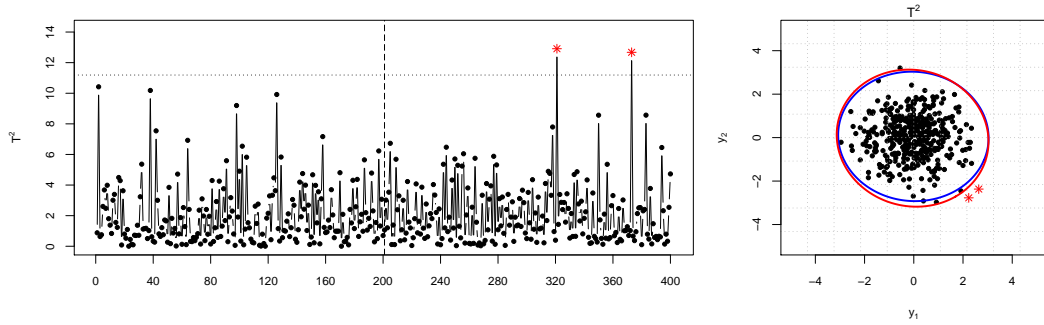
(a) Hotelling's T^2
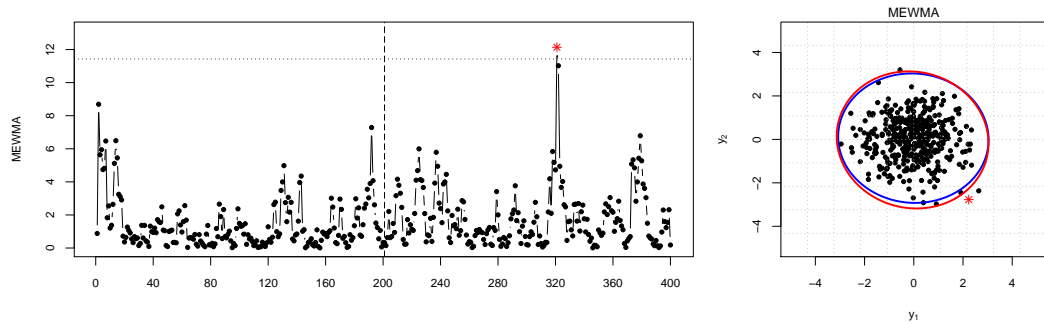
(b) MEWMA

(c) Lin-MEWMA

Figure 4.2: Hotelling's T^2, MEWMA and Lin-MEWMA control chart patterns with scatter plots ($p = 2, \lambda = 0.1, d = 0$)

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. Figure 4.2 show those estimated PCA boundaries, seen in the
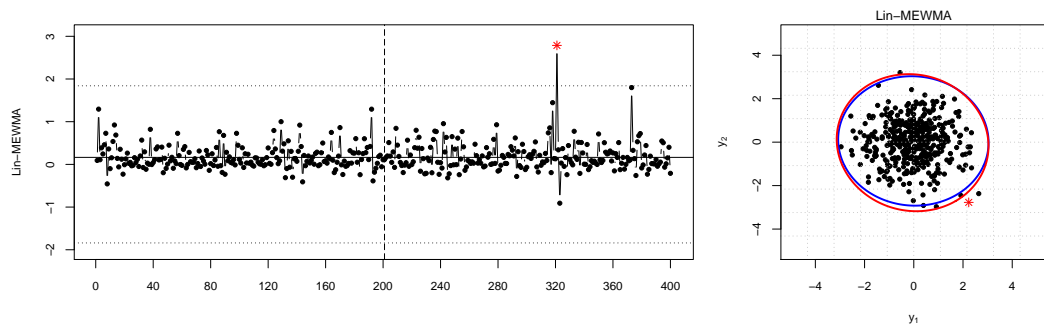
scatter plots on the right side of each respective control chart in blue and red. The blue circle is the estimated ellipsoid with a confidence level of 99.5% for the IC process, and the red circle is the estimated ellipsoid for the OC process.

Figure 4.3 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 approximately normal statistic, varying symmetrically 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 Figure 4.4, it can be seen that a shift in the mean vector ($d = 1$) can be detected by all compared control charts. This change is not perceived as well 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 statistic.

