



Departamento de Engenharia Mecânica e Industrial

**DESENVOLVIMENTO DA CARTA MULTIVARIADA DMPCA  
PARA DADOS AUTOCORRELACIONADOS – COMPARAÇÃO  
COM AS CARTAS  $T^2$  E DPCA**

DIOGO MIGUEL DA FONSECA FERREIRA

Dissertação apresentada na Faculdade de Ciências e Tecnologia da  
Universidade Nova de Lisboa no âmbito do Mestrado Integrado em Engenharia  
e Gestão Industrial

Orientadora: Prof. Ana Sofia Leonardo Vilela de Matos

Lisboa

Julho 2012



## **AGRADECIMENTOS**

A realização deste trabalho só foi possível graças à colaboração e ao apoio de um grande número de pessoas, relativamente às quais exprimo os meus mais sinceros agradecimentos.

Deste modo deixo expresso o meu reconhecimento:

À Professora Ana Sofia Matos, orientadora da presente dissertação, pela oportunidade da realização deste trabalho. Os meus mais sinceros agradecimentos por todo o apoio, incentivo, motivação, recomendações e cordialidade demonstrada.

Um agradecimento muito especial à minha família, nomeadamente os meus pais, pelo carinho e pelo suporte incondicional manifestado ao longo da elaboração desta dissertação.

Ao meu irmão, que considero um exemplo a seguir, por todo o apoio, carinho, amizade e conselhos sábios que sempre me guiaram pela minha vida.

À Margaret Ivens Ferraz por todo o carinho, amizade, disponibilidade, paciência e incessante solidariedade e apoio durante a realização deste trabalho.

Ao Colégio Militar, e a todas as pessoas que fizeram parte da minha vivência nesta instituição de ensino durante 8 anos, por toda a formação que me deram como pessoa, estudante e profissional.

Aos meus colegas de curso, Ana Filipa Barreira, Bruno Mendonça, César Santos, Diogo Moreira e Diogo Santos, pela amizade, incentivo e camaradagem, durante todo o percurso académico.



## SUMÁRIO

Nos últimos anos, os métodos de controlo de processos estatísticos multivariados (*MSPC*), nomeadamente a análise de componentes principais (*PCA*), tem evidenciado uma abordagem poderosa para a deteção e isolamento de condições anormais em indústrias de processo com as variáveis altamente correlacionadas. A presente dissertação tem como objetivo apresentar uma nova carta de controlo que segue a mesma filosofia da carta de análise de componentes principais dinâmicas (*DPCA*), mas introduz uma matriz com uma nova estrutura, onde as colunas apresentam um desfasamento permitindo o cancelamento de autocorrelação nas variáveis dos componentes principais. Um estudo comparativo entre o desempenho da conhecida carta de controlo de Hotelling  $T^2$  (usando resíduos ou previsões futuras), da carta de controlo de componentes principais dinâmicas (*DPCA*) e da nova carta de controlo proposta designada por *DMPCA* (Matriz desfasada dos componentes principais) é apresentado. A abordagem desenvolvida para a comparação das cartas é descrita em detalhes, usando o número médio de observações ao fim do qual se destaca uma situação fora de controlo (*ARL*) e o correspondente *SDRL* (desvio padrão da distribuição) como indicadores de desempenho. O estudo comparativo é efectuado com recurso a simulações de Monte Carlo, onde são geradas três variáveis diferentes em estruturas de dados autocorrelacionados (*AR* e *ARMA*) e sem correlação cruzada entre elas. As principais vantagens e desvantagens das cartas são apontadas, na perspectiva prática de quem aplica cartas de controlo multivariadas para monitorar processos dinâmicos contínuos em que o número de variáveis é reduzido. Este estudo revela melhorias consideráveis em relação ao uso da carta de controlo *DMPCA* para detectar pequenas ou moderadas alterações no parâmetro do processo quando comparado com a carta *DPCA* e grandes melhorias quando comparado com carta  $T^2$  de Hotelling.

**Palavras chave:** cartas de controlo multivariadas; dados autocorrelacionados, *ARL*, carta *DPCA*, carta de matriz desfasada com componentes principais, carta  $T^2$  de Hotelling



## ABSTRACT

Over the last years, multivariate statistical process control (*MSPC*) methods and namely principal component analysis (*PCA*) have shown to provide a powerful approach to detection and isolation of abnormal conditions in process industries with highly correlated variables. The present dissertation presents a new control chart that follows the same philosophy as *DPCA* chart but introduces a new matrix structure with deployed columns that allows canceling autocorrelation in the score variables. A comparative performance study between the well-known Hotelling's  $T^2$  control chart (using residuals or one-step-ahead predictions), the dynamic *PCA* (*DPCA*) chart and the new proposed control chart named as *DMPCA* (Deployed Matrix *PCA*) is presented. The approach developed to compare those charts is described in detail, using the average run length (*ARL*) and correspondent *SDRL* (standard deviation run length) as a performance indicator. Monte Carlo experiments are used to simulate three variables following different autocorrelated structures (*AR and ARMA*) and without cross correlation between them. The main advantages and disadvantages of each chart are pointed out, in the practical perspective of those who intent to use *MSPC* to monitor dynamic continuous processes with a small number of variables to be controlled. This study reveals considerable improvements regarding the use of *DMPCA* to detect small to moderate shifts in the mean process parameter when compared with *DPCA* and huge improvements when compared with  $T^2$  Hotelling's Chart.

**Key words:** multivariate control chart; autocorrelated data; *ARL*, *DPCA* chart; Deployed Matrix of principal components;  $T^2$  Hotelling's Chart



## LISTA DE ABREVIATURAS E SIMBOLOGIA

### Abreviaturas

ARIMA (p,d,q)	Modelo autorregressivo de ordem p, diferenciação de ordem d e de médias móveis de ordem q	<i>AutoRegressive Integrated Moving Average</i>
ARL	Número médio ao fim do qual se destaca uma situação fora de controlo	<i>Average Run Length</i>
ARX	Modelo autorregressivo com variável Exógena	<i>Autorregressive Model with Exogene Variable</i>
AR (p)	Modelo autorregressivo de ordem p	<i>Autorregressive Model</i>
CP	Componente Principal	<i>Principal Component</i>
Carta CCC	Carta de Causas Comuns	<i>Comum Cause Chart</i>
Carta CUSUM	Carta de Somas Acumuladas	<i>Cumulative Sums Chart</i>
Carta EWMA	Carta de Média Móvel Exponencialmente Amortecida	<i>Exponential Weight Moving Range Chart</i>
Carta EWMAST	Carta de Média Móvel Exponencialmente Amortecida para dados autocorrelacionados	
Carta MCEWMA	Carta EWMA com limites móveis	<i>Moving Center Line EWMA Chart</i>
Carta MEWMA	Carta EWMA Multivariada	<i>Multivariate EWMA Chart</i>
Carta MCUSUM	Carta CUSUM Multivariada	<i>Multivariate CUSUM Chart</i>
Carta MRRC		<i>Multivariate Ridge Residual Chart</i>
DPCA	Análise de Componentes Principais Dinâmicos	<i>Dynamic Principal Component Analysis</i>
FAC	Função de Autocorrelação	<i>Autocorrelation Function</i>
FACE	Função de Autocorrelação Estimada	<i>Estimate Autocorrelation Function</i>
FACP	Função de Autocorrelação Parcial	<i>Partial Autocorrelation Function</i>
FACPE	Função de Autocorrelação Parcial Estimada	<i>Estimate Partial Autocorrelation Function</i>
LC	Linha Central	<i>Center Line</i>
LIC	Limite Inferior de Controlo	<i>Lower Control Limit</i>
LSC	Limite Superior de Controlo	<i>Upper Control Limit</i>
MSE	Erro Médio Quadrático	<i>Mean Squared Error</i>
MA (q)	Modelo de Médias Móveis de ordem q	<i>Moving Average Model</i>
MR	Amplitudes Móveis	<i>Moving Range</i>
NIPALS		<i>Nonlinear Iterative Partial Least Square</i>
PCA	Análise de Componentes Principais	<i>Principal Component Analysis</i>
PLS	Regressão pelos Quadrados Mínimos Parciais	<i>Partial Least Square</i>
SDRL	Desvio padrão da distribuição de “Run Length”	<i>Standard Deviation of Run Length Distribution</i>
SPC	Controlo Estatístico do Processo	<i>Statistical Process Control</i>

### Símbolos

A	Número de componentes principais retidos
---	--

$c_j^{(Q)}$	Contribuição de cada variável relativamente à estatística $Q$
$c_j^{(ti)}$	Contribuição de cada variável relativamente à estatística $T_A^2$
$c_j^{(T_A^2)}$	Contribuição total da variável relativamente à estatística $T_A^2$
$C_{R,\alpha}$	Ponto crítico
$d$	Número de ordem de diferenciação
$E$	Matriz residual
$F_{\alpha;p;m-p}$	Percentil à direita, para uma probabilidade $\alpha$ , da Distribuição de Fisher, com $p$ e $m - p$ grau de liberdade
$g$	Factor multiplicativo
$h$	Limite de controlo da carta MCUSUM; número de graus de liberdade da distribuição Qui-Quadrado
$H$	Limite de controlo da carta MEWMA
$I$	Matriz identidade
$m$	Número de observações recolhidas para cada variável, quando se considera o estudo de cartas multivariadas
$n$	Dimensão de cada amostra
$p$	Número de variáveis, quando se considera o estudo de cartas multivariadas; número de ordem do modelo autorregressivo
$P$	Matriz dos vectores próprios
$p_{ij}$	Elemento da $i$ -ésima linha e $j$ -ésima coluna da matriz $P$
$q$	Número de ordem do modelo médias móveis
$Q$	Estatística para a determinação dos resíduos obtidos pela PCA
$T$	Matriz dos componentes principais
$T^2$	Estatística de Hotelling
$T_A^2$	Estatística de Hotelling baseada nos $A$ componentes principais
$T^2_{(i)}$	Valor da estatística de Hotelling calculada a partir de $(p - 1)$ variáveis, para o método da estatística $d$
$T_j^2$	Termos incondicionais obtidos da decomposição de $T^2$ em componentes ortogonais
$T_{j+1,1,2,\dots,j}^2$	Termos condicionais obtidos da decomposição de $T^2$ em componentes ortogonais
$T_t^2$	Valor da estatística da carta MEWMA, no instante $t$
$t_i$	Valor do componente principal $i$
$R$	Diagonal entre as constantes de amortecimento das $p$ variáveis, considerando a carta MEWMA
$S$	Matriz de covariância
$X$	Matriz de dados composta pelas $p$ variáveis, em que cada variável tem $m$ observações, cujas variáveis possuem média zero e variância unitária
$X_i$	Vector constituído pelas observações das $p$ variáveis no instante $i$
$X_{j,novo}$	Valor da observação da variável $j$
$\hat{X}_{j,novo}$	Estimativa da variável $j$ resultante do modelo PCA
$X_t$	Observações da variável no instante $t$
$Y_t^*$	Valor da estatística da carta MCUSUM, no instante $t$
$z_\alpha$	Percentil à direita, para uma probabilidade $\alpha$ , da Distribuição Normal Reduzida
$\alpha$	Nível de significância ou erro do tipo I
$\beta$	Erro do tipo II
$\beta_{\alpha,(\frac{p}{2}, \frac{m-p-1}{2})}$	Percentil à direita, para uma probabilidade $\alpha$ , da Distribuição Beta, com parâmetros $p/2$ e $(m - p - 1)/2$
$\omega$	Valor de uma determinada estatística (contexto do SPC)
$\mu$	Vector das médias
$\Sigma$	Matriz das covariâncias
$l$	Atraso incorporado na matriz de dados na DPCA
$\chi^2$	Estatística do Qui-Quadrado

$\varepsilon_t$	Ruído branco
$\delta$	Alteração no parâmetro médio do processo
$\delta^2(\mu)$	Parâmetro de não centralidade
$\sigma_\varepsilon$	Desvio padrão da série de ruído branco
$\sigma_\varepsilon^S$	Desvio padrão da série de ruído branco, estimado por simulação
$\lambda_i$	Valores próprios dos $i$ componentes principais
$\lambda_j$	Constante de amortecimento para variável $j$ , na carta MEWMA
$\theta$	Parâmetro de médias móveis
$\theta_i$	Valor obtido com base nos valores próprios da matriz das covariâncias dos resíduos
$\phi$	Parâmetro autorregressivo
$\phi_{kk}$	Coefficiente de autocorrelação parcial de ordem $k$
$\rho_k$	Coefficiente de correlação de desfasamento $k$ (lag $k$ )
$\gamma_k$	Autocovariância de desfasamento $k$
$\gamma_0$	Autocovariância de desfasamento $k = 0$



## ÍNDICE GERAL

<b>Sumário .....</b>	<b>v</b>
<b>Abstract .....</b>	<b>vii</b>
<b>LISTA DE ABREVIATURAS E SIMBOLOGIA .....</b>	<b>ix</b>
<b>ÍNDICE DE GERAL .....</b>	<b>xiii</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>xv</b>
<b>ÍNDICE DE TABELAS .....</b>	<b>xix</b>
<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1. <i>Considerações Gerais</i> .....	1
1.2. <i>Relevância</i> .....	1
1.3. <i>Objetivos</i> .....	3
1.4. <i>Metodologia Geral</i> .....	4
1.5. <i>Organização da Dissertação</i> .....	5
<b>2. FUNDAMENTOS TEÓRICOS .....</b>	<b>7</b>
2.1. <i>Introdução</i> .....	7
2.2. <i>Controlo estatístico do processo</i> .....	7
2.3. <i>Princípios das cartas de controlo</i> .....	10
2.4. <i>Medidas de Desempenho das Cartas de Controlo</i> .....	14
2.5. <i>Recolha de Dados</i> .....	17
2.6. <i>Condições de Aplicabilidade das Cartas de Controlo</i> .....	19
2.6.1. <i>Aleatoriedade dos dados</i> .....	19
2.6.2. <i>Autocorrelação dos dados</i> .....	20
2.6.3. <i>Normalidade dos dados</i> .....	29
2.7. <i>Cartas de Controlo Univariadas para Dados Autocorrelacionados</i> .....	30
2.8. <i>Cartas de Controlo Multivariadas</i> .....	31
2.8.1. <i>Carta <math>T^2</math></i> .....	33
2.8.2. <i>Carta CUSUM</i> .....	36
2.8.3. <i>Carta EWMA</i> .....	37
2.8.4. <i>Interpretação da Carta <math>T^2</math></i> .....	39
2.9. <i>Cartas de Controlo Multivariadas baseadas em Métodos de projeção</i> .....	43
2.9.1. <i>Interpretação da Carta PCA</i> .....	50

2.9.2.	Cartas de Controlo Multivariadas Baseadas na Análise de Componentes Principais Dinâmicas.....	53
2.10.	<i>Proposta de nova Carta de Controlo Multivariada, DMPCA</i> .....	55
2.10.1.	Carta DPCA e a Nova Proposta DMPCA.....	56
2.11.	<i>Síntese do capítulo</i> .....	57
<b>3.</b>	<b>METODOLOGIA</b> .....	<b>61</b>
3.1.	<i>Metodologia proposta</i> .....	61
3.2.	<i>Construção do Programa de Simulação</i> .....	63
3.3.	<i>Comparação do Desempenho das Cartas de Controlo</i> .....	71
3.4.	<i>Interpretação das Cartas de Controlo Multivariadas</i> .....	71
3.5.	<i>Aplicação Prática</i> .....	72
<b>4.</b>	<b>ANÁLISE DOS RESULTADOS</b> .....	<b>73</b>
4.1.	<i>Determinação dos limites de controlo</i> .....	73
4.2.	<i>Comparação do desempenho das cartas de controlo multivariadas</i> .....	74
4.3.	<i>Validação dos pressupostos das cartas de controlo multivariadas</i> .....	79
4.4.	<i>Ganhos da nova carta proposta, DMPCA, relativamente às outras cartas de controlo multivariadas</i> .....	83
4.5.	<i>Vantagens e Desvantagens das cartas multivariadas</i> .....	86
<b>5.</b>	<b>CONCLUSÃO E RECOMENDAÇÕES</b> .....	<b>91</b>
	<b>REFERÊNCIAS</b> .....	<b>95</b>
	<b>ANEXOS</b> .....	<b>101</b>

## ÍNDICE DE FIGURAS

Figura 2.1 - Carta de Controlo.....	11
Figura 2.2 - Regras de deteção de causas especiais - Norma ISO 8258:1991 .....	16
Figura 2.3 - Exemplo de um Gráfico de Contribuição .....	50
Figura 3.1 - Construção da carta DPCA e determinação do ARLEC e SDRLEC .....	64
Figura 3.2 - Estudo do desempenho da carta DPCA a alterações no parâmetro médio .	66
Figura 3.3 - Construção da carta DMPCA e determinação do ARLEC e SDRLEC.....	68
Figura 3.4 - Estudo do desempenho da carta DMPCA a alterações no parâmetro médio.....	70
Figura 4.1 – Cenário 1: (a) curvas de ARL para a carta $T^2$ de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de $0,5 \delta$ .....	76
Figura 4.2 – Cenário 2: (a) curvas de ARL para a carta $T^2$ de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de $0,5 \delta$ .....	77
Figura 4.3 – Cenário 3: (a) curvas de ARL para a carta $T^2$ de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de $0,5 \delta$ .....	77
Figura 4.4 – Cenário 4: (a) curvas de ARL para a carta $T^2$ de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de $0,5 \delta$ .....	78
Figura 4.5 – Cenário 5: (a) curvas de ARL para a carta $T^2$ de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de $0,5 \delta$ .....	78
Figura 4.6 - FAC e FACP da carta DPCA para a estatística $T^2$ (Cenário 1) .....	80
Figura 4.7 - FAC e FACP da carta DPCA para a estatística Q (Cenário 1) .....	81

Figura 4.8 - FAC e FACP da carta DMPCA para a estatística $T^2$ (Cenário 1).....	81
Figura 4.9 - FAC e FACP da carta DMPCA para a estatística Q (Cenário 1).....	82
Figura 4.10 - Cenário 1: Os ganhos obtidos com a nova carta proposta DMPCA quando comparada com a carta $T^2$ de Hotelling e a carta DPCA.....	83
Figura 4.11 - Cenário 2: Os ganhos obtidos com a nova carta proposta DMPCA quando comparada com a carta $T^2$ de Hotelling e a carta DPCA.....	84
Figura 4.12 - Cenário 3: Os ganhos obtidos com a nova carta proposta DMPCA quando comparada com a carta $T^2$ de Hotelling e a carta DPCA.....	84
Figura 4.13 - Cenário 4: Os ganhos obtidos com a nova carta proposta DMPCA quando comparada com a carta $T^2$ de Hotelling e a carta DPCA.....	85
Figura 4.14 - Cenário 5: Os ganhos obtidos com a nova carta proposta DMPCA quando comparada com a carta $T^2$ de Hotelling e a carta DPCA.....	85
Figura I.1 - FAC e FACP da carta DPCA para a estatística $T^2$ (Cenário 2).....	101
Figura I.2 - FAC e FACP da carta DPCA para a estatística Q (Cenário 2).....	101
Figura II.117 - FAC e FACP da carta DMPCA para a estatística $T^2$ (Cenário 2).....	102
Figura II.2 - FAC e FACP da carta DMPCA para a estatística Q (Cenário 2).....	102
Figura III.1 - FAC e FACP da carta DPCA para a estatística $T^2$ (Cenário 3).....	103
Figura III.2 - FAC e FACP da carta DPCA para a estatística Q (Cenário 3).....	103
Figura IV.1 - FAC e FACP da carta DMPCA para a estatística $T^2$ (Cenário 3).....	104
Figura IV.2 - FAC e FACP da carta DMPCA para a estatística Q (Cenário 3).....	104
Figura V.1 - FAC e FACP da carta DPCA para a estatística $T^2$ (Cenário 4).....	105
Figura V.2 - FAC e FACP da carta DPCA para a estatística Q (Cenário 4).....	105
Figura VI.1 - FAC e FACP da carta DMPCA para a estatística $T^2$ (Cenário 4).....	106
Figura VI.2 - FAC e FACP da carta DMPCA para a estatística Q (Cenário 4).....	106

Figura VII.1 - FAC e FACP da carta DPCA para a estatística $T^2$ (Cenário 5).....	107
Figura VII.2 - FAC e FACP da carta DPCA para a estatística Q (Cenário 5).....	107
Figura VIII.1 - FAC e FACP da carta DMPCA para a estatística $T^2$ (Cenário 5).....	108
Figura VIII.2 - FAC e FACP da carta DMPCA para a estatística Q (Cenário 5).....	108



## ÍNDICE DE TABELAS

Tabela 2.1 - Regras da norma ISO 8258:1991 para detecção de causas especiais.....	16
Tabela 2.2 - Comportamento das Funções de Autocorrelação e Autocorrelação Parcial para os diferentes.....	24
Tabela 2.3 - Decomposição de $T^2$ em componentes ortogonais para $p = 3$ .....	41
Tabela 3.1 – Configuração do Processo de Simulação dos cinco cenários de estudo com os respectivos parâmetros do processo. ....	62
Tabela 4.1 – Configuração do Processo de Simulação e o Limite Superior de Controlo (LSC) para a estatística Q ( $ARL_{Em\ Controlo} = 370$ ) .....	74
Tabela 4.2 - Comparação das medidas de desempenho, ARL e SDRL, considerando os cinco cenários escolhidos com vários graus de autocorrelação .....	75



## 1. INTRODUÇÃO

### 1.1. CONSIDERAÇÕES GERAIS

Atualmente a competitividade e a globalização provocam uma procura contínua pela melhoria de desempenho nas organizações industriais, o que obriga estas a investigar novos processos e novas metodologias, que permitam produzir de forma eficiente, maximizando os recursos da organização, minimizando os custos e excedendo as expectativas dos clientes. Para que hoje uma organização se sinta competitiva e global é imprescindível que nas suas estratégias esteja presente o termo, processo, filosofia de Qualidade, a qual assume cada vez mais um papel determinante na concepção de produtos ou serviços, e na definição da oferta aos clientes.

Na base da Qualidade estão os métodos estatísticos que desempenham um papel fundamental na sua avaliação, controlo, auxiliando na medição, interpretação e modelação da variabilidade, mesmo quando a disponibilidade de dados é limitada. A análise estatística é uma ferramenta imprescindível para melhor compreender a essência, extensão e causas da variabilidade, e principalmente para obter a melhor solução ou por outro lado, prever todos os problemas que possam estar associados a tal variabilidade. A análise estatística é assim, a base mais fiável e eficaz na tomada de decisão.

Na análise estatística da Qualidade, destaca-se o controlo do processo com base em métodos estatísticos, designada como Controlo Estatístico do Processo (*SPC – Statistical Process Control*). Este método permite uma avaliação preliminar do processo, a estimação dos seus parâmetros, a avaliação da sua capacidade face à especificação técnica e controlo/monitorização do processo. Na essência, esta ferramenta possibilita a deteção de variações invulgares, atuando sobre elas, de forma a reduzir custos, perda de material, mão-de-obra, desgaste de material, equipamentos e reparações.

### 1.2. RELEVÂNCIA

A melhoria contínua da qualidade dos produtos e serviços nas organizações, é hoje um fator muito importante na tomada de decisão. Existe, desta forma, uma preocupação por partes das organizações em utilizar técnicas estatísticas adequadas para que os processos produtivos sejam estáveis e que a variabilidade destes seja a menor possível. A utilização de ferramentas

de controlo estatístico do processo, (*SPC*), é atualmente uma forma das organizações monitorizarem/controlarem o desempenho dos processos produtivos ao longo do tempo, detetarem falhas e ocorrências de causas especiais de variação. A implementação de cartas de controlo é uma das ferramentas de controlo estatístico do processo, que permite reduzir a variabilidade de uma dada característica na qualidade do processo.

Na maioria das indústrias, os processos produtivos apresentam inúmeras variáveis, as quais são necessárias monitorizar recorrendo a técnicas multivariadas, utilizando meios tecnológicos modernos de medição em tempo real para controlar diversas variáveis ao mesmo tempo. Normalmente, a aplicabilidade destas ferramentas só acontece em organizações que apresentam processos modernos, contínuos, com monitorização automática e obtenção de dados no instante de tempo em que se pretende, ou seja, em tempo real. No entanto, na maioria dos processos industriais automatizados, as variáveis apresentam dados autocorrelacionados, uma observação num dado instante de tempo depende de observações ocorridas em instantes anteriores. Assim, o controlo estatístico de processos multivariado com dados autocorrelacionados revelou-se uma ferramenta extremamente poderosa para indústrias com processos modernos detetarem variações nas suas operações e consequentemente combaterem-nas através da melhoria contínua dos processos.

A presente dissertação centrou-se no desenvolvimento de uma nova carta que permite corrigir a violação dos pressupostos subjacentes às cartas de controlo, conduzida pelas cartas *DPCA*, a que se chamou *DMPCA* (*Deployed Matrix Principal Components Analysis*) e no estudo comparativo do desempenho de cartas de controlo multivariadas, com dados autocorrelacionados – nomeadamente entre a carta  $T^2$  de Hotelling, desenvolvida por Hotelling em 1947 e a carta *DPCA* (*Dynamic Principal Component Analysis*), desenvolvida por Ku *et al.*(1995).

A carta desenvolvida por Hotelling,  $T^2$ , representa a carta de controlo multivariada mais conhecida na literatura, e desta forma é a mais utilizada e recomendada para processos que em que se pretendem controlar várias características da qualidade, no entanto, Montgomery (2005) alerta para o facto de só dever ser aplicada quando o número de variáveis é muito reduzido. Esta carta de controlo também pode ser aplicada em processos com dados autocorrelacionados.

Por outro lado, para o estudo comparativo em causa, seleccionou-se uma carta de controlo baseada em métodos de projeção, assente na análise de componentes principais dinâmica.

Esta opção deve-se ao facto desta ferramenta ser bastante poderosa na monitorização de processos complexos e com muitas características da qualidade, possibilitando desta forma, a redução significativa do número de variáveis a serem monitorizadas e também por ser possível a sua aplicabilidade a dados autocorrelacionados.

Para o estudo comparativo em causa, os dados utilizados para a componente prática foram simulados no *software* MATLAB.

Este estudo comparativo demonstra a necessidade das organizações procurarem de forma exhaustiva a otimização dos seus processos para daí desenvolverem o seu negócio em termos de recursos e rentabilidade. Para que tal aconteça, é necessário utilizar as ferramentas corretas, e principalmente as mais eficazes e que acarretam mais valor para atingir o objetivo. Acima de tudo, este estudo comparativo de cartas multivariadas com dados autocorrelacionados assenta em princípios diferentes, ou seja, cada carta apresenta os seus próprios princípios e, é necessário saber qual a carta que apresenta melhor desempenho quando o processo é sujeito a perturbações de diferentes magnitudes ao nível da média numa ou mais variáveis.

Na verdade esta análise comparativa, torna-se também de extrema relevância, uma vez que há escassez de estudos comparativos desta natureza recorrendo à simulação, que analisam o desempenho da carta de  $T^2$  de Hotteling, a carta *DPCA* e outras cartas multivariadas através das medidas de desempenho, *ARL* (*Average Run Length*).

Por fim, é importante reter as principais vantagens e desvantagens que cada uma das cartas apresentadas tem, para que este estudo seja relevante para a indústria no momento da escolha de uma carta de controlo para monitorizar processos multivariados com dados autocorrelacionados.

### 1.3. OBJETIVOS

Este trabalho parte da necessidade da conceptualização e desenvolvimento de uma nova carta de controlo para processos multivariados e com dados autocorrelacionados, *DMPCA* e centra-se também num estudo comparativo do desempenho de outras duas cartas de controlo multivariadas com dados autocorrelacionados, a carta de Hotteling desenvolvida em 1947,  $T^2$ , e a carta *DPCA*, desenvolvida por Ku *et al.* (1995).

Os objetivos principais a que se propõe o trabalho são:

1. Revisão bibliográfica sobre o Controlo Estatístico do Processo, nomeadamente multivariado, com especial incidência em processos com dados autocorrelacionados;
2. Interpretação e análise das cartas de controlo multivariadas com dados autocorrelacionados;
3. Desenvolver uma carta de controlo multivariada (nova), designada de carta *DMPCA* (*Deployed Matrix Principal Components Analysis*) que permita corrigir uma violação dos pressupostos subjacentes às cartas de controlo, conduzida pelas cartas *DPCA* (*Dynamic Principal Component Analysis*);
4. Avaliar o comportamento da nova carta *DMPCA* e estabelecer uma comparação do seu desempenho com as cartas  $T^2$  de Hotelling e *DPCA*, tendo por base o mesmo modelo matemático e as mesmas variáveis, de forma a identificar qual a carta que melhor permite detetar alterações no processo;
5. Facultar toda a informação necessária e essencial para a construção da nova carta, *DMPCA*, e procedimentos para a sua aplicabilidade, apresentando as suas vantagens e desvantagens para as indústrias com processos que apresentem várias características da qualidade e com dados autocorrelacionados;
6. Conclusões e recomendações para projetos futuros de desenvolvimento do estudo das cartas de controlo multivariado com dados autocorrelacionados.

#### **1.4. METODOLOGIA GERAL**

A concretização deste trabalho baseou-se na divisão em duas partes, procurando atingir objetivos propostos ao longo do mesmo.

Optou-se por dividir o trabalho em duas partes, que pelas suas características assim se impunha. A parte teórica como pilar fundamental para uma realização sustentada e sistematizada do trabalho prático a desenvolver.

## **Parte Teórica**

Para iniciar, foi realizado um levantamento bibliográfico, baseado numa recolha documental de tudo o que pudesse revelar-se de alguma utilidade para o trabalho a desenvolver (livros, estudos, artigos científicos, etc.), destes foram selecionados os que potencialmente poderiam trazer uma mais valia a este trabalho construindo-se assim esta parte de interesse nuclear para a realização de um projeto desta natureza.

## **Parte Prática**

Nesta parte, o importante é refletir os conhecimentos que foram retirados da 1ª parte, “Parte Teórica”, para aprofundar o estudo do controlo estatístico dos processos multivariados com dados autocorrelacionados para posteriormente ser possível a realização do estudo comparativo das cartas e retirar as devidas ilações para a melhoria dos processos produtivos e principalmente para a sua monitorização.

### **1.5. ORGANIZAÇÃO DA DISSERTAÇÃO**

Esta dissertação encontra-se dividida em cinco capítulos, sendo o primeiro capítulo, denominado de Introdução, inteiramente dedicado à apresentação do trabalho a desenvolver, noções como a relevância encontrada para o tema em desenvolvimento, ou os objetivos propostos para o sucesso da dissertação.

O segundo capítulo, denominado de Fundamentos Teóricos, dá foco ao enquadramento teórico na área do controlo estatístico do processo, onde apresenta resumidamente, os princípios das cartas de controlo, as principais cartas de controlos, as regras para deteção de causas especiais e as medidas de desempenho destas mesmas cartas. Para além da base referida, existe também um enorme foco no controlo estatístico do processo com dados autorrelacionados onde são apresentados os modelos *ARIMA* através da metodologia de Box e Jenkins, assim como apresentação e descrição de cartas univariadas e multivariadas concebidas para este tipo de processos, dando mais importância às últimas devido ao objetivo do estudo em causa. Dentro destas cartas multivariadas, dá-se destaque à carta  $T^2$  de Hotelling, à carta baseada na análise de componentes principais, tanto para o caso estático

*PCA* como para o caso dinâmico *DPCA*, e apresenta-se por fim a proposta de nova carta, denominada de *DMPCA* para combater o problema das cartas anteriormente referidas, a autocorrelação nas variáveis dos componentes principais. É este capítulo que tal como anteriormente referido integra a “Parte Teórica” da dissertação, que serve então essencialmente para sustentar as opções tomadas ao longo do caso de estudo no que se refere, que é o objetivo último deste trabalho de investigação.

A “Parte Prática” da dissertação será constituída por dois capítulos. O Capítulo 3, em que se explica a metodologia adotada para atingir os objetivos propostos, onde se indica e explica as opções tomadas e a caracterização dos métodos escolhidos em cada caso. Neste capítulo, o foco principal é apresentar a metodologia que serve como base ao estudo comparativo das três cartas de controlo multivariadas com dados autocorrelacionados, tendo com base os desenvolvimentos apresentados no capítulo 2, “Parte Teórica”. O capítulo seguinte, Capítulo 4, oferece uma visão da parte experimental realizada neste estudo. Aqui são apresentadas as variáveis relevantes para a monitorização do processo estatístico multivariado com dados autocorrelacionados para as cartas  $T^2$ , *DPCA* e *DMPCA*. O estudo comparativo entre elas é feito com recurso a simulação, com vários tipos de intervenção de autocorrelação, deste a autocorrelação forte até à autocorrelação fraca, sujeitos a perturbações em uma ou mais variáveis em simultâneo. No fim deste capítulo, dedica-se uma parte à análise e discussão dos resultados, que contém os resultados obtidos durante a realização do caso de estudo, bem como uma análise aprofundada a esses mesmos resultados. Entre eles destacam-se os pontos: 4.1 onde são determinados os limites de controlo das cartas *DPCA* e *DMPCA* para cada um dos processos estudados tendo em conta todos os cenários propostos, o ponto 4.2 onde é realizada uma análise de desempenho detalhada das três cartas tendo como base comparativa as medidas de desempenho *ARL* e *SDRL*, o ponto 4.3 que verifica se os pressupostos das cartas de controlo multivariadas com dados autocorrelacionados não são violados e o ponto 4.4 onde se realizou um estudo dos ganhos que a nova carta proposta *DMPCA* traz em relação às restantes.

Por fim existe subcapítulo no capítulo 4 inteiramente dedicado a resumir as vantagens e desvantagens das três cartas e um capítulo final, capítulo 5, inteiramente dedicado às conclusões e considerações desta dissertação assim como propostas para trabalhos futuros.

## 2. FUNDAMENTOS TEÓRICOS

### 2.1. INTRODUÇÃO

Este capítulo tem como objetivo apresentar os conceitos e técnicas para a aplicação do Controlo Estatístico do Processo, base da presente dissertação.

Objetivando o estudo desta dissertação, serão abordadas técnicas para o estudo de dados autocorrelacionados, com incidência nas cartas de controlo multivariadas, concebidas para a análise de processos com várias características da qualidade, fundamentalmente, para processos complexos presentes na indústria.

O Controlo Estatístico Multivariado representa desta forma uma mais-valia para as organizações, permitindo a eliminação de possíveis análises erróneas.

Em relação às cartas multivariadas que foram desenvolvidas para processos com dados autocorrelacionados, vai evidenciar-se mais a carta  $T^2$  de Hotelling e a carta baseada na análise de componentes principais dinâmica (DPCA), que serão base de comparação com a carta multivariada que se propõe desenvolver, DMPCA (*Deployed Matrix Principal Components Analysis*).

### 2.2. CONTROLO ESTATÍSTICO DO PROCESSO

O Controlo Estatístico do Processo, ou SPC (*Statistical Process Control*) foi inicialmente desenvolvido por Walter Shewhart em 1931. Esta ferramenta é extremamente poderosa para o controlo de processos de empresas das mais variadas áreas, tendo sido alvo de enriquecedoras obras por parte de diversos investigadores, passando-se a citar entre outros Burr (1976), Duncan (1986), Juran e Gryna (1993), Wheeler (1995), Pitt (1994), Dotty (1996), Quesenberry (1997), Montgomery (2005), Oakland (2008), Pereira e Requeijo (2008), entre outros.

Segundo Montgomery (2005), o SPC é um conjunto de instrumentos que têm como objetivo a resolução de problemas relacionados com a estabilidade do processo e melhoria da capacidade dos mesmos, através da redução da variabilidade.

As características da qualidade, ou seja, os elementos que em determinado processo são estudados podem ser de três tipos, como descreve Montgomery (2001: físicos (peso, tensão, viscosidade, comprimento), sensoriais (cor, aparência, tato) e orientação temporal (fiabilidade, durabilidade).

Estas características da qualidade podem interagir com o processo isoladamente ou em conjunto com as diversas dimensões da qualidade, assim como podem estar relacionadas entre si, caso em que os dados são autocorrelacionados.

Para definir as características da qualidade como as que devem ser analisadas, deve-se à priori fazer uma análise minuciosa ao processo produtivo, partindo o processo produtivo em pequenos processos para ser mais fácil a escolhas das características críticas/ processos mais críticos. Após este levantamento, é pertinente a construção, por exemplo, dos chamados planos de controlo. Nos planos de controlo identificam-se as características suscetíveis de se estudar estatisticamente, com informações de dimensões de amostras, especificações técnicas, nível de frequência de amostragem e equipamento de medição.

O *SPC*, mais precisamente no que concerne à abordagem tradicional, baseia-se na análise da informação obtida que é avaliada em representações gráficas, designadas por cartas de controlo que permitem avaliar estabilidade dos processos.

Esta técnica é extremamente importante para qualquer indústria. A sua aplicação é relativamente simples, podendo ser aplicada por qualquer pessoa que interage com o processo. Se a ferramenta for aplicada nestas circunstâncias, a consequência é redução do tempo com que se consegue detetar mudanças no comportamento do processo, permitindo a tomada de ações corretivas atempadamente, e evitando assim maiores prejuízos.

Segundo Doty (1996), o *SPC* permite identificar possíveis fontes de ocorrência de variações nos processos: equipamento, matéria-prima, inspeção de qualidade, mão-de-obra, meio ambiente, métodos, metrologia, engenharia e gestão (sistema ou erros organizacionais).

Estas fontes podem resultar em causas de variação que se dividem em causas especiais e causas comuns. As causas especiais provocam um padrão de variações que varia com o tempo de forma imprevisível. As causas comuns produzem uma variação controlada com um padrão estável e consistente ao longo do tempo Wheeler (1995).