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 Figure 4.4(c), the linear statistic monitored using the Lin-MEWMA chart displays OC observations outside of the threshold projection in the direction of the shift. As $\mathbf{M}_1 = (1,0)$, 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
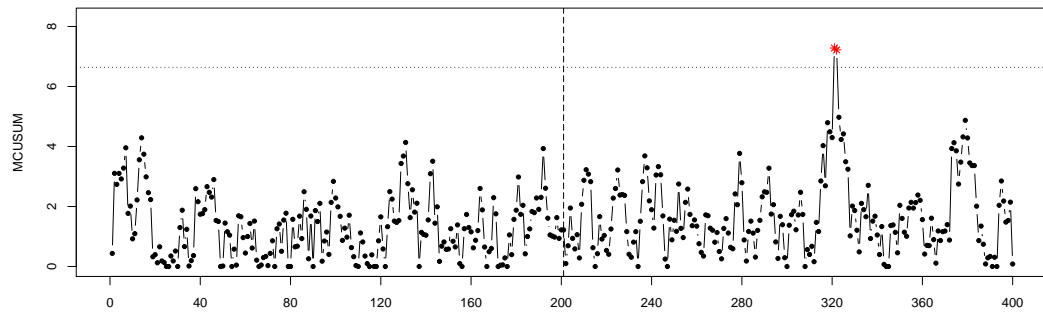
(a) CUSUM
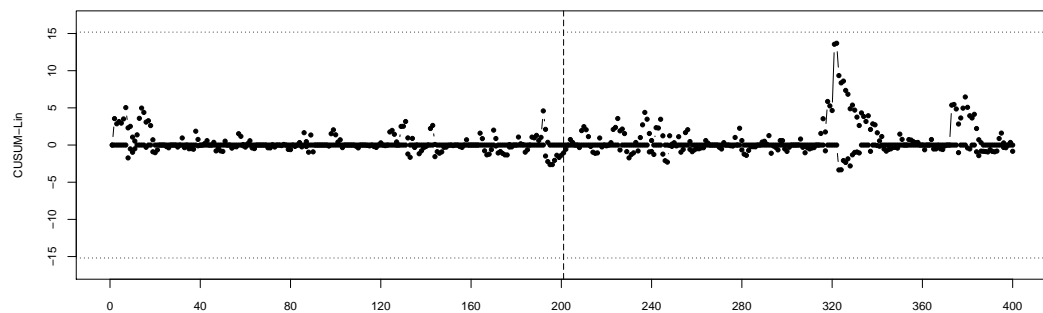
(b) CUSUM-Lin

Figure 4.3: MCUSUM and CUSUM-Lin control chart patterns ($p = 2, \lambda = 0.1, k = 0.5, d = 0$)

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 Figure 4.5, a non-Shewhart control chart, i.e., utilising the past process information, can also selectively tag observations vectors even if the current mean vector is already OC.

Increasing the distance of the OC process to $d = 2$, Figure 4.5 shows the behavior of the Hotelling's T^2, MEWMA and Lin-MEWMA charts. While the Hotelling's T^2
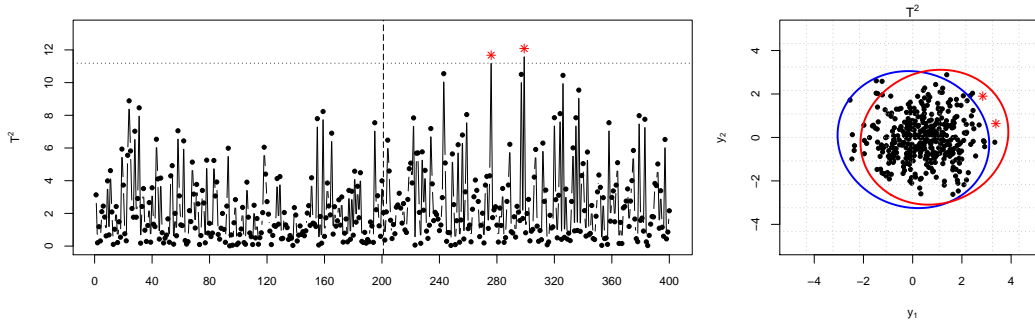
(a) Hotelling's T^2
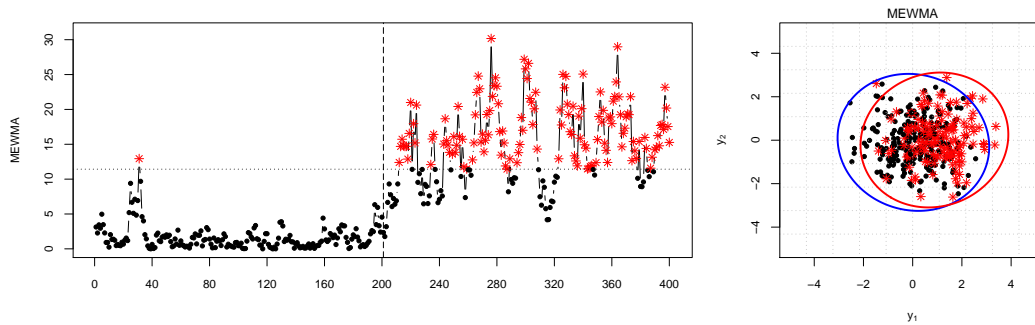
(b) MEWMA
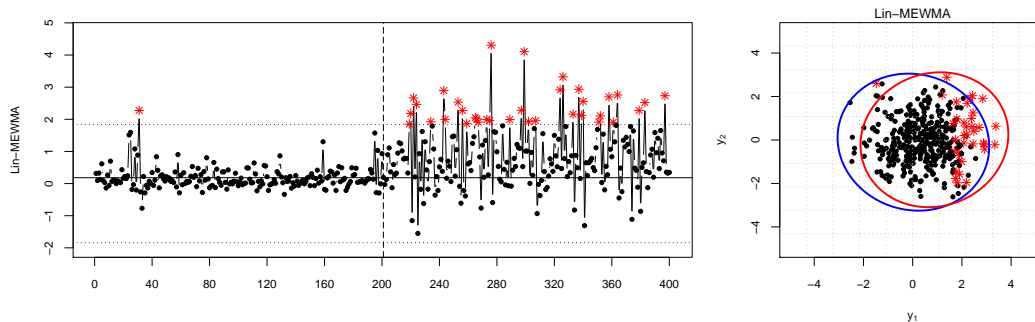
(c) Lin-MEWMA