Relativamente às fontes de variação que podem afetar um processo, Shewhart classificou-as em dois tipos:

- **Causas comuns**: são causas aleatórias, variações que afetam os processos, mesmo quando estes estão sob controlo estatístico e que caracterizam os dados aleatórios de um processo segundo uma distribuição de probabilidade. As causas comuns têm por norma, uma probabilidade baixa de ocorrerem, uma vez que são normalmente controladas através de decisões da gestão de topo. As decisões tomadas podem alterar o processo do sistema, exemplos disso são: instalação de novos equipamentos, programa de formação e treino dos colaboradores e alteração dos fornecedores);
- **Causas especiais**: são causas esporádicas, variações que afetam mais o desempenho do processo quando comparadas com as causas comuns. Quando aparece uma causa especial diz-se que o processo não está sob controlo estatístico. Estas causas são normalmente detetadas por operacionais do processo específico em causa e são removidas de imediato.

No entanto, é indispensável para uma boa gestão do processo, dar formação a todos os colaboradores nos diversos níveis, disputando o espírito de melhoria contínua dos processos e mecanismos eficientes de combate à variabilidade destes.

O Controlo Estatístico do Processo, *SPC*, pode ser aplicado a qualquer processo utilizando sete ferramentas principais:

- Histograma;
- Folhas de Registo e Verificação;
- Diagrama Causa e Efeito;
- Fluxograma;
- Diagrama de Dispersão;
- Diagrama de Pareto;
- Cartas de Controlo.

Segundo Montgomery (2005), as ferramentas supracitadas constituem uma parte importante do Controlo Estatístico do Processo e para uma melhoria contínua da qualidade e da produtividade.

O principal objetivo do Controle Estatístico do Processo é detetar a ocorrência de causas especiais de variação no processo, para que se possa investigar e introduzir ações corretivas, permitindo reduzir o fabrico de unidades não conformes. As ferramentas que melhor se adequam a este objetivo são as cartas de controlo. Estas permitem uma monitorização do processo em tempo real, podem ser utilizadas para estimar os parâmetros do mesmo e conduzir a estudos da capacidade do processo (Montgomery, 2005). As cartas de controlo são determinantes na redução da variabilidade.

### 2.3. PRINCÍPIOS DAS CARTAS DE CONTROLO

As cartas de controlo são a principal ferramenta utilizada no *SPC*, representam a base de apoio ao controlo da qualidade de um processo. Os principais objetivos subjacentes às cartas de controlo são: monitorizar o processo, estimar parâmetros do processo, reduzir a quantidade de produtos fora das especificações e, consequentemente, os custos de produção. Quando as cartas de controlo são utilizadas nas condições apropriadas, permitem ao operador um controlo contínuo do processo, levando a que se produza de forma consistente, com eficiência, com qualidade e com custos adequados.

Segundo Ledolter e Burril (1999), as cartas de controlo são um método gráfico para determinar se o processo é estável, isto é, se o processo está sob controlo estatístico durante um período de tempo.

O intuito principal e aspeto típico de uma carta de controlo é uma representação gráfica de valores, que ilustram a evolução de uma estatística ( $\omega$ ) ao longo do tempo, referente a uma determinada característica da qualidade. Na carta de controlo tradicional, apresentada na Figura 2.1, observa-se uma linha central (*LC*) representativa do valor médio da estatística  $\omega$ , e duas linhas simetricamente colocadas acima e abaixo da linha central, designadas como limite superior (*LSC*) e limite inferior (*LIC*) de controlo. A área delimitada pelos limites de controlo define a variação aleatória do processo. Quando todos os pontos estiverem contidos entre estes limites, assume-se que o processo está sob controlo estatístico. Se forem detetados pontos fora dos limites de controlo, ou se por outro lado existirem evidências de sequências especiais, sistemáticas e não aleatórias, significa que o processo está fora de controlo, sujeito a causas especiais de variação. Quando as causas especiais aparecem numa carta de controlo, é de extrema importância analisar/investigar a sua origem, para se tomar medidas corretivas

de modo a eliminar causas responsáveis por tal ocorrência, com intuito de melhorar o processo produtivo e não se cair em perdas produtividade e custos.

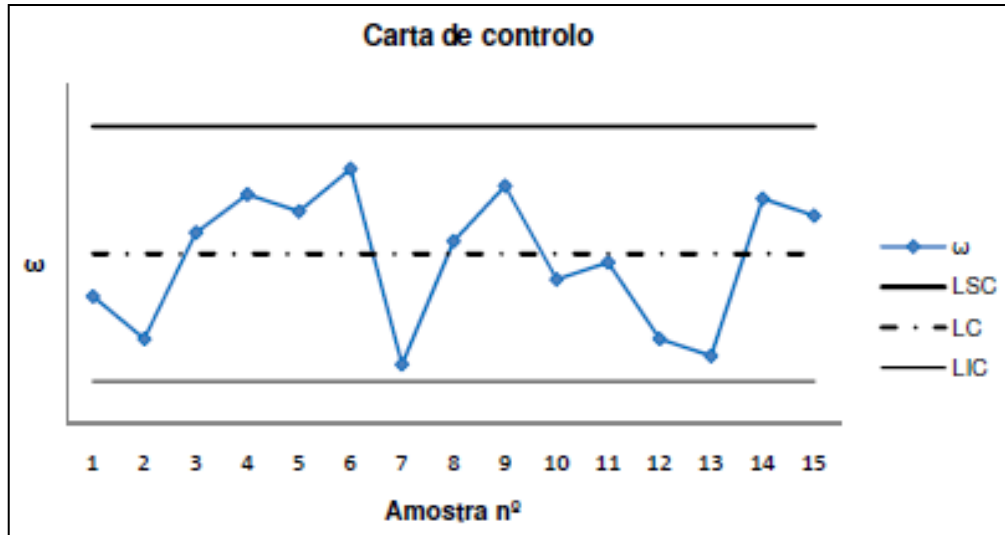


Figura 2.1 - Carta de Controle

Um modelo geral de uma carta de controle pode ser descrito. Considerando-se que os valores da estatística ( $\omega$ ), seguem uma distribuição Normal, com média  $\mu_\omega$  e desvio padrão  $\sigma_\omega$ , a linha central, e os limites superior e inferior de controle são dados por (Montgomery e Runger, 2006):

$$LSC_\omega = \mu_\omega + L\sigma_\omega \quad (2.1)$$

$$LC = \mu_\omega \quad (2.2)$$

$$LIC_\omega = \mu_\omega - L\sigma_\omega \quad (2.3)$$

em que  $L$  representa a distância dos limites de controle à linha central, e normalmente o valor usual é de  $L = 3$  ( $\alpha = 0,27\%$ ). Esta teoria geral de carta de controle foi proposta por Dr. Walter S. Shewhart, pelo que as cartas de controle desenvolvidas de acordo com estes princípios são habitualmente denominadas por Cartas de Controle de Shewhart.

A distância dos limites de controle à linha central é de  $\pm 3\sigma_\omega$ , o que é análogo a afirmar que, considerando que as observações são Normalmente distribuídas, a probabilidade de qualquer ponto se encontrar dentro dos limites de controle é de 99,73%. Assim, a probabilidade de um

ponto pertencente à distribuição de  $\omega$ , exceder um dos limites de controlo quando o processo se encontra sob controlo estatístico, é de 0,27%. Quando um ponto transpõe um dos limites de controlo, indica uma causa especial de variação, logo esse ponto acusa que os valores da estatística não pertencem à distribuição considerada.

A análise relativa ao processo estar sob controlo estatístico ou não, está fortemente relacionada com um teste de hipóteses (Montgomery, 2001). A hipótese a ser estudada  $H_0$ , será assumida como verdadeira a menos que os dados da amostra apresentem evidência contrária. Assim, no caso do controlo estatístico é como se fizesse um teste de hipóteses a cada ponto estar ou não fora do intervalo definido pelos limites de controlo.

No que concerne ao teste de hipóteses, existem erros associados ao interpretar as cartas de controlo, um exemplo é a probabilidade de 0,27% que está associada à probabilidade de um ponto pertencer à distribuição estatística que se está a controlar e que ao mesmo tempo está fora do intervalo definido pelos limites de controlo. Esta probabilidade pode ser designada pelo risco do produtor, isto é, a probabilidade de se cometer um erro do tipo I ( $\alpha = P(\text{Rejeitar } H_0 | H_0 \text{ Verdadeira})$ ).

Os riscos associados à análise das cartas de controlo podem ser do tipo I ou tipo II. O risco tipo I verifica-se quando se conclui que um processo está fora de controlo, quando na verdade este está sob controlo, também denominado de falso alarme. Por outro lado, o risco tipo II acontece quando se considera o processo sob controlo, quando este está fora de controlo. Desta forma, quando se afastam os limites de controlo da linha central, diminui a possibilidade de ocorrência de um erro tipo I, mas aumenta a probabilidade de ocorrência de um erro tipo II.

Algumas referências na área, como é o caso de Montgomery (2005), Quesenberry (1997) e Pereira e Requeijo (2008) consideram que o processo de implementação das cartas de controlo deve ser dividido em duas fases: Fase I e Fase II. A Fase I inicia-se quando se desconhecem os parâmetros do processo, procede-se à recolha dos dados e constrói-se a carta de controlo. Nesta fase, quando é verificada uma causa especial, deve-se procurar identificar a origem dessa causa e eliminar o ponto responsável. Uma vez eliminado o ponto, deve-se recalcular os limites de controlo e construir uma nova carta de controlo. Contudo, pode-se verificar que ao longo da carta existem vários pontos fora de controlo, nestes casos deve-se investigar e corrigir as causas que conduziram a essa situação e posteriormente deve-se recolher novos dados. Quando finalmente o processo se encontrar sob controlo estatístico

estima-se os parâmetros do processo ( $\mu$  e  $\sigma$ ) e procede-se à análise da capacidade do mesmo. Concluída a Fase I e verificada a capacidade do processo para produzir de acordo com as especificações, segue-se para a Fase II. A Fase II, tem como objetivo a monitorização do processo. Nesta fase a ocorrência de uma causa especial de variação deve ser analisada e posteriormente devem ser aplicadas medidas corretivas.

Já, Palm (2000), discorda com esta divisão do *SPC*, sugerindo que este se desenvolva em três Estados:

- Estado A – estado inicial, início do processo, fase retrospectiva;
- Estado B - estado intermédio, fase de melhoria;
- Estado C - estado final, fase de monitorização do processo.

Muitos estudos foram feitos ao longo dos anos e é notória a potencialidade da aplicação das cartas de controlo à indústria, de seguida apresentam-se alguns factos que segundo Montgomery (2005) contribuíram para o seu sucesso:

- As cartas de controlo são uma técnica para a melhoria da produtividade. Uma correta aplicação das cartas de controlo, permite reduzir desperdícios e evitar o “fazer de novo”, o que reflete num aumento da capacidade de produção e na redução de custos;
- As cartas de controlo são eficazes na prevenção de defeitos. Ajudam o processo a manter-se estável, sob controlo, o que é consistente com a filosofia “fazer bem à primeira”, reduzindo nos custos de seleção e reprodução;
- As cartas de controlo evitam ajustes no processo desnecessário, permitindo distinguir entre “ruído de fundo” (causas comuns) e uma “variação anormal” (causas especiais). Quando não existem as cartas de controlo, devido ao “ruído de fundo”, são efetuados ajustes no processo desnecessários. Os “ruídos de fundo” podem resultar na deterioração do desempenho do processo;
- As cartas de controlo fornecem informação de diagnóstico através do padrão dos dados. A análise e o conhecimento dos padrões presentes numa carta de controlo, permite implementar correções no processo e melhorar o seu desempenho;
- As cartas de controlo fornecem informações sobre os parâmetros do processo e possibilitam a análise da capacidade do processo.

## 2.4. MEDIDAS DE DESEMPENHO DAS CARTAS DE CONTROLO

Um conceito importante associado às cartas de controlo é a noção de *ARL* (*Average Run Length*). Segundo Montgomery (2005), a partir do conceito da probabilidade de 0,27%, valor definido por Shewhart, e conseqüente risco do produtor pode-se introduzir a noção de *ARL*, que representa o número médio de pontos representados numa carta antes de ocorrer um ponto fora do intervalo definido por [*LIC*; *LSC*].

Quando um processo está sob controlo estatístico deseja-se que o  $ARL_{Em\ controlo}$  seja o maior possível para que a ocorrência de falsos alarmes não seja frequente. Pelo contrário, quando o processo não está sob controlo, interessa detetar mais rapidamente as causas especiais e como tal o  $ARL_{Fora\ de\ controlo}$  deverá ser o menor possível.

Sendo o  $ARL_{Em\ controlo}$ , igual a 370, significa que de 370 em 370 pontos irá ocorrer um ponto fora dos limites. A equação será dada por:

$$ARL_{Em\ controlo} = \frac{1}{\alpha} \quad (2.4)$$

Com um desvio padrão  $SDRL_{Em\ controlo}$  igual a:

$$SDRL_{Em\ controlo} = \frac{\sqrt{1-\alpha}}{\alpha} \quad (2.5)$$

onde  $\alpha$  representa a probabilidade de se cometer um erro do tipo I. Nesta situação, considera-se erradamente que um ponto não pertence à distribuição que se está a controlar, ocorrendo assim a existência do chamado falso alarme.

Por outro lado, um  $ARL_{Fora\ de\ controlo}$  já será dado pela equação:

$$ARL_{Fora\ de\ controlo} = \frac{1}{1-\beta} \quad (2.6)$$

em que o respetivo desvio padrão  $SDRL_{Fora\ de\ controlo}$  é dado por:

$$SDRL_{Fora\ de\ controlo} = \frac{\sqrt{\beta}}{1-\beta} \quad (2.7)$$

o valor de  $\beta$  representa a probabilidade de se cometer o erro do tipo II, isto é, ( $\beta = P(\text{Não Rejeitar } H_0 | H_0 \text{ Falsa})$ ).

A noção de *ARL*, ou seja, esta medida de desempenho *ARL*, é normalmente a ferramenta mais poderosa para comparar o desempenho tanto de cartas univariadas como de cartas multivariadas. Desta forma, para se ter o *ARL* sob controlo estatístico, deve-se ajustar os limites de controlo de modo a estabelecer-se uma comparação de *ARL*'s para as várias dimensões de desvio ao processo.

A presente dissertação tem como objetivo desenvolver um estudo comparativo do desempenho de duas cartas de controlo multivariadas com dados autocorrelacionados com a carta que se propõe desenvolver *DMPCA*, assim apresentam-se de seguida algumas referências de estudos comparativos de cartas de controlo multivariadas.

Lowry et al. (1992) apresenta um estudo comparativo de 4 cartas de controlo multivariadas, nomeadamente, a carta  $T^2$  de Hotelling, a carta *MCUSUM* desenvolvida por Crosier (1988), a carta multivariada *CUSUM* (*MC1*) desenvolvida por Pignatiello e Runger, em 1990, a carta *MEWMA* no qual se utiliza a matriz covariância desenvolvida por MacGregor e Harris, em 1990 e a carta *MEWMA* no qual se utiliza a matriz covariância desenvolvida por Tsui e Woodall, em 1991, tendo como medida de desempenho o *ARL*.

Já Javaheri e Houshmand (2001) apresentaram um estudo comparativo do desempenho de 5 métodos de controlo da qualidade multivariada, nomeadamente, a carta de Hotelling  $T^2$ , a carta multivariada de Shewhart desenvolvida por Houshmand et al., em 1998, a Análise Discriminante desenvolvida por Murphy, em 1987, a decomposição de  $T^2$  desenvolvida por Mason, Tracy e Young (1995) e a carta *MRRC* (*Multivariate Ridge Residual*) desenvolvida por Houshmand e Javaheri, em 1998. O estudo é realizado através a metodologia de simulação Monte Carlo e utiliza como medida de desempenho o *ARL*.

Ghute e Shirke (2008) publicaram um artigo onde apresentam um estudo comparativo do desempenho da carta  $T^2$  sintética com a carta  $T^2$  desenvolvida por Hotelling, e com a carta  $T^2$  com regras de funcionamento suplementar, desenvolvida por Aparisi et al. (2004), através do *ARL*. Os valores de *ARL* das cartas  $T^2$  sintética e  $T^2$  de Hotteling são determinados através do desenvolvimento de um programa computacional usando o software MATLAB. Já os valores da carta  $T^2$  com regras suplementares são dados por Apirisi et al. (2004). Desta forma, para medir o desempenho das cartas de controlo existe a necessidade de detetar causas especiais e otimizar a interpretação das cartas de controlo definiram-se regiões nas cartas de controlo.

Na generalidade da literatura para detetar causas especiais de variação utilizam-se as regras da *Western Electric*, em que são sugeridas seis zonas de divisão das cartas denominadas pelas zonas A, B e C de cada lado da linha central (Figura 2.2).

Pitt (1994) apresenta uma série de critérios para deteção de causas especiais, onde se destaca os critérios estabelecidos na norma ISO 8258:1991 (Tabela 2.1). Esta norma fundamenta-se nas regras da *Western Electric*.

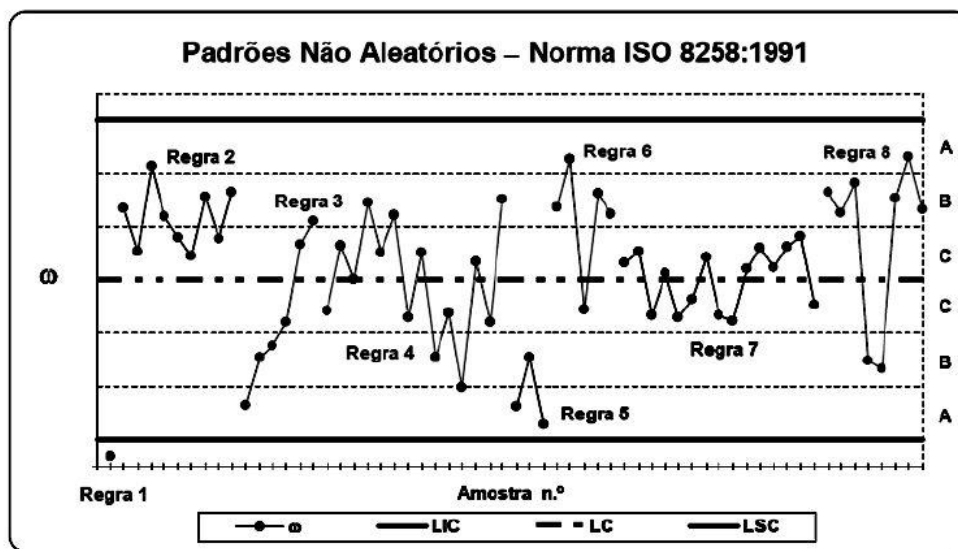


Figura 2.2 - Regras de deteção de causas especiais - Norma ISO 8258:1991 (Pereira e Requeijo, 2008)

Tabela 2.1 - Regras da norma ISO 8258:1991 para deteção de causas especiais

Regras da norma ISO 8258:1991 para deteção de causas especiais	
Regra 1	Um ponto fora do intervalo definido pelos limites de controle
Regra 2	Nove pontos consecutivos do mesmo lado da linha central
Regra 3	Seis pontos consecutivos no sentido descendente ou ascendente
Regra 4	Catorze pontos consecutivos crescendo e decrescendo alternadamente
Regra 5	Dois de três pontos consecutivos na zona A do mesmo lado da linha central
Regra 6	Quatro de cinco pontos consecutivos na zona B ou A do mesmo lado da linha central
Regra 7	Quinze pontos consecutivos na zona C acima e abaixo da linha central
Regra 8	Oito pontos consecutivos de ambos os lados da linha central, sem nenhum na zona C

## 2.5. RECOLHA DE DADOS

Para a construção das cartas de controlo é necessária proceder-se a uma perfeita recolha de dados, caso contrário, o sucesso do estudo poderá ser colocado em causa, uma vez que a análise pode estar enviesada e não representar corretamente a realidade, fomentando conclusões erradas do processo em estudo.

A construção das cartas de controlo deve basear-se numa recolha de dados de acordo com os subgrupos racionais (amostras), título dado por Shewhart. Estes subgrupos racionais devem ser selecionados com o intuito de maximizar a probabilidade de deteção de diferenças entre subgrupos, e por outro lado minimizar a probabilidade de deteção de diferenças dentro dos subgrupos. Para tal, existem diversos métodos de seleção de amostras. Montgomery (2005) refere que deverá ser implementada uma carta de controlo por cada máquina de um processo, assim como para o caso de fornecedores, uma carta de controlo para cada um deles por processo, evitando desta forma a mistura de populações. Uma carta de controlo deve conter dados, sempre que possível, de subgrupos homogéneos. Para tal, é espectável que as unidades tenham sido produzidas consecutivamente e de forma análoga.

Mas, em algumas situações, a mesma carta de controlo pode ser utilizada, não sendo necessário definir uma carta de controlo distinta para cada máquina ou para cada fornecedor, desde que estudos preparatórios tenham sido realizados, baseados em análises de variância sobre os parâmetros das características da qualidade em estudo ou em testes de hipóteses, por exemplo (Pereira e Requeijo, 2008).

Quando a complexidade e especificidade dos processos é elevada, pode não ser possível criar subgrupos. Quando tal acontece, a análise deve ser realizada com base em observações individuais.

O controlo da dispersão do processo, com apenas uma observação, é impraticável. Assim, para colmatar esta lacuna, o procedimento mais utilizado é determinar a medida de dispersão com base na observação recolhida nesse instante e uma ou mais observações recolhidas em instante(s) imediatamente anterior(es), construindo uma “amostra fictícia” que é denominada de amplitude móvel.

No entanto, esta condição torna de certa forma impossível o controlo da dispersão do processo pelo método tradicional.

Definidos os subgrupos, é fundamental definir o número de amostras ( $m$ ) a serem recolhidas na Fase I, a frequência com que estas são recolhidas e a respetiva dimensão das mesmas ( $n$ ). Na Fase I é necessário ter em atenção o número de amostras que vão ser recolhidas para a sua realização, uma vez que vão ser estimados os parâmetros e os limites de controlo. Por isso, é necessário recolher um número suficiente de amostras de tal forma que as observações sejam representativas do processo.

Vários autores referem que no caso em que os dados são observações individuais, devem ser recolhidas no mínimo 100 observações. Já Quesenberry (1997) estabelece uma relação entre o número de amostras e a dimensão da amostra, equação abaixo apresentada. Afirma também que quando são utilizadas observações individuais o número mínimo de dados a recolher na Fase I deverá ser de 300 observações.

$$m \geq \frac{400}{(n-1)} \quad (2.8)$$

Por norma, numa fase inicial, a recolha dos dados para a amostra tende a ser frequente em intervalos curtos e regulares, de forma a verificar se ocorrem muitas alterações no processo, e consequentemente ocorrência de causas especiais. No entanto, a frequência da amostragem deve sim ser realizada em intervalos de tempo mais longos e regulares, com o objetivo que as observações recolhidas possam espelhar o melhor possível o comportamento do processo para mais fácil deteção das pequenas falhas no processo.

Quando o processo se encontra estatisticamente controlado, a frequência de recolha tende a diminuir, desde que continue a ser representativa do processo (Pereira e Requeijo, 2008). No entanto, a frequência de amostragem é tanto melhor quanto maior for o seu valor. O mesmo acontece no que concerne à dimensão da amostra, um aumento na dimensão da amostra reflete-se numa maior sensibilidade na deteção de alterações nos parâmetros do processo. Contudo, geralmente esta prática não é exequível, uma vez se torna incomportável por motivos económicos ou operacionais.

## 2.6. CONDIÇÕES DE APLICABILIDADE DAS CARTAS DE CONTROLO

As cartas de controlo têm como requisitos que os dados sejam independentes e identicamente distribuídos com distribuição Normal, média  $\mu$  e variância  $\sigma^2$ . Nos casos em que os dados não cumprem tais condições e se prossegue com o controlo estatístico, pode-se chegar ao ponto de retirar conclusões erradas no que diz respeito à estabilidade dos processos. Na prática, muitas vezes a suposição de independência é violada pelo facto das variáveis serem autocorrelacionadas, isto é, o valor da variável num dado instante depende, em parte, de um ou de vários valores anteriores da mesma variável. Assim, as cartas de controlo não funcionam corretamente pois se for assumida a independência dos dados, quando na realidade existe a presença de autocorrelação, tanto os limites de controlo como os parâmetros do processo serão estimados incorretamente. Quando a autocorrelação é positiva, o desvio padrão é subestimado, o que resulta num estreitamento dos limites de controlo originando um aumento do número de falsos alarmes. Segundo Wieringer (1999), quando a autocorrelação é negativa, o desvio padrão é sobrestimado, produzindo um efeito oposto, isto é, aumento dos limites de controlo seguido dum perda de sensibilidade às variações da média do processo.

A implementação do SPC só se deve realizar, quando os pressupostos são cumpridos, ou seja, só aplicar quando os dados são independentes e normalmente distribuídos, só desta forma é possível à posteriori controlar os dados.

Desta forma, de seguida realiza-se uma abordagem a métodos para identificar a aleatoriedade, a independência e a normalidade dos dados.

### 2.6.1. Aleatoriedade dos dados

Para a verificação da aleatoriedade dos dados Pereira e Requeijo (2008) mencionam os seguintes métodos: teste de Sequências, teste de Sequências Ascendentes ou teste de Sequências Descendentes e o teste Modificado do Quadrado Médio das Diferenças Significativas. Quando se verifica a não aleatoriedade dos dados esta pode dever-se a imensos fatores, é um dos exemplos a correlação existente entre observações consecutivas ou desfasadas no tempo.

### 2.6.2. Autocorrelação dos dados

Dizer que os dados são autocorrelacionados é o mesmo que afirmar que estes não são independentes, sendo em muitos casos a principal causa de não aleatoriedade dos dados. No entanto, pode-se afirmar que um processo pode originar dados intrinsecamente autocorrelacionados devido à sua dinâmica natural nos processos em causa.

A problemática de correlação entre os dados surge muitas vezes quando os dados são gerados em pequenos intervalos de tempo, sendo que Montgomery (2001) aconselha que a autocorrelação seja obrigatoriamente verificada em todos os processos, onde diversas características apresentam inércia, isto porque a ocorrência de um dado processo num dado instante de tempo pode alterar todo o processo adiante, podendo repetir-se de  $k$  em  $k$  observações a repetição de determinado padrão. Esta repetição pode ser a causa da ocorrência de pontos fora do intervalo definido por  $[LIC; LSC]$  e o processo estar sob controlo estatístico, ou seja, ocorrência de falsos alarmes. Portanto o estudo da autocorrelação é de extrema importância nos processos cujo procedimento tenha influência no mesmo processo passado um período de tempo.

Sempre que o pressuposto da independência dos dados é violado, isto é, os dados apresentados são autocorrelacionados, aconselha-se a aplicação do método de *Box e Jenkins* para modelar o comportamento dinâmico dos dados.

#### **Metodologia de Box-Jenkins**

A metodologia de *Box-Jenkins* é uma ferramenta de fácil aplicação para previsão de variáveis baseadas em séries temporais. Segundo Miranda (2002), a metodologia de *Box-Jenkins* baseia-se no ajuste de modelos matemáticos às observações recolhidas, denominados por modelos *ARIMA* (*Auto Regressive Integrated Moving Average*), onde a diferença entre os valores gerados pelos modelos e os valores observados resulte em séries de resíduos, de comportamento aleatório em torno de zero, denominados de ruído branco.

Os modelos *ARIMA* são originados pela combinação de três componentes denominados de “filtros: o componente auto-regressivo (*AR*), o filtro de integração (*I*) e o componente de médias móveis (*MA*). Qualquer série pode ser modelada pelos três filtros supracitados ao

mesmo tempo, ou apenas por um subconjunto deles, podendo assim originar várias alternativas de modelos.

A metodologia de *Box-Jenkins* apresentada por Box, Jenkins e Reinsel, na sua obra mais recente, Box et al. (2008), tem como base três etapas principais:

- 1) Identificação do Modelo;
- 2) Estimação de Parâmetros;
- 3) Avaliação do Modelo.

De seguida, faz-se uma descrição de cada uma das etapas acima descritas.

### **1. Identificação do Modelo:**

A etapa de identificação do modelo consiste em identificar qual o modelo *ARIMA* ( $p, d, q$ ) que melhor representa o comportamento do processo em causa. Inicialmente, o mais importante é descobrir quais são os filtros do modelo *ARIMA* que fazem parte do processo gerador da série a ser estudada (*AR, I e MA*), assim como os valores de  $p$ ,  $d$  e  $q$  que identificam a ordem dos respetivos filtros.

Existem vários métodos de deteção de autocorrelação dos dados, o método mais utilizado e mais popular no entanto, é a construção do correlograma, que tem por base o cálculo da Função de Autocorrelação (*FAC*) e da Função de Autocorrelação Parcial (*FACP*). Através dos correlogramas resultantes destas duas funções, consegue-se obter uma primeira aproximação do modelo a ser utilizado.

### **Função de Autocorrelação (FAC)**

Quando o valor de uma determinada variável num dado instante depende em parte do(s) valor(es) anterior(es) da mesma variável, diz-se que esta variável é autocorrelacionada. Para medir o grau de dependência de  $k$  observações utiliza-se o coeficiente de autocorrelação de

desfasamento  $k$  ( $\rho_k$ ). Ao conjunto de coeficientes de autocorrelação ( $\rho_k$ ), com  $k = 0, 1, 2, 3$ , designa-se por Função de AutoCorrelação (*FAC*).

$$\rho_k = \frac{\text{Cov}(X_t - X_{t+k})}{\text{Var}(X)} = \frac{\gamma_k}{\gamma_0}, \quad k = 0, 1, 2, \dots \quad (2.9)$$

$\text{Cov}(X_t - X_{t+k})$  – Covariância entre as observações com desfasamento  $k$ ;

$\text{Var}(X)$  - Variância de  $X$ ;

$\gamma_k$  – autocovariância de desfasamento  $k$ ;

$\gamma_0$  - autocovariância de desfasamento  $k = 0$ ;

Segundo Box et al. (2008), o valor estimado mais satisfatório de  $\rho_k$  é dado por  $r_k$ .

$$r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n-1} (X_t - \bar{X})^2}, \quad k = 0, 1, 2, \dots \quad (2.10)$$

O conjunto formado pelos coeficientes de correlação estimados  $r_k$  é designado por Função de Autocorrelação Estimada (*FACE*).

A representação gráfica da *FACE* em função de  $k$  permite averiguar se os dados da variável  $X$  são autocorrelacionados, sendo esta representação gráfica designada por correlograma.

O método de averiguação num correlograma é verificar se todos os valores estão contidos entre um determinado intervalo de confiança, calculado com base no valor esperado ( $E(r_k) = 0$ ) e na variância de  $\text{Var}(r_k)$ .

$$\text{Var}(r_k) \approx \frac{1}{N} \sum_{v=-\infty}^{+\infty} (\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2) \quad (2.11)$$

em que como estimador de  $\text{Var}(r_k)$  define-se por:

$$\hat{v} = \text{Var}(r_k) \approx \frac{1}{N} (1 + 2 \sum_{v=1}^{k-1} r_k^2) \quad (2.12)$$

O intervalo de confiança, para um nível de significância  $\alpha$ , é definido pela equação 2.13, sendo que  $r_k$  segue uma distribuição normal reduzida.

Caso todos os valores estejam contidos dentro do referido intervalo os dados não são previsivelmente autocorrelacionados.

### Função de Autocorrelação Parcial (FACP)

A função autocorrelação parcial é utilizada para ajudar a escolher o modelo que melhor descreve o processo, quando o comportamento da função de autocorrelação (*FAC*), não é conclusiva.

Define-se o coeficiente de autocorrelação parcial de ordem  $k$  ( $\phi_{kk}$ ) pela correlação entre  $X_t$  e  $X_{t+k}$  com os efeitos das observações ( $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ ) removidas.

Sendo o conjunto de valores de  $\phi_{kk}$  caracterizado por Função de AutoCorrelação Parcial (*FACP*), em que  $k = 1, 2, \dots$

Para modelos  $AR(p)$  (autorregressivos de ordem  $p$ ), a autocorrelação parcial é obtida recorrendo às equações de Yule-Walker (Box et al., 2008).

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.13)$$

Por outro lado, também se pode recorrer à notação matricial:

$$\phi_k = P_k^{-1} \rho_k \quad (2.14)$$

Os coeficientes de autocorrelação parcial não são conhecidos, têm de ser estimados, pelo que os valores  $\rho_k$  são substituídos pelos valores de  $r_k$ , obtendo-se  $\hat{\phi}_{kk}$ . O conjunto formado pelos coeficientes de autocorrelação parcial estimados,  $\hat{\phi}_{kk}$ , denomina-se por Função de Autocorrelação Parcial Estimada (*FACPE*).

Nem sempre é fácil distinguir, através dos correlogramas, qual o processo que está em questão (*AR*, *MA* ou *ARMA*). Na Tabela 2.2 estão descritas as características das *FAC* e *FACP* dos modelos *AR*(*p*), *MA*(*q*) e *ARMA*(*p*, *q*), segundo Box et al. (2008).

Tabela 2.2 - Comportamento das Funções de Autocorrelação e Autocorrelação Parcial para os diferentes processos

Processo	<i>FAC</i>	<i>FACP</i>
<i>AR</i> ( <i>p</i> )	<b><u>Infinita:</u></b> Decresce exponencialmente para zero e/ou decresce segundo uma sinusoidal amortecida ( <i>Trails off</i> )	<b><u>Finita:</u></b> Decai bruscamente para zero depois do <i>lag p</i> ( <i>Cuts off after lag p</i> )
<i>MA</i> ( <i>q</i> )	<b><u>Finita:</u></b> Decai bruscamente para zero depois do <i>lag q</i> ( <i>Cuts off after lag q</i> )	<b><u>Infinita:</u></b> Maioritariamente decresce exponencialmente para zero e/ou decresce segundo uma sinusoidal amortecida ( <i>Tails off</i> )
<i>ARMA</i> ( <i>p</i> , <i>q</i> )	<b><u>Infinita:</u></b> Decresce exponencialmente para zero e/ou decresce segundo uma sinusoidal amortecida, após o primeiro <i>lag p</i> ou <i>q</i> ( <i>Tails off</i> )	<b><u>Infinita:</u></b> Maioritariamente decresce exponencialmente para zero e/ou decresce segundo uma sinusoidal amortecida, após o primeiro <i>lag p</i> ou <i>q</i> ( <i>Tails off</i> )

A função de autocorrelação (*FAC*) pode dificultar a escolha do melhor modelo para descrever o processo pois veja-se que num modelo *AR*(*p*), a *FAC* decresce sem nunca atingir o valor zero, enquanto num modelo *MA*, a *FAC* decresce e atinge o valor zero a partir de determinada ordem de desfasamento.

Para a escolha do melhor modelo a aplicar o melhor modelo deve ser parcimonioso, isto é, deve-se escolher o modelo que apresentar um menor conjunto de parâmetros a ser estimado.

## **2. Estimação dos parâmetros:**

Depois de terem sido determinados os valores de  $p$  e  $q$ , é necessário estimar os parâmetros autorregressivos,  $\phi$ , e os  $q$  parâmetros  $\theta$  (parâmetros da média móvel) do modelo selecionado. Os métodos mais utilizados para estimar os parâmetros são o método dos mínimos quadrados ou o método da máxima verossimilhança.

## **3. Avaliação do modelo:**

Após a escolha do melhor modelo *ARIMA* e estimados os seus parâmetros, o passo seguinte é avaliar se a seleção do modelo foi a mais acertada, verificando se este se ajusta aos dados da série temporal do processo em causa. Um dos métodos mais populares na literatura, é o teste de Box-Ljung. Este teste identifica se o resíduo é um ruído branco, isto é, se o resíduo é independente e identicamente distribuído. Caso o resíduo não apresente esta característica, conclui-se que o modelo escolhido não é o mais indicado para a série em causa, tornando-se necessário ir ao encontro de um novo modelo que descreva de melhor forma o processo.

Para se aplicar a metodologia de Box-Jenkins é fundamental primeiro identificar a série que se quer analisar e remover a não estacionaridade, uma vez que os modelos introduzidos por Box e Jenkins descrevem exclusivamente séries estacionárias, ou seja, com média e variância constantes no tempo. Isto permite fixar os parâmetros do modelo válidos para previsão do futuro a partir do passado.

Desta forma, vai apresentar-se de seguida os modelos para séries estacionárias, modelos *ARMA*, e para séries não-estacionárias, modelos *ARIMA*. Os modelos para séries estacionárias são ajustados à série original e os modelos para séries não-estacionárias são transformadas pelo método das diferenças de ordem  $d$ , isto é, cujas séries originais são séries não-estacionárias.

### **Processos estacionários – Modelos *ARMA* ( $p, q$ )**

A combinação de  $p$  e  $q$  originam um grande número de modelos matemáticos para descrever séries temporais.