Figure 4.4: Hotelling's T^2, MEWMA and Lin-MEWMA control chart patterns with scatter plots ($p = 2, \lambda = 0.1, d = 1$)

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
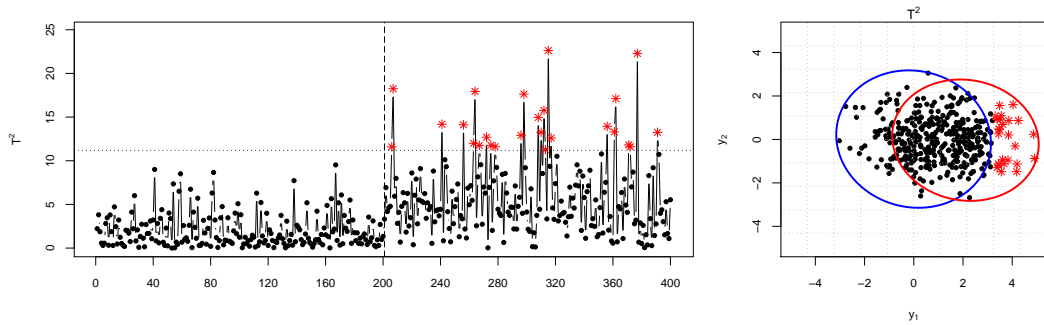
(a) Hotelling's T^2
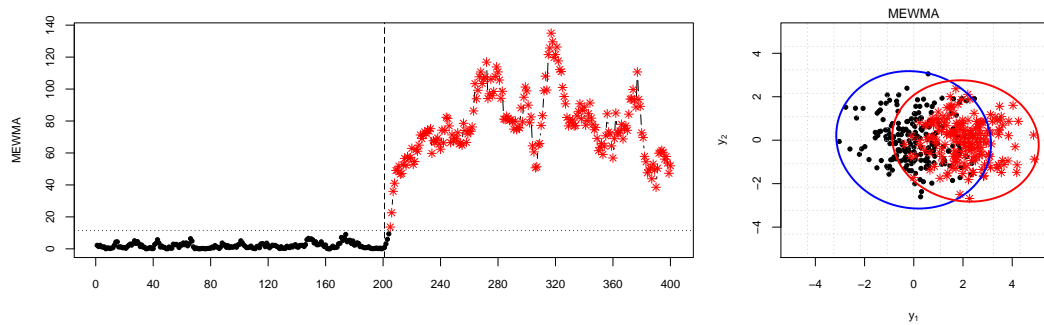
(b) MEWMA
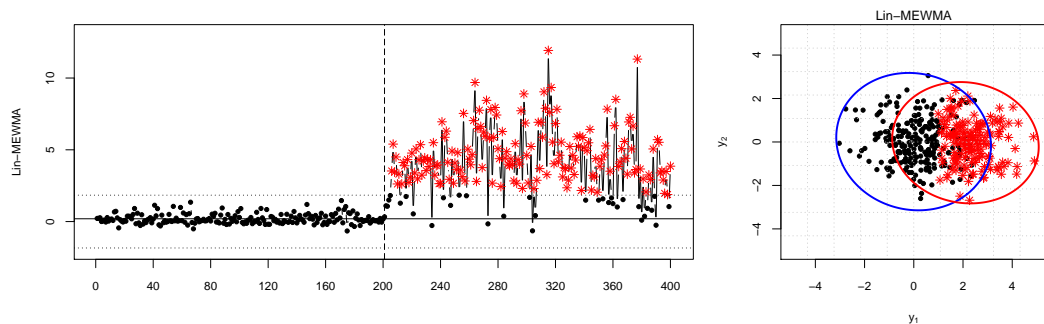
(c) Lin-MEWMA

Figure 4.5: Hotelling's T^2, MEWMA and Lin-MEWMA control chart patterns with scatter plots ($p = 2, \lambda = 0.1, d = 2$)

projected linear boundaries.

As seen in Figures 4.4-4.5, 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
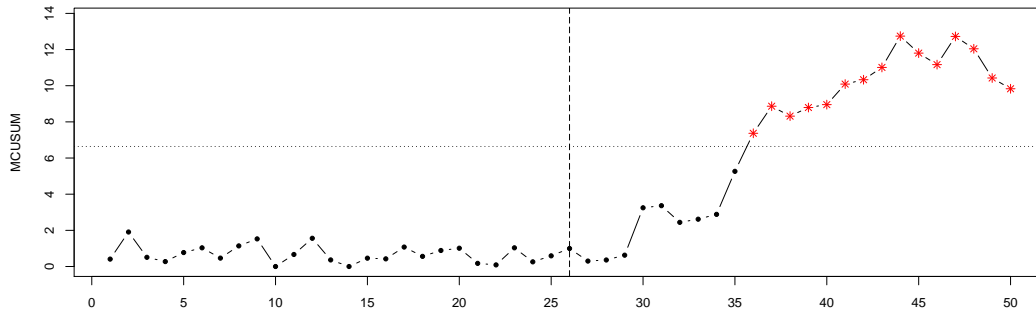
(a) CUSUM
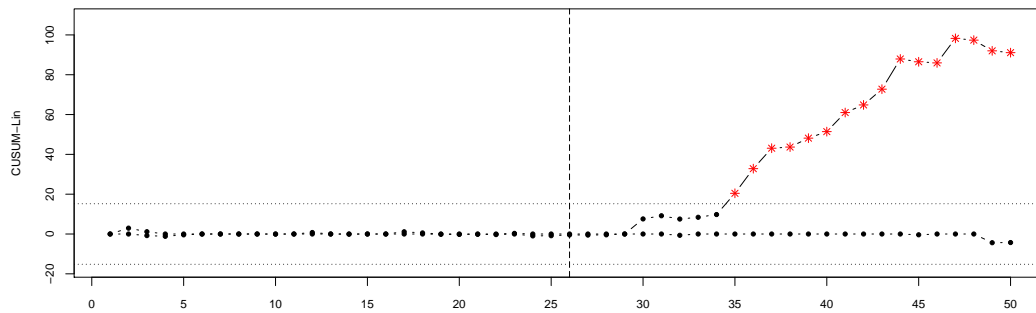
(b) CUSUM-Lin

Figure 4.6: MCUSUM and CUSUM-Lin control chart patterns ($p = 2, \lambda = 0.1, k = 0.5, d = 1$)

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.

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. Stan-

dardising the linear statistic through a CUSUM procedure renders the statistic to be a quadratic measure again, resulting in the loss of some important information on the vector location, as described above.

Figure 4.6 illustrates a simulated example with a reduced run length of $m = 50$ observations. In this case, a change occurs at position $t = 51$, and the CUSUM-Lin chart detects the change before the MCUSUM chart. In the example simulated in Figure 4.7, increasing the distance of the OC process to $d = 2$ results in the same performance for both MCUSUM and CUSUM-Lin control charts.
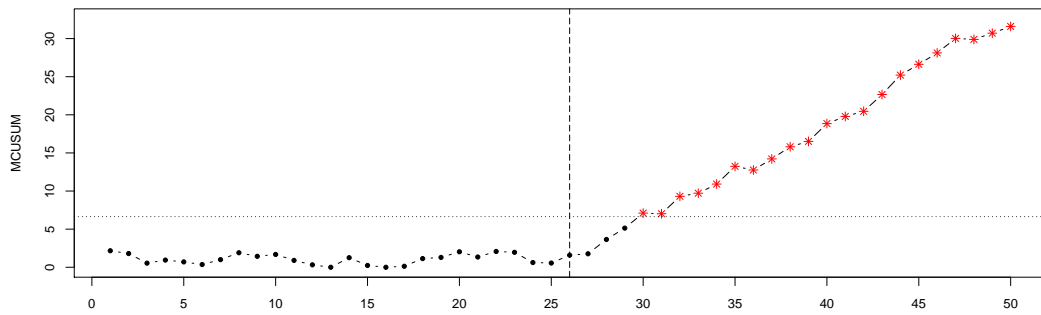
(a) CUSUM
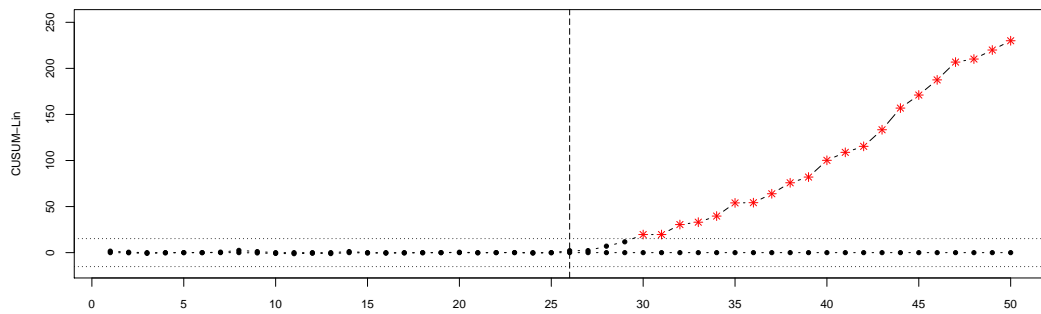
(b) CUSUM-Lin

Figure 4.7: MCUSUM and CUSUM-Lin control chart patterns ($p = 2, \lambda = 0.1, k = 0.5, d = 2$)

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 quadratic CUSUM-Lin chart after the smoothed linear transformation is also studied by the following ARL comparison.