Processos autorregressivos – AR(p)

Os processos autorregressivos de ordem  $p$ ,  $AR(p)$  ou  $ARMA(p, q)$  podem ser apresentados pela seguinte expressão:

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.15)$$

Onde:

$\xi$  – Parâmetro que determina a média do processo;

$\phi_i$  - Parâmetro da componente autorregressiva, com  $i = 1, 2, 3, \dots, p$ ;

$X_{t-p}$  – Valor variável no instante  $t - p$ ;

$\varepsilon_t$  – Valor do resíduo no instante  $t$ ;

Por exemplo, para um processo autorregressivo de primeira ordem  $AR(1)$ , este pode ser descrito da seguinte forma

$$X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t \quad (2.16)$$

Em que a média do processo  $AR(1)$  é dada por

$$\mu = \frac{\xi}{1-\phi_1} \quad (2.17)$$

a função de autocorrelação (FAC) por

$$\rho_k = \phi_1^k, \text{ com } k = 0, 1, \dots \quad (2.18)$$

Processos Média Móvel–  $MA(q)$ 

Os processos de média móvel de ordem  $q$ ,  $MA(q)$  ou  $ARMA(q,0)$  podem ser descritos através da seguinte expressão:

$$X_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \quad (2.19)$$

Onde:

$\mu$  – Média do processo;

$\theta_i$  - Parâmetro da componente média móvel, com  $i = 1, 2, 3, \dots, q$ ;

$\varepsilon_{t-q}$  – Valor do resíduo no instante  $t - q$ ;

Se se estiver perante o processo de média móvel de primeira ordem  $MA(1)$ , este pode ser representado pela seguinte forma:

$$X_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad (2.20)$$

Em que a função de autocorrelação (FAC) é dada por

$$\rho_k = \begin{cases} \frac{-\theta}{1+\theta^2}, & k = 1 \\ 0, & k > 0 \end{cases} \quad (2.21)$$

Processos autorregressivos de Média Móvel –  $ARMA(p, q)$ 

Os processos autorregressivos de média móvel,  $ARMA(p, q)$ , podem ser apresentados pela seguinte expressão:

$$X_t = \xi + \phi_1X_{t-1} + \phi_2X_{t-2} + \dots + \phi_pX_{t-p} - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} + \varepsilon_t \quad (2.22)$$

Se estivermos perante um processo  $ARMA(1,1)$ , este pode ser representado da seguinte forma

$$X_t = \xi + \phi_1 X_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (2.23)$$

Em que a média é dada pela expressão (2.17) e a função de autocorrelação por

$$\rho_k = \begin{cases} \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\theta_1\phi_1}, & k = 1 \\ \phi_1\rho_{k-1}, & k > 1 \end{cases} \quad (2.24)$$

#### Processos não-estacionários – Modelos $ARIMA(p, d, q)$

Tal como foi descrito anteriormente, quando se está perante séries não-estacionárias, é necessário tornar estas séries em estacionárias, para este caso utiliza-se o método das diferenças de ordem  $d$ . Neste método, para fazer com que as séries se tornem estacionárias é necessário definir a ordem de integração, que não é mais do que o número de diferenças ( $d$ ). Um exemplo de como se chega ao número de diferenças ( $d$ ), é imagine-se por exemplo a série original  $X_t$  que não é estacionária, mas  $Y_t = \Delta X_t = X_t - X_{t-1}$  é estacionária, para estes casos diz-se que  $X_t$  é integrada de primeira ordem ( $d = 1$ ). Deste modo, a ordem de integração da série depende do número de diferenças que esta necessita para se tornar estacionária. Assim, por exemplo se a série original ( $X_t$ ) se tornar estacionária e a série resultante ( $Y_t$ ) for representada por um modelo  $ARMA(p, q)$ , diz-se que  $X_t$  é descrita por um modelo  $ARIMA(p, d, q)$  representada por

$$Y_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.25)$$

Para este tipo de processos, existem, também, os modelos  $ARI(p, d)$  ou  $ARIMA(p, d, 0)$  e  $IMA(d, q)$  ou  $ARIMA(0, d, q)$ .

### 2.6.3. Normalidade dos dados

Para aplicar o controlo estatístico do processo tradicional, é necessário que uma das principais condições seja cumprida, que os dados sigam uma distribuição normal. Ao se assumir que os dados seguem uma distribuição normal sem se analisar, pode-se incorrer em conclusões que não representam a realidade, uma vez que os dados podem só aproximar-se a uma distribuição normal.

Como se sabe e como já foi citado, os limites de controlo das cartas de controlo são baseadas num pressuposto de normalidade ( $\mu \pm 3\sigma_w$ ), o que quer dizer que se estiverem em causa dados não normais, a consequência poderá passar por deteção tardia de causas especiais, o que pode por em causa um processo industrial e consequentemente perda de tempo e custos desnecessários por falta de controlo.

Para verificar a normalidade dos dados, existem vários métodos. De entre os mais conhecidos no controlo estatístico de processo destacam-se o teste do Qui-Quadrado e o teste de Kolmogorov-Smirnov.

Um facto extremamente importante de realçar, é o conceito definido por Shewhart e exposto também por Spedding e Rawlines (1994) e Quesenberry (1997), que em situações em que se pretende fazer um controlo a uma característica cujas amostras tenham dimensão superior ou igual a quatro ( $n \geq 4$ ), não é necessário verificar a normalidade dos dados, pois segundo o Teorema do Limite Central (TLC) a distribuição de medidas de amostras é normal.

Quando a dimensão das amostras é inferior a quatro ( $n < 4$ ), principalmente nas cartas de observações individuais, a distribuição normal dos dados é de extrema importância, no entanto, caso este pressuposto não seja cumprido, os dados devem seguir um dos seguintes métodos para tratar a não normalidade destes:

- Caracterização da distribuição dos dados;
- Método da Variância Ponderada;
- Transformação dos dados, método de Box-Cox ou do Sistema de Distribuições de Johnson (SDJ).

## 2.7. CARTAS DE CONTROLO UNIVARIADAS PARA DADOS AUTOCORRELACIONADOS

A independência dos dados é um pressuposto fundamental na aplicação das cartas de controlo, contudo este nem sempre é possível. Devido a vários fatores muitos processos apresentam autocorrelação. Desta forma, para a monitorização de processos cujos dados apresentam autocorrelação, Montgomery (2005) sugere duas abordagens para a construção de cartas de controlo. A primeira abordagem consiste na utilização das cartas de controlo tradicionais com limites de controlo modificados, tendo como base a autocorrelação que se verifica no processo. A segunda abordagem passa por determinar um modelo matemático *ARIMA* (*Autoregressive Integrated Moving Average*) que se ajuste aos dados autocorrelacionados e, desta forma construir as cartas de controlo com os resíduos ou dos erros de previsão do modelo ajustado.

Existe na literatura inúmeros estudos e trabalhos relativos à temática destas duas abordagens, relativamente à primeira abordagem, apresentam-se de seguida os considerados mais relevantes, que tiveram por base a modificação dos limites das cartas de controlo. Vasilopoulos e Stamboulis em 1978, introduziram as cartas de controlo de limites modificados para processos autorregressivos de segunda ordem, *AR*(2). Já Yang e Hancock em 1990, desenvolveram limites modificados para as cartas  $\bar{X}$ , *R*, *S* e  $S^2$ , com base na função de autocorrelação  $\rho_k$ . Montgomery e Mastrangelo em 1991 aconselharam o uso da estatística *EWMA* como um valor de previsão, de forma a eliminar a autocorrelação dos dados, através de pequenas alterações na média do processo. Na sua proposta definem um procedimento muito simples, que consiste na aplicação de duas cartas em simultâneo, uma carta *EWMA*, aplicada a observações originais, e uma carta de erros de previsão ou resíduos determinados através de uma variável *EWMA*. A aplicação das duas cartas é fundamentada pelos autores, pelo facto da carta de resíduos não permitir a obtenção direta de conclusões acerca dos acontecimentos do processo, complementando-se com a aplicação de uma carta que possibilite a visualização dinâmica do processo. Para além disso, também desenvolveram a carta *MCEWMA* (*Moving Center-line EWMA*) com limites não constantes, que permite analisar, simultaneamente, a evolução do comportamento do processo e detetar causas especiais de variação. Mais recentemente, Zhang (1998) propôs o uso da estatística *EWMA* para processos estacionários, carta *EWMAST* (*EWMA for Stationary Processes*), preferencialmente quando a autocorrelação não é muito elevada, esta carta recorre à função de autocorrelação para modificar os limites de controlo da carta *EWMA*.

No que concerne à segunda abordagem, que consiste em ajustar o melhor modelo matemático *ARIMA* aos dados recolhidos, esta teve uma maior aceitação e foi inclusive alvo de mais estudos por parte dos investigadores que a primeira abordagem.

Entre os autores mais relevantes, há que destacar o trabalho desenvolvido por Alwan e Roberts (1988), que propõem duas novas cartas, a carta *CCC* (*Common-Cause Chart*) e a carta *SCC* (*Special-Cause Chart*). A carta *CCC* tem por base, somente, a construção dos valores previstos, obtidos pelo ajuste dos dados através do modelo *ARIMA* conhecido por Box e Jenkins. A constituição desta carta não contempla a existência de limites de controlo pelo que apenas é usada para estimar o nível atual do processo. Já a carta *SCC* tem por base as cartas de Shewart, é uma carta de resíduos. Alwan e Roberts (1988), conseguiram mostrar com as suas investigações que esta carta é muito mais vantajosa que a carta de Shewart tradicional no que concerne às alterações da média do processo. Para além destes autores, esta última carta, esta também foi investigada por vários outros autores entre eles, Harris e Ross (1991), Wardell et al. (1994), Runger e Willemain (1995), Lin e Adams (1996), Vander Wiel (1996a), Reynolds e Lu (1997), Zhang (1998) e English et al. (2000).

Existem também outras cartas que podem ser utilizadas para monitorizar os resíduos, são as cartas de controlo *CUSUM* e *EWMA*, que permitem detetar rapidamente alterações pequenas e moderadas na média do processo. Tal como nas cartas anteriormente referidas, foram também vários os autores que apresentaram desenvolvimentos para obtenção destas duas cartas: para a carta *CUSUM* temos Harris e Ross (1991), Yashchin (1993), Runger et al. (1995) e Lu e Reynolds (2001) e para carta *EWMA* temos Reynolds e Lu (1997), Lu e Reynolds (1999a, 1999b).

Para a presente dissertação, as cartas apresentadas no ponto anterior não serão desenvolvidas, uma vez que o âmbito da dissertação é o estudo e desenvolvimento de cartas de controlo multivariadas para dados autocorrelacionados.

## **2.8. CARTAS DE CONTROLO MULTIVARIADAS**

Quando se está perante um processo em que existe apenas uma única característica para controlar a qualidade do produto, utiliza-se as cartas de controlo univariadas. No entanto, nos dias de hoje, na maior parte das indústrias, os processos produtivos são complexos e os

produtos apresentam inúmeras características da qualidade para controlar. Nestes casos, não se deve aplicar o controlo estatístico do processo com cartas de controlo univariadas para cada uma das características da qualidade, ou seja, utilizar as cartas de controlo univariadas para cada característica separadamente, uma vez que a sua interpretação e diagnóstico separadamente pode resultar em conclusões falsas do processo no seu global. Uma das razões para que as conclusões não sejam as mais corretas, é que as variáveis podem ser dependentes entre elas, e ao analisar-se individualmente, pode fazer com que nenhuma delas defina apropriadamente a qualidade desse mesmo produto ou processo.

Pode também suceder-se que os processos apresentem múltiplos parâmetros e são monitorizados em simultâneo, nestes casos as cartas de controlo univariadas também não devem ser utilizadas, porque as características da qualidade podem estar correlacionadas entre si, o que pode prejudicar o desempenho das cartas na identificação de alterações no processo.

Para todos os casos supracitados, onde o objetivo é controlar mais do que uma característica da qualidade de um produto, devem ser utilizadas cartas de controlo multivariadas. As cartas de controlo multivariadas permitem agrupar em tempo real todos os dados. Estes dados tornam-se cruciais para extrair informação importante de forma a criar e desenvolver esquemas eficientes para uma ótima monitorização do desempenho do processo.

Nos processos em que se deteta a presença de autocorrelação nos dados multivariados, tem que se escolher e ajustar o melhor modelo matemático *ARIMA* para os dados em análise e construir as cartas de controlo com os resíduos ou dos erros de previsão do modelo ajustado, tal e qual como no controlo univariado. Assim, para realizar a construção das cartas multivariadas é fundamental verificar, numa primeira instância, se as variáveis são autocorrelacionadas, pois em caso afirmativo, num segundo passo deve-se determinar os resíduos correspondentes as  $p$  variáveis e só depois aplicar as cartas de controlo multivariadas.

Ao longo dos anos, muitas propostas foram feitas de cartas de controlo para monitorizar mais do que uma característica/variável da qualidade de um produto ou processo. A carta  $T^2$  de Hotelling (1947) é uma das mais populares e é a base de muitas outras propostas para o estudo multivariado.

Desta forma, e tal como já foi referido para o estudo univariado, há que distinguir duas fases bem distintas para a construção das cartas multivariadas, uma Fase 1 e uma Fase 2. A Fase 1,

corresponde ao estudo retrospectivo do processo, no qual determinam-se os limites de controlo e verifica-se se o processo está sob controlo estatístico. A Fase 2, corresponde à monitorização do processo.

Nos pontos seguintes, apresenta-se as cartas de controlo multivariadas para o controlo da média de diversas variáveis.

### 2.8.1. Carta $T^2$

A carta introduzida por Hotelling em 1947, a carta  $T^2$ , é considerada como uma extensão multivariada da carta de controlo de Shewhart univariada, baseada na monitorização das médias de amostras independentes.

Para a carta  $T^2$ , aconselha-se um mínimo de 100 medições de cada característica da qualidade na análise em causa do produto ou processo.

Para construir a carta  $T^2$ , tem que à partida se conseguir distinguir duas situações: 1) casos em que os dados são agrupados em amostras; 2) casos em que os dados são observações individuais.

A utilização apropriada da estatística  $T^2$ , ou  $\chi^2$ , pode ser dividida em quatro categorias, segundo Lowry e Montgomery (1995):

- 1) Fase 1 e  $n = 1$ , utiliza-se observações individuais;
- 2) Fase 1 e  $n > 1$ , utiliza-se subgrupos racionais;
- 3) Fase 2 e  $n = 1$ , utiliza-se observações individuais;
- 4) Fase 2 e  $n > 1$ , utiliza-se subgrupos racionais;

Na presente dissertação, vão ser apenas abordadas as situações em que os dados correspondem a observações individuais.

Desta forma, a estatística  $T^2$  para as  $p$  características da qualidade é dada pela expressão abaixo indicada (Pereira e Requeijo, 2008), onde cada uma delas contem  $m$  observações individuais, no instante  $i$ :

$$(T^2)_i = (X_i - \bar{X})S^{-1}(X_i - \bar{X}) \quad (2.26)$$

O vetor  $X_i$ , é constituído pelas observações das  $p$  características no instante  $i$ , é dado por

$$X_i = (X_{1i}, X_{2i}, \dots, X_{pi})^T \quad (2.27)$$

desta forma assume-se que  $X_1, X_2, \dots, X_p$  são independentes e identicamente distribuídas (i.i.d.) segundo  $N_p(\mu, \Sigma)$ .

O vetor média  $\bar{X}$ , é definido por

$$\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)^T \quad (2.28)$$

onde  $\bar{X}_j$ , com  $j = 1, 2, \dots, p$ , é a média das  $m$  observações, para a característica da qualidade  $j$ .

Já a matriz de covariância  $S$  é dada por

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1p} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ S_{p1} & S_{p2} & S_{p3} & \cdots & S_{pp} \end{bmatrix} \quad (2.29)$$

neste caso onde os elementos  $S_{jj}$  da matriz  $S$ , correspondem às variâncias amostrais de cada uma das  $p$  características e os elementos  $S_{ji}$ , correspondem às covariância entre duas características distintas.

$$S_{jj} = S_j^2 = \frac{1}{m-1} \sum_{k=1}^m (X_{ji} - \bar{X}_j)^2 \quad (2.30)$$

$$S_{ji} = \frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)(X_{jh} - \bar{X}_h); j \neq h \quad (2.31)$$

Quanto aos limites de controlo da carta  $T^2$ , para a Fase 1, segundo Tracy et al. (1992), estes devem ser baseados na função da distribuição Beta, sendo definidos pela seguinte expressão:

$$LSC_{T^2} = \frac{(m-1)^2}{m} \beta_{\alpha; \left(\frac{p}{2}, \frac{m-p-1}{2}\right)} \quad (2.32)$$

$$LIC_{T^2} = 0 \quad (2.33)$$

neste caso  $\beta_{\alpha; \left(\frac{p}{2}, \frac{m-p-1}{2}\right)}$ , representa o percentil à direita, para uma probabilidade  $\alpha$ , da distribuição Beta com parâmetros  $\frac{p}{2}$  e  $\frac{m-p-1}{2}$ . Podem, no entanto, ocorrer casos onde não se tem acesso a *software* que permite calcular o valor da distribuição Beta, desta forma, para estas situações, Tracy et al. (1992) sugere a sua determinação em função da distribuição de Fisher:

$$\beta_{\alpha; \left(\frac{p}{2}, \frac{m-p-1}{2}\right)} = \frac{\frac{p}{m-p-1} F_{\alpha; p, (m-p-1)}}{1 + \frac{p}{m-p-1} F_{\alpha; p, (m-p-1)}} \quad (2.34)$$

onde

$F_{\alpha; p, (m-p-1)}$  corresponde ao percentil à direita para uma probabilidade  $\alpha$ .

Para a Fase 2, no que concerne à carta  $T^2$ , a estatística é dada pela expressão 2.26 e os seus limites de controlo por:

$$LSC_{T^2} = \frac{p(m+1)(m-1)}{m(m-p)} F_{\alpha; p, m-p} \quad (2.35)$$

$$LIC_{T^2} = 0 \quad (2.36)$$

Nos casos em que os parâmetros  $\mu$  e  $\Sigma$ , são estimados a partir de um grande número de amostras preliminares ( $m > 100$ ) o limite superior desta carta pode ser dado pela seguinte fórmula, sugerida por Montgomery (2005):

$$LSC_{\chi^2} = \chi_{\alpha; p}^2 \quad (2.37)$$

$$LSC_{\chi^2} = 0$$

Assim tanto nas cartas univariadas como nas presentes cartas multivariadas, a deteção de uma causa especial de variação é feita quando um valor de  $(T^2)_i$  ou  $(\chi^2)_i$  excede os limites de controlo da respetiva carta.

Por outro lado, tal como no caso univariado, as cartas de controlo multivariadas do tipo Shewhart ( $T^2$  ou  $\chi^2$ ) apenas utilizam a informação das observações recentes, o que as tornam pouco sensíveis para a deteção das pequenas e moderadas alterações que vão afetando os parâmetros das  $p$  características de um processo. Para ir de encontro à resolução destes problemas, foi necessário desenvolver as cartas *CUSUM* e *EWMA* para o caso multivariado.

Tal como as cartas *CUSUM* e *EMWA* univariadas, estas também dependem do  $ARL_{EC}$  e do desvio que se pretende detetar.

Uma vez que se está num caso multivariado,  $p$  características, o desvio é dado pelo parâmetro de não centralidade  $\delta^2(\mu)$  ou  $\delta(\mu)$  designado como “distância estatística”, e definido por (Pereira e Requeijo, 2008)

$$\delta^2(\mu)\Sigma^{-1}(\mu - \mu_0) \quad (2.38)$$

onde:

$\mu_0$  – Vetor média inicial do processo;

$\delta^2(\mu)$  - Parâmetro de não centralidade.