4.3.4   Performance comparison

To effectively compare the performances of the control charts, the charts were calibrated to $\text{ARL}_0 = 200$. For the control chart calibration, the parameter estimates of $(\mathbf{M}_0, \boldsymbol{\Sigma}_0)$ were calculated using a Phase I stage with $m = 25$ observations of the IC process in the two-dimensional case $(p = 2)$ and with $m = 50$ for $p = 4$. 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.

Table 4.2: Control chart performance comparison $(p = 2, \lambda = 0.1)$

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
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

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 Hotelling's 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 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 4.2, for example, shifts bigger than $d > 0.5$ are detected using all the non-Shewhart methods in less than 15 observations.

Table 4.3: Control chart performance comparison ($p = 2, \lambda = 0.4$)

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

The experiments of Table 4.2 were designed to detect large shifts, and the best performance for the Hotelling's T^2 chart is achieved when $d \geq 3$. 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 $d \geq 2$ and behaves most similarly to the MEWMA chart when $d = 1.5$.

Table 4.4: Control chart performance comparison $(p = 4, \lambda = 0.1)$

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

By transforming the ARLs in a logarithmic scale, as shown in Figure 4.8, 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 ln(ARL) in Figure 4.8. For both the MCUSUM and MEWMA charts, the ARL curves in Figure 4.8(a) 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 ln(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 Figure 4.8(a) that the Lin-MEWMA and even the CUSUM-Lin charts are less affected by the problem of inertia than the MCUSUM and MEWMA charts.

Table 4.5: Control chart performance comparison ($p = 4, \lambda = 0.4$)

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

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 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.

(a) $p = 2, \lambda = 0.1$ (b) $p = 2, \lambda = 0.4$

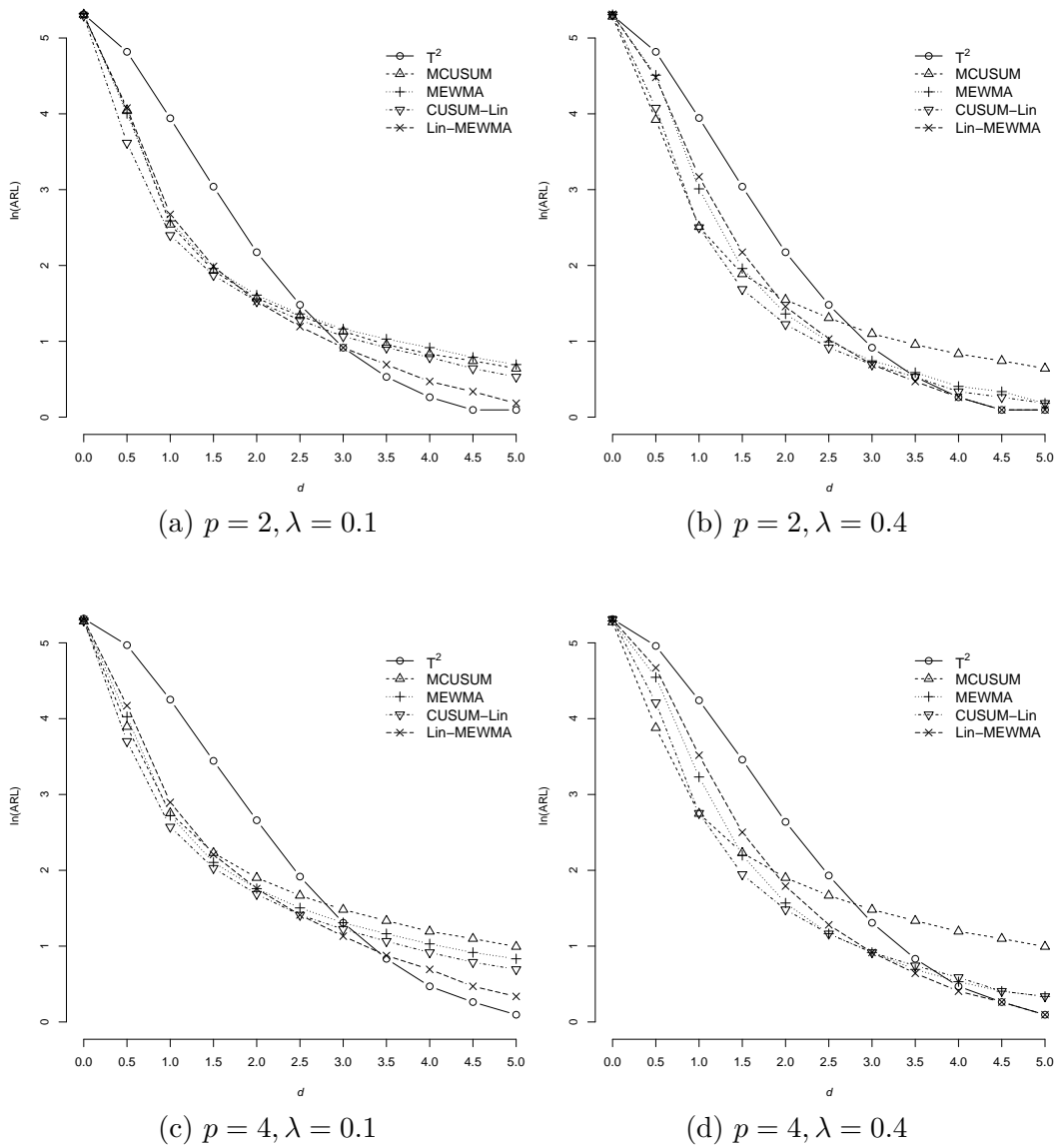(c) $p = 4, \lambda = 0.1$ (d) $p = 4, \lambda = 0.4$

Figure 4.8: Control chart ARL comparison on the logarithmic scale

To avoid the problem of inertia, the results in Table 4.3 for $\lambda = 0.4$ indicate that the Lin-MEWMA control chart has a performance superior to that of the MEWMA chart for $d \geq 3$. 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 4.3, 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 $p = 4$ are shown in Tables 4.4-4.5. 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 4.2-4.4, 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 4.4). Figures 4.8(c)-(d) present the ARL results of Tables 4.2-4.5 on the logarithmic scale. When $\lambda = 0.4$ (Table 4.5), 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 4.5 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) Lin-MEWMA $p = 2$ (b) Lin-MEWMA $p = 4$

(c) CUSUM-Lin $p = 2$ (d) CUSUM-Lin $p = 4$
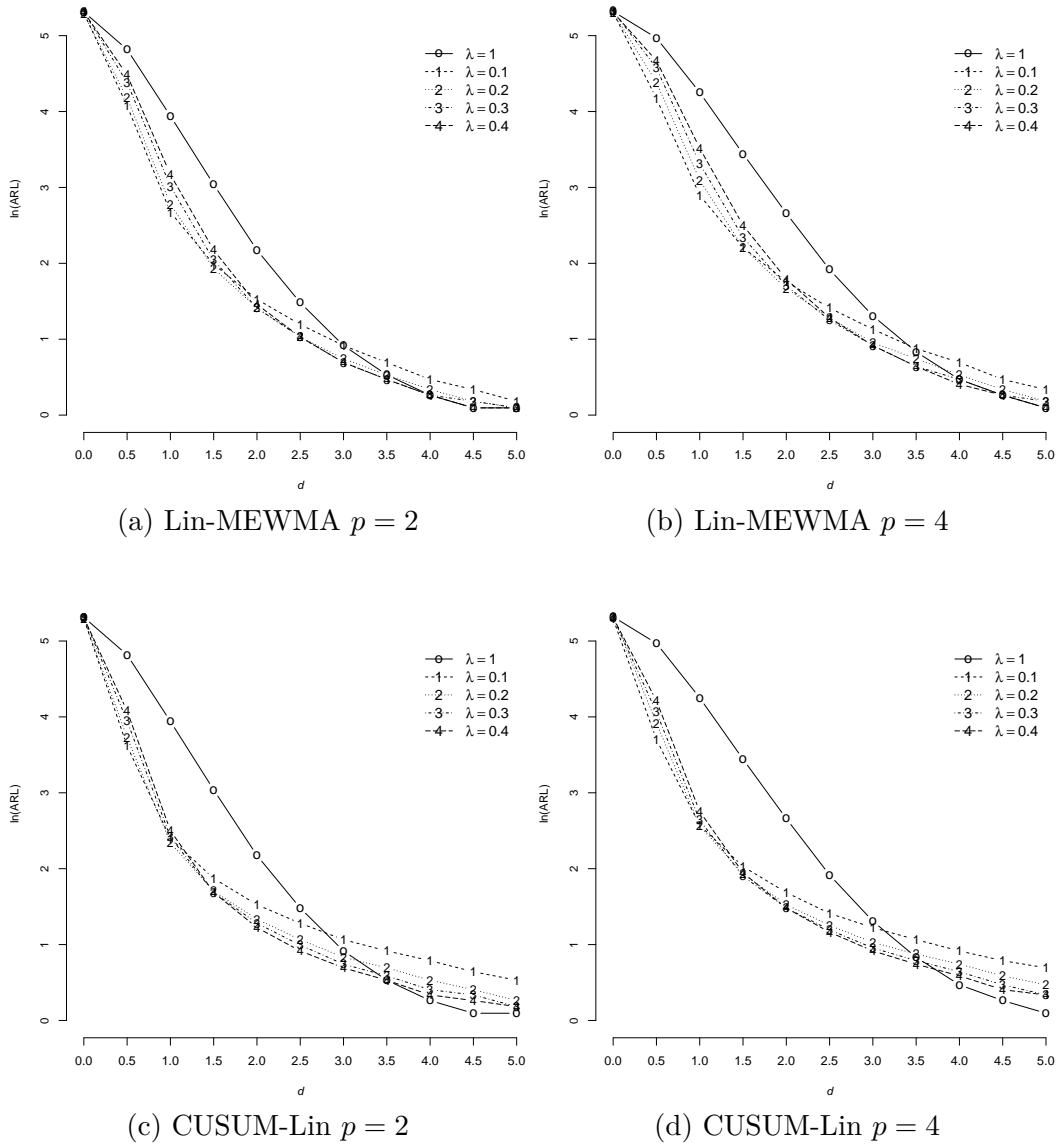
Figure 4.9: ARL comparison of the Lin-MEWMA and CUSUM-Lin control charts on the logarithmic scale