### 2.8.2. Carta CUSUM

A carta *CUSUM* multivariada (*MCUSUM*), é representada pela estatística  $Y^*$ . Esta estatística é definida, no instante  $t$ , por (Pereira e Requeijo, 2008):

$$Y_t^* = (C_t \Sigma^{-1} C_t)^{\frac{1}{2}} \quad (2.39)$$

Onde

$$C_t = (C_{t-1} + X_t - \mu_0) \left(1 - \frac{k}{C_t^*}\right) \text{ se } C_t^* \leq k \quad (2.40)$$

$$C_t = 0 \text{ se } C_t^* > k \quad (2.41)$$

Em que

$$C_t^* = [(C_{t-1} + X_t - \mu_0)^T \Sigma^{-1} (C_{t-1} + X_t - \mu_0)]^{\frac{1}{2}} \quad (2.42)$$

$$C_0 = 0 \text{ e } k > 0 \quad (2.43)$$

Na carta *MCUSUM*, deteta-se a ocorrência de uma causa especial de variação sempre que um valor de  $Y_t^* > h$ , no qual  $h$  corresponde ao limite de controlo. Como a determinação do *ARL*, para esta carta, depende dos parâmetros de não centralidade, Croisier (1988) recomenda  $k = \frac{\delta(\mu)}{2}$ , onde  $\delta^2(\mu)$  é dado pela expressão (2.38).

Crosier (1988) conclui que a carta *MCUSUM* é mais sensível que a carta  $T^2$  e a carta  $\chi^2$ , no que diz respeito à deteção de alterações pequenas e moderadas do vetor média.

### 2.8.3. Carta EWMA

A carta *EWMA* multivariada (*MEWMA*), é representada pela estatística  $T^2$  e definida, no instante  $t$ , por (Pereira e Requeijo, 2008)

$$T_t^2 = Z_t^T \Sigma_Z^{-1} Z_t \quad (2.44)$$

onde

$$Z_t = RX_t + (I - R)Z_{t-1}, \quad Z_0 = 0 \quad (2.45)$$

$$\Sigma_Z = \frac{\lambda}{2-\lambda} [1 - (1 - \lambda)^{2t}] \Sigma \quad (2.46)$$

Normalmente,  $\lambda_1 = \lambda_2 = \dots = \lambda_p$ .

Quando isto acontece,  $Z_t$  passa a ser definido por

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1} \quad (2.47)$$

onde

$\Sigma_Z$  – Matriz covariância de  $Z$ ;

$I$  – Matriz identidade;

$R = \text{dig}(\lambda_1, \lambda_2, \dots, \lambda_p)$  com  $0 \leq \lambda_j \leq 1$ ;

$\lambda_j$  - Constante de amortecimento para a variável  $j$  ( $j = 1, 2, \dots, p$ ).

Para as cartas *MEWMA*, sempre que um valor de  $T_t^2 > H$ , deteta-se uma situação fora de controlo. O valor de  $H$  é obtido através do recurso à simulação, de forma a encontrar um  $ARL_{EC}$  específico.

Segundo Bersimis et al. (2006), o desempenho do *ARL*, depende apenas dos parâmetros de não centralidade, caso raras exceções, como nos casos em que se utiliza constantes com pesos diferentes, o *ARL* passa a depender da direção das alterações.

A aproximação da expressão (2.40), dada por  $\Sigma_Z = \left(\frac{\lambda}{2-\lambda}\right)\Sigma$ , é a sugestão de muitos autores. Porém, a utilização da expressão 2.40 conduz a uma resposta inicial rápida para a carta *MEWMA* (Bersimis et al., 2006).

#### 2.8.4. Interpretação da Carta $T^2$

Uma das tarefas mais difíceis na engenharia de processos e melhoria contínua, é a identificação de uma causa especial, isto porque a sua causa pode ser de inúmeras naturezas, entre elas, relação com uma única variável ou com um subconjunto de variáveis. Este problema de deteção da(s) variável(eis) que está(ão) associada(s) à existência de uma causa especial de variação foi alvo de grande investigação por vários autores. Entre eles, Alt, em 1985, tendo proposto a utilização de cartas univariadas das  $p$  variáveis com limites baseados nas desigualdades de Bonferroni. Jackson também em 1985, utilizou a decomposição de  $T^2$  ou  $(\chi^2)$  em componentes principais. Hauter e Tsui, em 1994, sugeriram o método do ponto crítico,  $C_{R,\alpha}$ . Mason, Tracy e Young em 1995 e 1997, utilizaram a decomposição de  $T^2$  ou  $(\chi^2)$  em componentes ortogonais. Já Runger, Alt e Montgomery (1996) e Montgomery (2005) usaram a estatística  $d$ . Nedumaran e Pignatiello, em 1998, sugeriram a utilização de cartas univariadas com limites de diagnóstico.

Para a presente dissertação, apenas serão abordados dois dos métodos anteriormente expostos, o método da estatística  $d$  e o da decomposição em componentes ortogonais, uma vez que foram os métodos mais desenvolvidos por diversos autores e que de certa forma, são mais eficazes e tradicionais.

#### **Método da Estatística $d$**

O Método da Estatística  $d$  tem como base a determinação da variável/eis responsável/eis por uma situação de fora de controlo.

Para tal, utiliza a estatística  $d$ , que é designada, para cada variável  $i$ , por (Pereira e Requeijo, 2008)

$$d_i = T^2 - T_{(i)}^2, \quad i = 1, 2, \dots, p \quad (2.48)$$

onde

$T^2$ - Valor da estatística, determinado com base nas  $p$  variáveis;

$T_{(i)}^2$ - valor da estatística calculada a partir de  $(p - 1)$  variáveis, tendo em conta que a variável em questão ( $i$ ) não entra para o cálculo.

Uma variável contribui significativamente para a situação de fora de controlo se

$$d_i > \chi_{\alpha;1}^2 \quad (2.49)$$

### Método de Decomposição em Componentes Ortogonais

O Método de Decomposição em Componentes Ortogonais, é sugerido por Mason, Tracy e Young em 1995, que tem por base a decomposição da estatística  $T^2$  em duas partes independentes, expressão (2.50). Cada uma das partes reflete a contribuição de uma variável individual. Esta decomposição ajuda os investigadores a detetar qual a/as variável/eis, com contribuição significativa, é/são a/as causadora/as do desvio existente.

$$T^2 = T_{p-1}^2 + T_{p,1,2,\dots,p-1}^2 \quad (2.50)$$

O primeiro termo da expressão (2.50) corresponde à estatística  $T^2$  das  $(p - 1)$  variáveis, no qual também pode ser separado em duas partes

$$T_{p-1}^2 = T_{p-2}^2 + T_{p-1,1,2,\dots,p-2}^2 \quad (2.51)$$

O raciocínio acima descrito pode também ser feito para as estatísticas  $T_{p-k}^2$ , obtendo-se assim a equação equivalente à (2.50)

$$T^2 = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 + T_{4,1,2,3}^2 + T_{p,1,2,\dots,p-1}^2 = T_1^2 + \sum_{j=1}^{p-1} T_{j+1,1,2,\dots,j}^2 \quad (2.52)$$

onde

$T_1^2$ - Estatística da primeira variável;

$\sum_{j=1}^{p-1} T_{j+1,1,2,\dots,j}^2$  - Estatística da variável  $(j + 1)$  ajustada pelas estimativas da média e desvio padrão da distribuição condicionada da variável  $(j + 1)$ , dada a primeira variável  $j$ .

Para decompor a estatística  $T^2$  existem  $p!$  combinações possíveis. Um exemplo, é o caso em que se dispõe de três variáveis  $p = 3$ , isto significa que existem seis possibilidades para decompor a estatística  $T^2$  (Tabela 2.3), no qual todas as combinações deverão ser analisadas (Pereira e Requeijo, 2008).

**Tabela 2.3 - Decomposição de  $T^2$  em componentes ortogonais para  $p = 3$  (adaptado de Pereira e Requeijo, 2008)**

Combinação	Ordenação das Variáveis			Termo Incondicional	Termos Condicionais	Decomposição em Componentes Ortogonais
	1	2	3			
1	$X_1$	$X_2$	$X_3$	$T_1^2$	$T_{2,1}^2$ e $T_{3,1,2}^2$	$T_1^2 + T_{2,1}^2 + T_{3,1,2}^2$
2	$X_1$	$X_3$	$X_2$	$T_1^2$	$T_{3,1}^2$ e $T_{2,1,3}^2$	$T_1^2 + T_{3,1}^2 + T_{2,1,3}^2$
3	$X_2$	$X_1$	$X_3$	$T_2^2$	$T_{1,2}^2$ e $T_{3,2,1}^2$	$T_2^2 + T_{1,2}^2 + T_{3,2,1}^2$
4	$X_2$	$X_3$	$X_1$	$T_2^2$	$T_{1,2}^2$ e $T_{3,2,1}^2$	$T_2^2 + T_{1,2}^2 + T_{3,2,1}^2$
5	$X_3$	$X_1$	$X_2$	$T_3^2$	$T_{1,3}^2$ e $T_{2,3,1}^2$	$T_3^2 + T_{1,3}^2 + T_{2,3,1}^2$
6	$X_3$	$X_2$	$X_1$	$T_3^2$	$T_{2,3}^2$ e $T_{1,3,2}^2$	$T_3^2 + T_{2,3}^2 + T_{1,3,2}^2$

Desta forma, os termos incondicionais ( $T_j^2$ ) e condicionais ( $T_{j+1,1,2,\dots,j}^2$ ) apresentados na tabela anterior podem ser dados pelas seguintes expressões

$$T_j^2 = \frac{(x_{ij} - \bar{x}_j)^2}{s_j^2} \quad (2.53)$$

$$T_{j+1,1,2,\dots,j}^2 = \frac{(x_{ip} - \bar{x}_{p,1,\dots,p-1})^2}{s_{p,1,\dots,p-1}^2}, \text{ no qual } p = j + 1 \quad (2.54)$$

onde

$$\bar{X}_{p.1,\dots,p-1} = \bar{X}_p + b_p^T (X_i^{(p-1)} - \bar{X}_i^{(p-1)}) \quad (2.55)$$

$$b_p = S_{XX}^{-1} s_{xX} \quad (2.56)$$

$$S_{p.1,\dots,p-1}^2 = S_x^2 - s_{xX}^T S_{XX}^{-1} s_{xX} \quad (2.57)$$

Para se identificar e analisar uma situação de fora de controlo, utilizando o Método de Decomposição em Componentes Ortogonais, comparam-se os valores dos termos incondicionais e dos condicionais com os seus valores críticos. Desta forma, considera-se que um termo condicional contribui significativamente para a situação de fora de controlo se

$$T_{j+1.1,1,2,\dots,j}^2 > \frac{m-1}{m} F_{\alpha;1,m-1} \quad (2.58)$$

Já nos casos dos termos incondicionais, considera-se que este contribui significativamente para uma situação de fora de controlo se

$$T_j^2 > \frac{m-1}{m} F_{\alpha;1,m-1} \quad (2.59)$$

Um dos problemas deste método é apresentar um elevado número de combinações, o que pode fazer com que haja vários termos condicionais significativos, o que irá dificultar a sua interpretação. Assim, uma forma rápida de reduzir o número de termos a analisar/interpretar, é a determinação da diferença entre o valor da estatística  $T^2$ , da observação em questão, e os termos incondicionais:

$$T^2 - T_1^2, T^2 - T_2^2, \dots, T^2 - T_p^2 \quad (2.60)$$

No próximo passo, irá comparar-se cada diferença  $(T^2 - T_j^2)$  com o valor crítico  $\left[ \frac{m-1}{m} F_{\alpha;1,m-1} \right]$ .

Desta forma, os termos condicionais  $(T_{j+1,1,1,2,\dots,j}^2)$  não contribuem significativamente a uma situação de fora de controlo, logo não precisam de ser analisados, se

$$T^2 - T_j^2 < \frac{m-1}{m} F_{\alpha;1,m-1} \quad (2.61)$$

Por sua vez, os termos condicionais  $(T_{j+1,1,1,2,\dots,j}^2)$  contribuem significativamente para a situação de fora de controlo, logo devem ser analisados, se

$$T^2 - T_j^2 > \frac{m-1}{m} F_{\alpha;1,m-1} \quad (2.62)$$

Já relativamente aos termos incondicionais  $(T_j^2)$  estes devem sempre ser comparados com o valor crítico  $\left[\frac{m-1}{m} F_{\alpha;1,m-1}\right]$ .

Por fim, em conclusão, se um termo incondicional apresentar um valor superior ao valor crítico significa que a situação de fora de controlo é devido a uma alteração no parâmetro da variável  $j$ . Já se um termo condicional apresentar um valor superior ao valor crítico, então a situação de fora de controlo é devido a uma alteração na correlação entre as variáveis representadas no termo em questão.

## 2.9. CARTAS DE CONTROLO MULTIVARIADAS BASEADAS EM MÉTODOS DE PROJEÇÃO

Quando o objetivo é controlar um número elevado de variáveis, as cartas multivariadas tradicionais ( $T^2$ , *MCUSUM*, *MEWMA*) tendem a perder a eficácia no que respeita à deteção de alterações no processo, uma vez que estas só são exequíveis quando o número de variáveis a monitorizar não é muito elevado (Matos, 2006). No entanto, existem métodos mais eficazes na deteção de alterações no processo quando se está perante um processo com um número elevado de variáveis, está-se a falar de métodos de projeção, nomeadamente, a análise de componentes principais (*PCA* – *Principal Component Analysis*) e a regressão pelos Quadrados Mínimos Parciais (*PLS* – *Partial Least Square*).

Na dissertação em causa, e uma vez que um dos objetivos é o estudo comparativo de duas cartas, em que uma delas tem como base a análise de componentes principais (*PCA*), apenas se vai desenvolver este método adiante.

A análise de componentes principais (adiante, designada por *PCA*) é uma ferramenta extremamente poderosa na monitorização de processos complexos pois permite avaliar o desempenho de toda a unidade do processo (tanto pode controlar as características da qualidade de um produto, como também controlar vários sensores de temperatura, entre outras características).

Os desenvolvimentos da *PCA*, como um instrumento de referência para a monitorização de processos foi devido a uma série de investigadores, que analisaram o método e chegaram a resultados fundamentais, entre eles destacam-se MacGregor e Kourti (1995), Kourti e MacGregor (1995), com a contribuição importante de Dunia e Quin, em 1998, Negiz e Cinar, em 1997, Raich e Cinar, em 1997, entre outros.

A *PCA* é uma técnica estatística multivariada utilizada para reduzir a dimensão do espaço de monitorização, projetando a informação das variáveis originais para subespaços dimensionais, pequenos, definidos por poucas variáveis independentes (os componentes principais), segundo Ferrer (2007).

Na base de aplicação da *PCA*, está uma matriz de dados  $X$  composta por  $p$  variáveis aleatórias. Para a construção da matriz, é necessário garantir a normalização das variáveis, uma vez que estas geralmente correspondem a medidas de diferentes unidades, isto tudo para se obter resultados satisfatórios da aplicabilidade da *PCA*. Assim, a normalização das variáveis não é mais do que em cada coluna subtrair-se a média e dividir-se pelo desvio padrão da respetiva variável, o que faz com que as variáveis passem a ter média zero e variância unitária (Ku et al., 1995).

Este passo é fundamental, porque normalmente melhora a interpretação do modelo uma vez que todas as variáveis possuem o mesmo valor de média.

Por sua vez, esta matriz de dados é transformada em  $A$  componentes principais.

Cada componente principal, denominado *score*, é formado por uma combinação linear das variáveis aleatórias originais, sendo esta combinação obtida a partir da matriz de covariância ( $\Sigma$ ). Cada componente principal descreve a porção da variância total apresentada pelo

conjunto das  $p$  variáveis aleatórias. Segundo MacGregor e Kourti (1995) e Kourti (2005), o primeiro componente principal, definido como uma combinação linear ( $t_1 = p_1^T X$ ), é determinado de forma a reter a maior variância possível, sujeita a  $p_1 = 1$ . Já o segundo componente principal ( $t_2 = p_2^T X$ ), terá a segunda variância mais alta, sujeita a  $p_2 = 2$ , mas submetido à condição de que não deve ser correlacionado com o primeiro componente. Os restantes componentes principais são definidos do mesmo modo, até esgotar toda a variância disponível no conjunto de variáveis. Para tal os vetores  $p_i$  correspondem aos vetores próprios da matriz covariância ( $\Sigma$ ) de  $X$ , dada pela expressão (2.63) e os valores próprios  $\lambda_i$  correspondentes transmitem a variância de cada componente principal, obtida pela expressão (2.64).

$$\text{cov}(X) = \frac{X^T X}{m-1} \quad (2.63)$$

$$\text{Var}(t_i) = \lambda_i \quad (2.64)$$

Normalmente, não é necessário calcular os  $p$  vetores próprios, uma vez que a maior parte da variabilidade nos dados consegue ser capturada nos primeiros componentes principais, por exemplo, 2 ou 3 dos componentes principais são suficientes para explicar grande parte da variância aplicada nos produtos. Assim, uma vez retidos os primeiros  $A$  componentes principais, a matriz de dados  $X$  é decomposta numa soma de matrizes, definida por (Kourti, 2005)

$$X = TP^T + E = \sum_{i=1}^k t_i p_i^T + E \quad (2.65)$$

onde

$T$ - Matriz dos componentes principais;

$P$ - Matriz dos vetores próprios;

$E$ - Matriz residual (representa a informação que não é explicada pelo modelo  $PCA$ );

Para se determinar os componentes principais e os respectivos vetores próprios pode-se recorrer a vários métodos. Segundo os principais autores/investigadores destes métodos, nomeadamente Kourti e MacGregor (1995) e Kourti (2005), o método que reúne maior consenso é o algoritmo *NIPALS* (*Nonlinear Iterative Pastial Least Square*) é o método ideal pois este permite determinar os componentes principais de uma forma sequencial, quando o número de variáveis é elevado.

Já no que diz respeito ao número de componentes principais necessários para representar de forma eficiente o conjunto de dados, este pode ser obtido através de vários métodos, nomeadamente, através do método de Kaiser, pelo método da percentagem de variância explicada, pela validação cruzada ou pelo gráfico de componentes versus os respetivos valores próprios.

Para o efeito, e seguindo mais uma vez o que os autores mais especializados nestes estudos referem, como é o caso de MacGregor e Kourti (1995), o método considerado mais adequado corresponde ao método da validação cruzada.

Assim no método da validação cruzada, através dos componentes principais, *scores*, e dos resíduos associados a cada observação, são derivadas duas estatísticas complementares (ortogonais ou independentes), a estatística  $T_A^2$  de Hotelling e a estatística Q ou SPE (erro de previsão quadrático).

A estatística  $T_A^2$ , que representa uma medida total da variação do processo, ou seja, utiliza os componentes principais retidos para controlar a variabilidade das principais fontes do processo.

Caso um dos valores desta estatística exceda os limites de controlo, significa que a observação correspondente apresenta valores extremos anormais em algumas, ou em todas, as suas  $p$  variáveis originais. Matematicamente, a estatística  $T_A^2$  pode ser construída com base nos primeiros  $A$  componentes principais (Kourti, 2005):

$$T_A^2 = \sum_{i=1}^A \frac{t_i^2}{\lambda_i} \quad (2.66)$$

onde

$t_i$  – Valor do componente principal  $i$ ;

$\lambda_i$  - Valor do próprio do componente principal  $i$ ;

O limite superior de controlo da estatística  $T_A^2$ , é baseado nos primeiros  $A$  componentes principais, e é obtido recorrendo à distribuição de Fisher, e representado pela seguinte expressão (Kourti, 2005):

$$LSC(T_A^2) = \frac{A(m^2-1)}{m(m-A)} F_{\alpha;(A,m-A)} \quad (2.67)$$

onde

$F_{\alpha;(A,m-A)}$  - Representa o ponto crítico da distribuição de Fisher, considerando um nível de significância de  $\alpha$ .

Para esta carta, também se pode aproximar o limite superior de controlo à distribuição Qui-Quadrado (expressão 2.37).

A segunda carta de controlo, designada por  $Q$ , ou estatística  $Q$ , mede a quantidade de variação que não é capturada pelo modelo *PCA* (Kano et al., 2001). Uma vez que, apesar da maior parte da variabilidade dos dados ser explicada pelos primeiros componentes principais, os restantes componentes podem fornecer informações úteis, porque estes também capturam variâncias apesar de pequenas.

Matematicamente, a estatística  $Q$  pode ser dada pela seguinte expressão (Kourti, 2005):

$$Q = \sum_{j=1}^k (X_{j,novo} - \hat{X}_{j,novo})^2 \quad (2.68)$$

onde

$X_{j,novo}$  - Valor da observação da variável  $j$ ;

$\hat{X}_{j,novo}$  - Estimativa da variável  $j$  resultante do modelo *PCA*;

Assumindo-se que os resíduos seguem uma distribuição normal multivariada, vários procedimentos podem ser utilizados para determinar o limite superior de controle da estatística  $Q$ .

Como, por exemplo, Jackson e Mudholkar, em 1979, mostraram que o limite de controle, pode ser dado por

$$LSC(Q) = \theta_1 \left[ \frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (2.69)$$

onde

$$\theta_i = \sum_{j=A+1}^m (\lambda_j)^i = TR(E^k), \quad \text{com } i = 1, 2, 3 \quad (2.70)$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_1^2} \quad (2.71)$$

Nas equações anteriores considera-se:

$z_\alpha$  – Desvio normal correspondente ao percentil à direita para  $\alpha$  de probabilidade;

$\alpha$  – Nível de significância;

$\lambda_j$  – Valores próprios da matriz de covariância dos resíduos;

$\theta_i$  – Valor obtido com base nos valores próprios da matriz de covariância dos resíduos.

Por sua vez, Nomikos e MacGregor em 1995, utilizaram uma aproximação baseada na distribuição Qui-Quadrada ponderada:

$$Q_i \sim g\chi_h^2 \quad (2.72)$$

onde

$h$  – Número de graus de liberdade;

$g$  – Fator multiplicativo

Nomikos e MacGregor sugeriram também um modo simples e rápido para estimar os parâmetros  $g$  e  $h$ , que é baseado na correspondência de momentos entre a distribuição  $g\chi_h^2$  e a distribuição da amostra de  $Q$ . A média ( $\mu = gh$ ) e a variância ( $\sigma^2 = 2g^2h$ ) da distribuição Qui-Quadrado ponderada são comparadas com a média ( $b$ ) e variância ( $v$ ) da amostra de  $Q$ .

Por isso, os parâmetros  $g$  e  $h$  são estimados através das seguintes equações:

$$\hat{g} = \frac{v}{2b} \quad (2.73)$$

$$\hat{h} = \frac{2b^2}{v} \quad (2.74)$$

Desta forma, o limite superior de controlo da estatística  $Q$ , para um nível de significância  $\alpha$  é dado por

$$LSC(Q) = \frac{v}{2b} \chi_{\left(\frac{2b^2}{v}\right), \alpha}^2 \quad (2.75)$$

Para verificar se o processo está sob controlo estatístico, a carta  $Q$  deve ser a primeira a ser avaliada. Caso todos os pontos estejam dentro dos limites de controlo das duas cartas [ $T_{Ai}^2 < LSC(T_A^2)$  e  $Q_i \leq LSC(Q)$ ], o processo é considerado sob controlo.

Para que a construção destas duas cartas seja útil, carta  $T^2$  e  $Q$ , segundo Ostyn et al. (2007), na prática, devem considerar-se duas fases importantes.

Numa primeira fase, fase de construção, o modelo  $PCA$  deve ser construído utilizando um conjunto de dados de calibração e os valores limites de cada uma das estatísticas  $T_A^2$  e  $Q$  são estabelecidos utilizando as equações (2.67) e (2.69).

Numa segunda fase, fase de validação, os novos dados são projetados no modelo  $PCA$  calibrado e as estatísticas  $T_A^2$  e  $Q$  são calculadas e comparadas com os valores-limites obtidos

na primeira fase. Se as estatísticas  $T_A^2$  ou  $Q$  traçadas apresentarem valores acima destes limites superiores de controlo, significa que existe um desvio da observação, dando origem a um alarme (Montgomery, 2005).

### 2.9.1. Interpretação da Carta PCA

Sempre que é seja detetado um ponto fora dos limites de controlo de uma das cartas, é necessário uma análise individual e isolada de cada variável original responsável pelo sinal fora de controlo. Existem vários métodos para identificar quais as variáveis que mais contribuem para que ocorra um alarme, sinal fora de controlo, um desses métodos e talvez o de mais fácil visualização e interpretação de qual é a variável mais responsável pelo ponto fora de controlo, é o gráfico de contribuição. Este método é bastante poderoso uma vez que a informação que utiliza para base de análise contem toda a informação de todas as variáveis medidas em simultâneo.

Um exemplo deste método, é gráfico de contribuição presente na Figura 2.3, que indica quais as variáveis que contribuem numericamente para uma condição de fora de controlo. Estes são gráficos de barras no qual são desenhados os valores observados das variáveis, no período em que foi visualizada/assinalada uma causa especial.

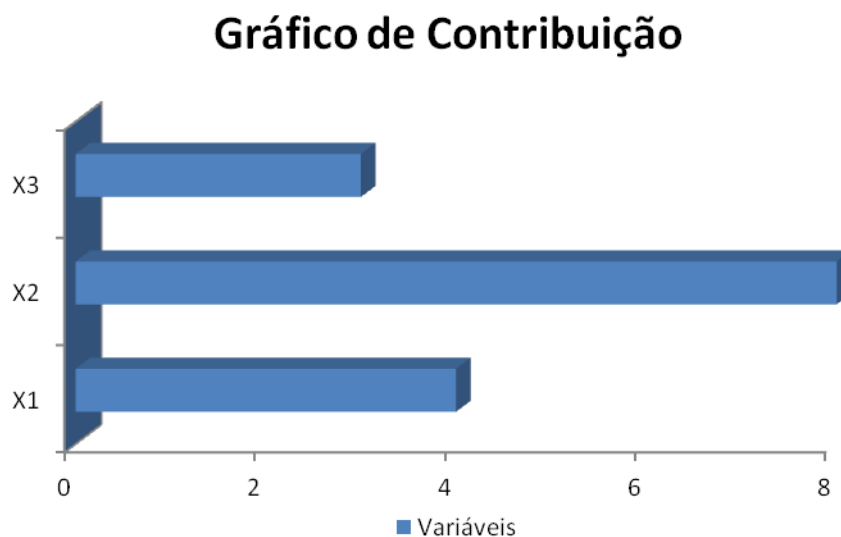


Figura 2.3 - Exemplo de um Gráfico de Contribuição

Os gráficos de contribuição podem ser calculados para os pontos “aberrantes”, ou seja, fora do controlo para as duas cartas, estatística  $T^2$  e  $Q$ .

Se a situação fora de controlo for detetada na carta  $Q$ , a contribuição de cada variável  $j$ , do conjunto de dados originais, é dada por (Kourti, 2005):

$$c_j^{(Q)} = (X_{j,novo} - \hat{X}_{j,novo})^2 \quad (2.76)$$

No caso, em que as variáveis apresentem valores elevados para as contribuições estas devem ser investigadas de imediato.

Se a situação fora de controlo for detetada na carta  $T_A^2$ , é construído um gráfico de barras dos componentes principais normalizado  $\left[\frac{(t_i)^2}{\lambda_i}\right]$ , da observação em questão, e o componente principal que apresentar maior valor normalizado é selecionado e investigado através do cálculo de contribuição da variável (Kourti, 2005).

O gráfico de contribuição da variável indica como é que cada variável envolvida no cálculo deste componente principal, contribui para o sinal fora de controlo. Assim, a contribuição de cada variável do componente principal é dada por (Kourti, 2005):

$$C_j^{(t_i)} = p_{ij}(X_j - X_j^*) \quad (2.77)$$

onde

$p_{ij}$  – Elemento da  $i$ -ésima linha e  $j$ -ésima coluna da matriz  $P$ .

Depois de determinada a contribuição de cada variável, é construído o gráfico de contribuição.

As variáveis, presentes neste gráfico, que tiverem valores elevados nas contribuições mas com o mesmo sinal que o componente principal, devem ser investigadas. Por outro lado, as contribuições que tiverem sinal oposto ao do componente principal, apenas farão com que este seja menor.

Para os casos em que existem vários componentes principais com valores elevados nas contribuições, deve-se determinar a “contribuição média total” por variável para todos os componentes principais selecionados, seguindo-se sempre as seguintes etapas (Kourti, 2005):

**1) Repetir para todos os  $k$  componentes principais elevados ( $k \leq A$ );**

- I. Determinar a contribuição da variável  $X_j$  para o componente principal selecionado:

$$c_j^{(t_i)} = \frac{t_i}{\lambda_i} p_{ij} (X_j - X_j^*) \quad (2.78)$$

- II. Colocar a contribuição igual a zero caso seja negativa.

**2) Calcular a contribuição total da variável  $x_k$ ;**

$$C_j^{(T_A^2)} = \sum_{i=1}^k (c_j^{(t_i)}) \quad (2.79)$$

**3) Investigar as variáveis com elevadas contribuições.**

O gráfico de contribuição é assim, um método ilustrativo principalmente para ajudar na interpretação do problema e auxiliar na tomada de decisões corretivas.

No entanto, é importante ter em mente, que o gráfico de contribuição pode não revelar a causa do evento especial, nestes casos é necessário a intervenção de alguém que contenha o conhecimento do processo, por exemplo um operador do processo em causa, de forma a poderem descobrir o que provocou as alterações nas variáveis identificadas.

## **2.9.2. Cartas de Controle Multivariadas Baseadas na Análise de Componentes Principais Dinâmicas**

Na atualidade, a maioria das indústrias processos, nomeadamente, industriais de produção fabril, estão bastante desenvolvidas no que concerne a novas tecnologias, sistemas de informação e otimização de processos. Está-se perante a era da modernização, a procura constante pela melhoria contínua, reengenharia de processos e eliminação de desperdícios. Toda esta modernização e desenvolvimento leva a que os processos produtivos sejam mais automáticos, de curta duração e de rápida execução. Todos os processos produtivos têm tarefas mais complexas que dependem de acontecimentos passados e aqui é que se encontra o grande desafio da atualidade nas indústrias modernas, ou seja, monitorizar processos complexos com multivariáveis com dados autocorrelacionados. Para processos com autocorrelação significativa, a aplicação do método *PCA* convencional pode não ser eficiente pois como a base estatística desse método é perdida, devido aos dados violarem a suposição da independência do tempo, podem-se obter resultados insatisfatórios que, por sua vez, poderão gerar excessivos falsos alarmes, especialmente para perturbações de tamanhos pequenos, segundo Tsung (2000). Por outro lado, quando ocorrem alterações no processo, a relação entre  $\varepsilon_t$  e  $X_t$  é sempre alterada. Assim, a aplicação direta do método *PCA* convencional nos dados, não irá revelar a relação exata entre as variáveis, mas sim uma aproximação estática linear. Por isso, não é muito eficaz caracterizar a relação entre as variáveis segundo um modelo *PCA* estático.

Para ultrapassar o problema da caracterização da relação entre as variáveis que não é muito eficaz através do modelo *PCA* estático, propõe-se a utilização do método *PCA* dinâmico, ao qual se vai denominar por *DPCA*, proposto inicialmente por Ku et al. (1995). Os investigadores em causa, sugeriram desta forma a utilização do método *PCA* na monitorização de processos para explicar a autocorrelação. Ao contrário do sistema estático, no sistema dinâmico os valores atuais das variáveis vão depender dos valores passados por isso, é necessário identificar, no mínimo, a relação linear entre  $X(t)$  e  $X(t - 1)$ . Na base, este método é igual ao original *PCA*, exceto a matriz dos dados ser composta por vetores replicados com atrasos de uma unidade de tempo por réplica.

Desta forma, dependendo do número de atrasos ( $l$ ), a matriz de dados é expressa da seguinte forma:

$$X(l) = \begin{bmatrix} X_t^T & X_{t-1}^T & \cdots & X_{t-l}^T \\ X_{t-1}^T & X_{t-2}^T & \cdots & X_{t-l-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ X_{t+l-m}^T & X_{t+l-m-1}^T & \cdots & X_{t-m}^T \end{bmatrix} \quad (2.80)$$

onde

$X_t^T = [X_{1t}, X_{2t}, \dots, X_{pt}]^T$  - é o vetor de dimensão  $p$ , no instante  $t$ .

No modelo *DPCA*, os resíduos que dele advém, devem ser caracterizados como perfeitamente independentes, formando desta forma, uma base estatística estável para a análise e estudo do processo dinâmico em causa, uma vez que foi feita uma escolha correta do valor inteiro a atribuir ao número de atrasos ( $l$ ), segundo Ku et al. (1995).

Depois de se ter construído a matriz de dados, utilizam-se as mesmas expressões aplicadas no ponto 2.9 e ponto 2.9.1 para determinar as estatísticas  $T_A^2$  e  $Q$ , assim como os respetivos limites de controlo bem como os gráficos de contribuições de forma a indicar as variáveis responsáveis pelas causas especiais.

Na aplicação da *PCA* na matriz de dados dada pela expressão (2.80), neste caso identifica-se um modelo multivariado autorregressivo (*AR*) que é extraído diretamente dos dados. Segundo Ku et al. (1995) no caso em que a matriz  $X(l)$  inclua as variáveis de entrada, o modelo obtido será do tipo *ARX*, modelo autorregressivo com variável exógena.

Já Russel et al. (2000), retira outras conclusões interessantes diferentes de Ku et al. (1995), indicando que se na matriz  $X(l)$  são incluídos um número de atrasos ( $l$ ) suficientes, então a carta  $Q$  é estatisticamente independente de um instante de tempo para outro e o limite de controlo desta carta, dado pela expressão (2.69) é teoricamente justificado.

## 2.10. PROPOSTA DE NOVA CARTA DE CONTROLO MULTIVARIADA, *DMPCA*

O interesse em métodos de controlo de qualidade multivariados tem aumentado consideravelmente nos últimos anos, impulsionado principalmente pelas necessidades das indústrias químicas, onde o aparecimento de instrumentos modernos e processos automáticos permitiram maior produção de grandes conjuntos de variáveis altamente correlacionadas, combinado com a crescente preocupação sobre a segurança, a manutenção, o rendimento e qualidade no processo.

Um dos métodos mais populares de controlo estatístico do processo é a análise de componentes principais (*PCA*), uma vez que pode lidar com um elevado número de variáveis, com dados autocorrelacionados, através da projeção de dados para subespaço dimensional que contém a maior parte da variância dos dados originais (J.-M . Lee et al., 2004). No entanto, o método *PCA* assume uma dinâmica insignificante do processo sob condições anormais de operação e só é adequado para a utilização em condições estáticas ou fracamente dinâmico, tal como já foi referido anteriormente. Isto faz com que a utilização do método *PCA* seja muito limitado, no que respeita aos processos químicos. Como resposta a esta limitação, o *PCA* foi estendido para incluir as estruturas de séries temporais de variáveis. Entre essas extensões, a análise de componentes principais dinâmica (*DPCA*), proposto por Ku et al. (1995), ganhou a atenção e muita popularidade, em parte pela sua simplicidade.

A análise de componentes principais dinâmica (*DPCA*) sugere a inserção de um modelo de estrutura linear de séries temporais para as variáveis do processo, aumentando a matriz com as variáveis de tempo desfasados. Muitos outros desenvolvimentos foram surgindo/criados a partir de/com este novo conceito, tal como muitos autores citam em alguns *papers*, entre eles Luo et al. (1999), Yoo et al., C.Lee et al. (2004) e Liu & Makis (2008) para a deteção de falhas de sensores, Treasure et al. (2004) para a identificação de subespaços, J.-M.Lee et al. (2004) com *DICA*, análise de componentes independentes dinâmica e Karim et al. (2007) com uma aplicação num processo químico.

Apesar do método de análise de componentes principais dinâmica (*DPCA*) ter mostrado uma boa eficiência na deteção de falhas de ocorrência em séries com dados autocorrelacionados, alguns autores alertam para o facto de que este método dinâmico pode não eliminar a autocorrelação e correlação transversal das variáveis, independentemente do tempo de latência que seja considerado (Xie et al, 2006; Kruger et al, 2004). Mesmo quando as

variáveis do processo não apresentam autocorrelação, nem correlação cruzada, as variáveis dos componentes principais obtidas através da estrutura *DPCA* exibem automaticamente autocorrelação. A fim de superar a limitação do método *DPCA* para eliminar a correlação automática e cruzada a partir de variáveis de componentes principais e valores de monitorização de  $T^2$ , este trabalho apresenta uma nova carta de controlo, baseada numa nova e diferente estrutura de matriz de dados. A nova carta de controlo é denominada por *DMPCA* (*Deployed Matrix PCA*) onde os dados das variáveis na matriz são desfasados em observações ímpares e pares agrupados por coluna, permitindo a anulação da autocorrelação nas variáveis dos dados originais.

### 2.10.1. Carta DPCA e a Nova Proposta DMPCA

A nova carta de controlo proposta, *DMPCA*, é semelhante à carta proposta por Ku et al. (1995), carta dinâmica *PCA* (*DPCA*), onde as principais diferenças entre carta proposta e esta última se centra na estrutura da matriz.

Enquanto a carta proposta por Ku et al. (1995), *DPCA*, organiza as variáveis do processo para formar uma estrutura autorregressiva (*AR*) (Xie et al., 2004), com o objectivo de extrair as relações dependentes do tempo nas medições (Ku et al., 1995) dadas pela equação (2.81), a estrutura da nova matriz proposta para a carta de controlo *DMPCA* tem a intenção de destruir qualquer autocorrelação dos dados existente a partir das variáveis de processo. Esta destruição da autocorrelação existente é conseguida através da separação de cada variável dos dados nas colunas pares e ímpares, tal como expresso na equação (2.82).