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

Table 4.6: Comparison of the performance of the CUSUM-Lin and Lin-MEWMA control charts ($p = 2$)

	CUSUM-Lin				Lin-MEWMA			
d	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$
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

the chart is to detect regular and large shifts early, the ARL curve incurs a loss of performance for 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 4.6, 4.7 and in Figure 4.9 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,

Table 4.7: Comparison of the performance of the CUSUM-Lin and Lin-MEWMA control charts ($p = 4$)

	CUSUM-Lin				Lin-MEWMA			
d	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$
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

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.

4.4 Discussion

It is important to realise that for multivariate Gaussian distributions, most samples fall into a doughnut-type ring-shaped region and that fewer samples fall into the centre, where the value of the density function is the largest. This effect increases with the process dimensionality, and the shift in the mean vector is considered to be significantly OC in high-dimension spaces for relatively larger distances. If the IC and OC processes are perfectly separable, the Hotelling's T^2 chart offers the best and simplest approach.

Given a significant overlap between the IC and OC processes, the selection of appropriate methods is of crucial importance.

As stated before, the performance of the control charts for mean vectors that are based on the noncentrality parameter are independent of the direction of the shift. If their performances were directionally dependent, the average run length could vary with the direction of change, performing better in one direction and worse in another direction.

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, 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 parameters. In fact,

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.

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 ($p = 2$ and $p = 4$), 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.

Future research on this topic may include comparisons with respect to the influence of the λ factor for higher dimensions, including hyperspaces ($p > 100$). 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.

Chapter 5

Conclusão

5.1 Considerações finais

Nesta Tese foi abordado o problema do monitoramento do vetor de médias de processos quando apenas uma observação do vetor multivariado está disponível a cada instante do tempo. A partir do estudo dos principais métodos aplicados no controle estatístico de processos (CEP), a saber os métodos MCUSUM, MEWMA e T^2 de Hotelling, foram identificados os padrões de atuação desses métodos bem como suas limitações.

A partir dessa investigação inicial foram levantados importantes aspectos quanto a essas medidas de distâncias, os quais deram origem a uma nova abordagem interpretativa na forma de limites de probabilidade, nomeadas como os Gráficos de Confiança. Os Gráficos de Confiança propostos, além de fornecerem uma maneira diferenciada na interpretação dos processos sob e fora de controle, foram úteis também na investigação de métodos alternativos para tentar reduzir o efeito inercial pronunciado no esquema MEWMA. Embora a abordagem alternativa com o uso de janelas deslizantes não tenha se mostrado efetiva no sentido de mitigar o efeito inercial resultante do esquema MEWMA, foi possível estabelecer uma forma empírica para a escolha adequada do tamanho das janelas deslizantes, parâmetro esse que têm sido também alvo de interesse

na literatura sobre CEP.

Classificando-se a magnitude das mudanças no vetor de médias em três tipos, considera-se que o primeiro tipo é a identificação de grandes mudanças no processo. Nesse contexto, grandes mudanças não são um problema de difícil solução, uma vez que os estados sob e fora de controle podem ser instantaneamente detectados por um gráfico de controle do tipo Shewhart, como o gráfico T^2 de Hotelling, por exemplo. O segundo e terceiro tipos de mudanças são classificados como de magnitudes pequenas e intermediárias, e são os tipos onde residem as maiores dificuldades das abordagens, tradicionalmente enfrentadas com gráficos de controle do tipo não-Shewhart. Conforme já exposto nos capítulos anteriores, o contínuo aumento na dimensionalidade dos processos acarreta em uma área de sobreposição entre os processos sob e fora de controle cada vez maior, pois a maior parte da densidade dos processos multivariados encontra-se nas caudas das distribuições. Isso caracteriza que mudanças em magnitudes intermediárias são muito importantes e devem abarcar um vasto número de aplicações práticas nos espaços multivariados.

Considera-se que o objetivo principal da Tese foi alcançado com a proposição de um critério diferenciado, utilizando projeções lineares adequadas para a classificação de dados. Embora já conhecidos na área de processamento de sinais, até então tais metodologias não haviam sido adaptadas ao problema temporal de controle de processos. Tal critério de projeção, ou transformação linear, visando reduzir a dimensão dos processos de forma ótima no sentido de maximizar a separabilidade entre os estados sob e fora de controle, resultou na minimização do tempo médio até a detecção das mudanças. Em primeiro lugar, a estatística proposta denominada Lin-MEWMA logrou bons resultados para mitigar o efeito inercial do método MEWMA em sua forma quadrática. Além disso, ao contrário do que acontece ao aplicar o esquema CUSUM sobre a estatística MEWMA quadrática, a aplicação do método denominado CUSUM-Lin mostrou-se enormemente eficiente para a rápida detecção de pequenas mudanças

nos processos simulados. Espera-se com isso que a proposta oferecida nessa Tese tenha grande aplicação em problemas práticos para a melhoria dos processos de controle de qualidade.

5.2 Trabalhos futuros

Dentre as várias possibilidades que se descortinaram a partir do trabalho iniciado nessa Tese, as principais vias de pesquisa para investigações futuras são:

- Disponibilização de um pacote no ambiente R contendo todas as técnicas propostas no presente trabalho;

- Estudo das formas fechadas para as estatísticas de primeira e segunda ordem da estatística proposta;

- Otimização da redução de dimensionalidade em processos multivariados visando a implementação de um gráfico de controle para o monitoramento da matriz de covariâncias;

- Unificação da proposta desenvolvida nessa Tese com o item acima, propondo um gráfico de controle para o monitoramento global de processos multivariados, isto é, controlar simultaneamente o vetor de médias e a matriz de covariâncias;

- O desenvolvimento de uma abordagem não-paramétrica para o controle de processos multivariados.

Bibliography

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

Bersemis, S., Psarakis, S., and 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.

Blazek, L., Novic, B., and Scott, D. (1987). Displaying multivariate data using polyplots. *Journal of Quality Technology*, 19(2):69–74.

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

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

Development Core Team, R. (2008). R: A language and environment for statistical computing. Technical report, R Foundation for Statistical Computing, Vienna, Austria.

Faraz, A. and Saniga, E. (2013). 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*, 29:407–415.

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

Guh, R. and 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. and 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. and 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.

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

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

Hwarng, H. and 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. and 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., The, S., and 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. and MacGregor, J. F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, 28:409–428.

Lowry, C. and Montgomery, D. (1995). A review of multivariate control charts. *II Transactions*, 6(27):800–810.

Lowry, C., Woodall, W., and Rigdon, S. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1):46–53.

Machado, M. and Costa, A. (2009). Monitoring the mean vector and the covariance matrix of bivariate processes. *Brazilian Journal of Operations and Production Management*, 5:47–62.

Mahalanobis, P. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1):49–55.

Mahmoud, M. and 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. and 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. and Noakes, L. (2005). Rao distances. *Journal of Multivariate Analysis*, 92:97–115.

Montgomery, D. (2001). *Introduction to Statistical Quality Control*. New York: Wiley.

Niaki, S. and Memar, A. (2009). 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*, 43:964–981.

Nikiforov, I. (2001). A simple change detection scheme. *Signal Processing*, 81:149–172.

Quinino, R., Costa, A., and Ho, L. L. (2012). A single statistic for monitoring the covariance matrix of bivariate processes. *Quality Engineering*, 24(3):423–430.

Rao, C. R. (1947). The problem of classification and distance between two populations. *Nature*, page 30.

Rao, C. R. (1949). On the distance between two populations. *Sankhya: The Indian Journal of Statistics*, 9:246–248.

Riaz, M. and Does, R. (2008). A process variability control chart. *Computational Statistics*, 24(2):345–368.

Seber, G. (1984). *Multivariate Observations*. Wiley: New York.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product.* American Society for Quality Control.

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

Tou, J. T. and Gonzalez, R. C. (1974). *Pattern Recognition Principles.* Addison-Wesley Publishing Company.

Tracy, N., Young, J., and Mason, R. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24:88–95.

Yeh, A., Lin, D., and McGrath, R. (2006). Multivariate control charts for monitoring covariance matrix: A review. *Quality Technology & Quantitative Management*, 3(4):415–436.

Zhang, J., Li, Z., and Wang, Z. (2010). A multivariate control chart for simultaneously monitoring process mean and variability. *Computational Statistics and Data Analysis*, 54:2244–2252.

Zou, C. and Tsung, F. (2008). Directional mewma schemes for multistage process monitoring and diagnosis. *Journal of Quality Technology*, 40(4):407–427.