- **Matriz DPCA**

$$\mathbf{X}(l) = \begin{bmatrix} \mathbf{X}_t^T & \mathbf{X}_{t-1}^T & \cdots & \mathbf{X}_{t-l}^T \\ \mathbf{X}_{t+1}^T & \mathbf{X}_t^T & \cdots & \mathbf{X}_{t-l+1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{t+m}^T & \mathbf{X}_{t+m-1}^T & \cdots & \mathbf{X}_{t+m-l}^T \end{bmatrix} \quad (2.81)$$

- **Matriz DMPCA**

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_t^T & \mathbf{X}_{t+1}^T \\ \mathbf{X}_{t+2}^T & \mathbf{X}_{t+3}^T \\ \dots & \vdots \\ \mathbf{X}_{t+(m-2)}^T & \mathbf{X}_{t+(m-1)}^T \end{bmatrix} \quad (2.82)$$

onde

$\mathbf{X}(l) \in \mathfrak{R}^{(m-1)(l+1)m}$  e  $\mathbf{X} \in \mathfrak{R}^{\left(\frac{m}{2}\right)(2m)}$  são os dados para as matrizes *DPCA* e *DMPCA* respetivamente para as amostras  $m$ ,  $m$  variáveis a serem monitorizadas e  $l$  número de medições desfasadas.

O procedimento utilizado para obter as estatísticas complementares, Hotelling  $T^2$  e  $Q$ , é essencialmente a mesma que a abordagem *PCA* original para ambas as cartas de controlo (*DPCA* e *DMPCA*), com a exceção da matriz de dados inicial.

## 2.11. SÍNTESE DO CAPÍTULO

O objetivo deste capítulo, foi apresentar primeiramente a base que está por detrás de todo o estudo a que se propõe esta dissertação. Foram igualmente apresentados outros pontos que, embora não tenham sido aplicados directamente, foram considerados como parte integrante do controlo estatístico univariado e multivariado. Iniciou-se por uma curta descrição resumida sobre o controlo estatístico do processo, evidenciando e caracterizando as cartas de controlo de Shewart, a carta *CUSUM* e a carta *EWMA*. A principal conclusão a retirar do levantamento bibliográfico realizado, é que as cartas de Shewart deixam de ser realmente eficazes e eficientes quando o objetivo é detetar pequenas e moderadas alterações dos parâmetros do processo, constatando-se que é uma ferramenta mais útil para situações em que as alterações são significativas. Ao contrário das cartas de Shewart, com a carta *CUSUM* e *EWMA* detetou-se que estas são mais sensíveis a pequenas e moderadas alterações no processo e como tal devem ser a aplicar nestes casos mais minuciosos.

De acordo com as necessidades do mercado dos dias de hoje, principalmente quando se fala de indústrias produtivas bastante desenvolvidas e com processos complexos, torna-se cada vez mais necessário o controlo das operações e análise do processo com dados reais em tempo real. É assim cada vez mais real, a presença de autocorrelação nos dados de um dado processo produtivo, uma vez que os processos são realizados cada vez mais rapidamente e deste modo, o espaço de tempo entre um acontecimento e próximo, é cada vez mais curto no tempo. Assim torna-se essencial um estudo do controlo estatístico do processo com dados autocorrelacionados. Neste capítulo, o estudo do controlo estatístico do processo com dados autocorrelacionados dividiu-se em duas partes. Uma primeira parte onde se expôs o caso do controlo estatístico do processo em que os dados são autocorrelacionados e se está perante apenas uma variável (característica da qualidade) e uma segunda parte, onde se tem os mesmos dados autocorrelacionados mas tem-se também multivariáveis, mais do que uma característica da qualidade para monitorizar.

Para o caso do estudo do controlo estatístico univariado com dados autocorrelacionados, foram consideradas duas abordagens para a construção das cartas de controlo. A primeira diz respeito à apresentação de várias cartas utilizadas para o efeito univariado, entre elas, a carta *MCEWMA* e a carta *EWMAST*. A segunda abordagem, necessita de um modelo matemático para determinar os valores previstos (ou resíduos), no qual se anunciaram algumas cartas, nomeadamente, a carta *SCC* e as cartas *CUSUM* e *EWMA* ambas de resíduos.

Para o caso do estudo do controlo estatístico multivariado com dados autocorrelacionados, foram consideradas as cartas tradicionais ( $T^2$ , *MCUSUM*, *MEWMA*) que mostraram ser de extrema importância para os processos industriais quando se pretende controlar mais do que uma variável, característica da qualidade, com dados autocorrelacionados. No entanto, apesar de conseguirem monitorizar várias variáveis, conclui-se também que estas perdem a eficiência e a eficácia quanto à deteção de alterações no processo, quando o número de variáveis a controlar é elevado. Para os casos em que seja necessário monitorizar muitas variáveis propõe-se a aplicação dos métodos de projeção em deferimento das cartas anteriormente referidas.

Já o estudo da carta de controlo baseada na análise de componentes principais dinâmica (*DPCA*), mostrou-se bastante útil para quando os dados são autocorrelacionados, uma vez que permite eliminar a autocorrelação dos dados, evitando assim a nova modelação do processo. Esta carta, *DPCA*, tornou-se também um sucesso comparada com a carta *PCA* estática, uma

vez que irá revelar a relação exata entre as várias variáveis. No entanto, muitos autores alertaram para o facto de que este método dinâmico pode não eliminar a autocorrelação das variáveis, independentemente do tempo de latência que seja considerado.

Desta forma, existiu a necessidade de propor e criar uma nova carta de controlo que permitisse ultrapassar este problema, uma nova carta baseada no método de análise de componentes principais. Esta carta foi denominada de *DMPCA*, e é semelhante à proposta por Ku et al. (1995), carta *DPCA*, mas com uma nova e diferente estrutura de matriz de dados. Na carta *DMPCA*, os dados das variáveis na matriz são desfasados em observações ímpares e pares agrupados por coluna, permitindo a anulação da autocorrelação nas variáveis dos dados originais. O objetivo desta nova proposta, através da sua nova estrutura matricial, é eliminar qualquer autocorrelação dos dados existentes a partir das variáveis de processo.



### 3. METODOLOGIA

No presente capítulo, pretende-se sugerir uma metodologia para a realização do estudo comparativo entre as três cartas multivariadas com dados autocorrelacionados, carta  $T^2$  de Hotelling, carta *DPCA* e a nova proposta de carta *DMPCA*. Para uma melhor compreensão e interpretação da metodologia adotada para o estudo apresentado, recorreu-se à utilização de fluxogramas para descrever as várias etapas da construção e desenvolvimento das cartas multivariadas com dados autocorrelacionados, isto tudo com base em toda a pesquisa bibliográfica presente no capítulo anterior (Capítulo 2).

#### 3.1. METODOLOGIA PROPOSTA

A fim de evidenciar a eficiência da nova carta *DMPCA* proposta, é realizada uma comparação não só com a carta *DPCA*, mas também com a carta multivariada para dados autocorrelacionados tradicional  $T^2$  de Hotelling bem conhecida (com resíduos). A carta de controlo  $T^2$  de Hotelling é também utilizada como um meio para estabelecer uma linha de base para o estudo do controlo estatístico do processo multivariado *MSPC* (pior caso). A medida de desempenho utilizada para comparar as três cartas de controlo foi o *ARL* e o correspondente *SDRL*.

Todo o estudo comparativo foi realizado através do processo de simulação implementado num *software*, MATLAB<sup>®</sup>. Um conjunto de três variáveis foram utilizadas neste estudo segundo dois processos dinâmicos diferentes, um autoregressivo de primeira ordem, *AR(1)*, e um autoregressivo de média móvel de primeira ordem, *ARMA(1,1)*.

Assumindo a incapacidade da estrutura da matriz destruir a possível correlação cruzada entre as variáveis como uma limitação conhecida desta nova carta *DMPCA*, este estudo assume que todas as três variáveis são autocorrelacionadas entre si mas não interligadas (sem correlação cruzada).

Com o intuito de tornar todas as cartas de controlo comparável de uma forma justa, o  $ARL_{Em\ Controlo}$ , foi considerado 370, que corresponde ao limite superior de controlo (*LSC*) da carta de Hotelling  $T^2$ , erro de Tipo I com  $\alpha$  de 0,27%. Tanto para a carta de controlo *DPCA* como para a *DMPCA*, o *LSC* da carta  $T^2$  também utiliza um  $\alpha$  igual 0,27%. No que diz

respeito à carta  $Q$ , o  $LSC$  foi obtido por simulação, a fim de garantir uma  $ARL_{Em\ Controlo}$  de 370 para a combinação da carta  $T^2$  e a carta  $Q$ . Tendo em consideração o esforço computacional subjacente às simulações necessárias para este estudo, as condições utilizadas para a obtenção  $LSC$  em simulação para a carta  $Q$  envolveu 50.000 ciclos, com 5.000 e 10.000 observações, respectivamente, para  $DPCA$  e  $DMPCA$  (o número de observações dobra para o caso de  $DMPCA$  gráfico porque a matriz é desfasada em observações pares e ímpares). Para calcular os valores  $ARL_{Fora\ de\ Controlo}$ , as condições de simulação foram reduzidas para 30.000 ciclos, com 3.000 e 6.000 observações, respectivamente, para  $DPCA$  e  $DMPCA$ .

A Tabela 3.1 apresenta cada um dos cinco cenários considerados e o seu modelo de série temporal correspondente e os parâmetros autorregressivos.

**Tabela 3.1 – Configuração do Processo de Simulação dos cinco cenários de estudo com os respetivos parâmetros do processo.**

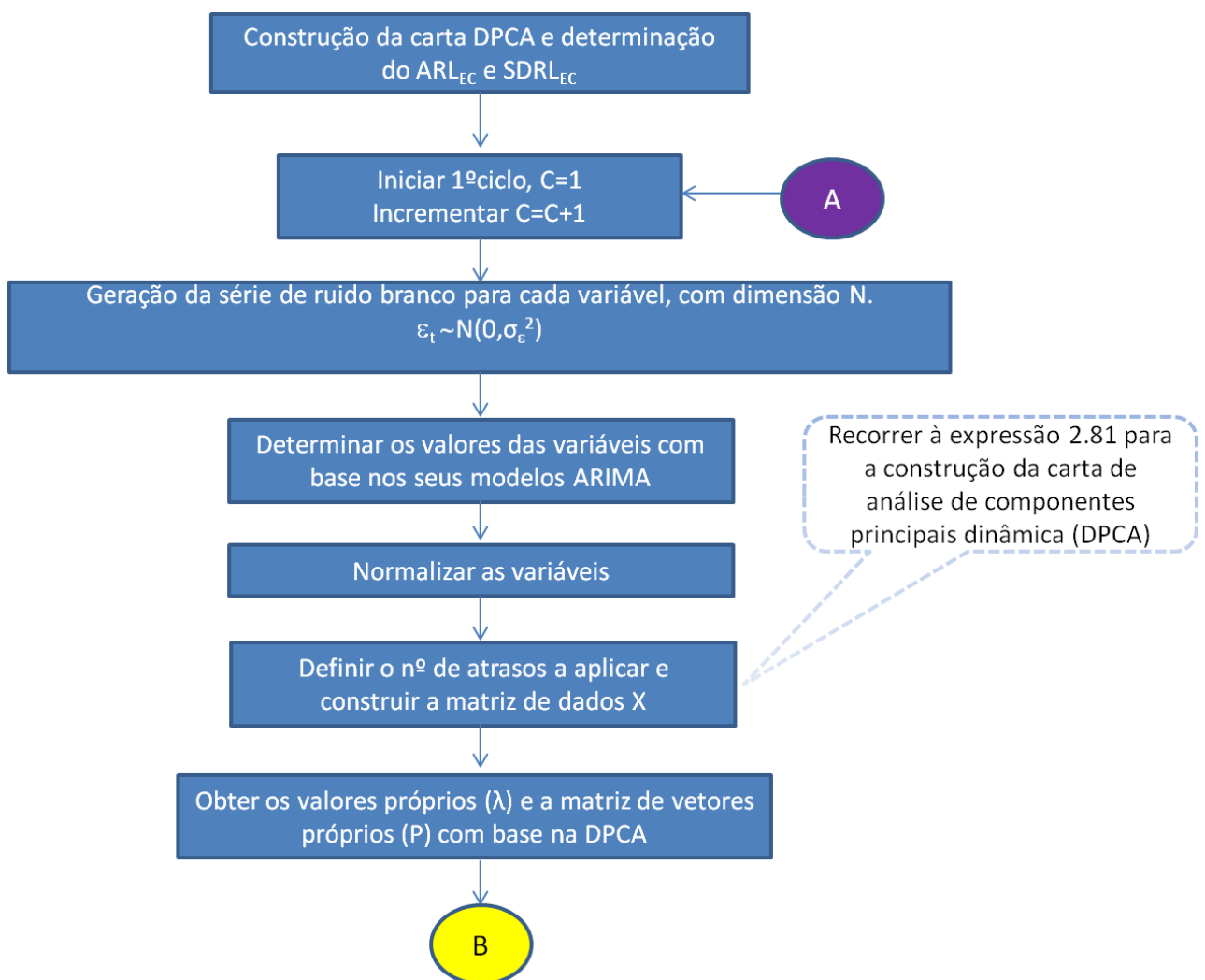
Cenários	Processos	Parâmetros do Processo
1	AR(1)	$\phi_1=\phi_2=\phi_3=0.8$ (autocorrelação muito forte)
2		$\phi_1=0.7, \phi_2=\phi_3=0.5$ (autocorrelação forte)
3		$\phi_1=0.7, \phi_2=0.5, \phi_3=0.2$ (autocorrelação forte)
4		$\phi_1=0.5, \phi_2=\phi_3=0.2$ (autocorrelação fraca)
5	ARMA(1,1)	$\phi_1=0.7, \phi_2=\phi_3=0.5$ (autocorrelação forte) $\theta_1=0.6, \theta_2=0.4, \theta_3=0.6$

Toda esta informação será utilizada como base para todos os estudos e análises referidas nos capítulos anteriores. Com esta base apresentada neste subcapítulo, o objetivo é facultar toda a informação necessária e essencial para a construção, desenvolvimento e análise da nova carta,  $DMPCA$ , e procedimentos para a sua aplicabilidade, apresentando as suas vantagens e desvantagens para as indústrias com processos que apresentem várias características da qualidade e com dados autocorrelacionados. Isto tudo para explicar e evidenciar a importância da escolha dos vários cenários e modelos referidos, para de certa forma tentar descrever o melhor possível a realidade industrial, uma vez que estes estudos apenas fazem sentido se contiverem alguma aplicabilidade na realidade e no dia-a-dia nos processos industriais com o fim da melhoria contínua dos negócios.

### 3.2. CONSTRUÇÃO DO PROGRAMA DE SIMULAÇÃO

Os modelos referidos ao longo do estudo serão produzidos por base em simulação através do *software* MATLAB<sup>®</sup>. No entanto, para a realização e desenvolvimento das simulações tem que se definir bem o número de ciclos a efetuar assim como, o número de observações de cada simulação, para se conseguir garantir qualidade nos dados e resultados a extrair para uma posterior boa análise da realidade.

Desta forma, apresenta-se abaixo um esquema resumido (Figura 3.1, Figura 3.2, Figura 3.3 e Figura 3.4), que permite indicar as várias etapas que irão caracterizar a metodologia a adotar para o desenvolvimento e estudo de desempenho da carta *DPCA* e da nova carta proposta *DMPCA* (*Deployed Matrix Principal Components Analysis*).



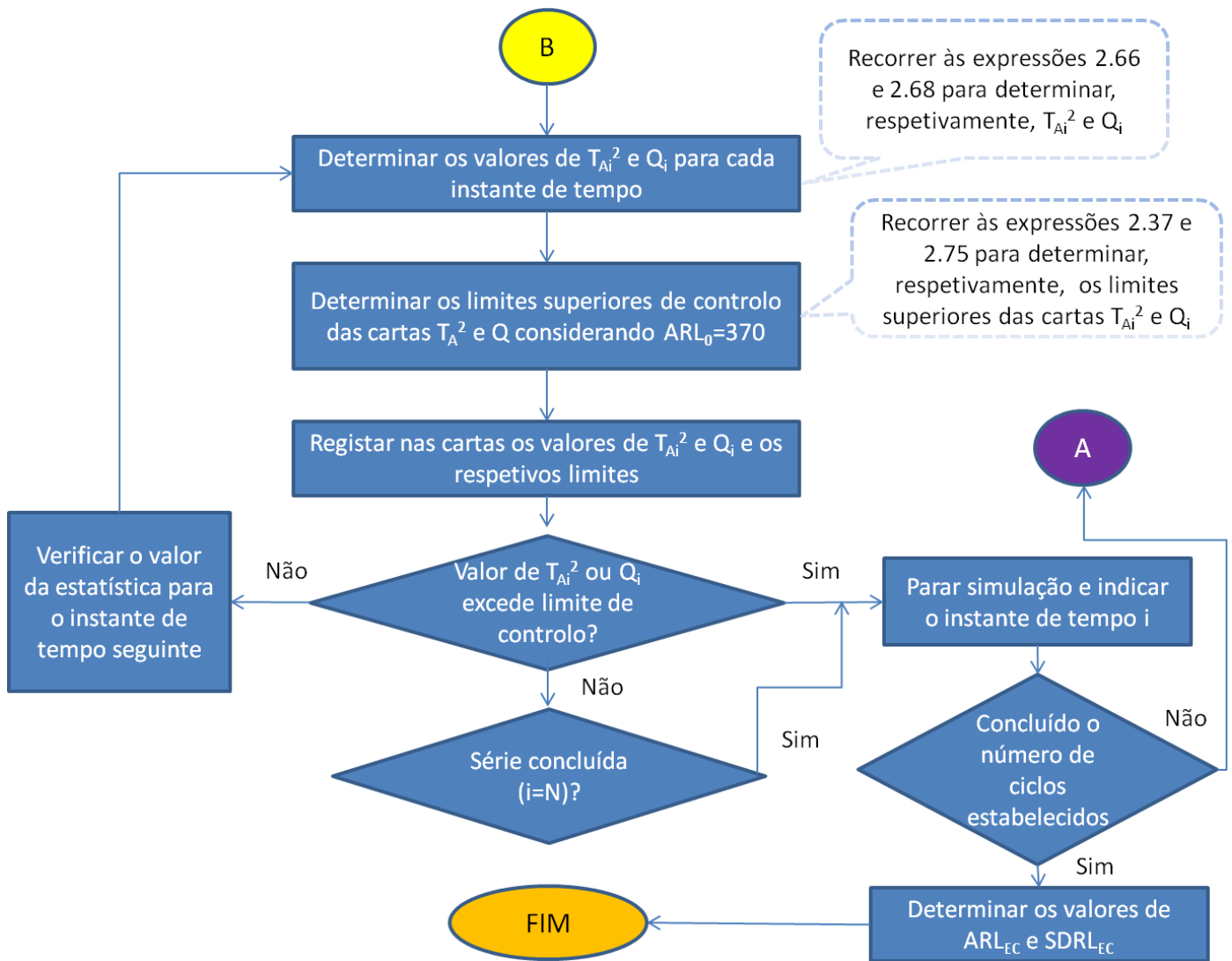
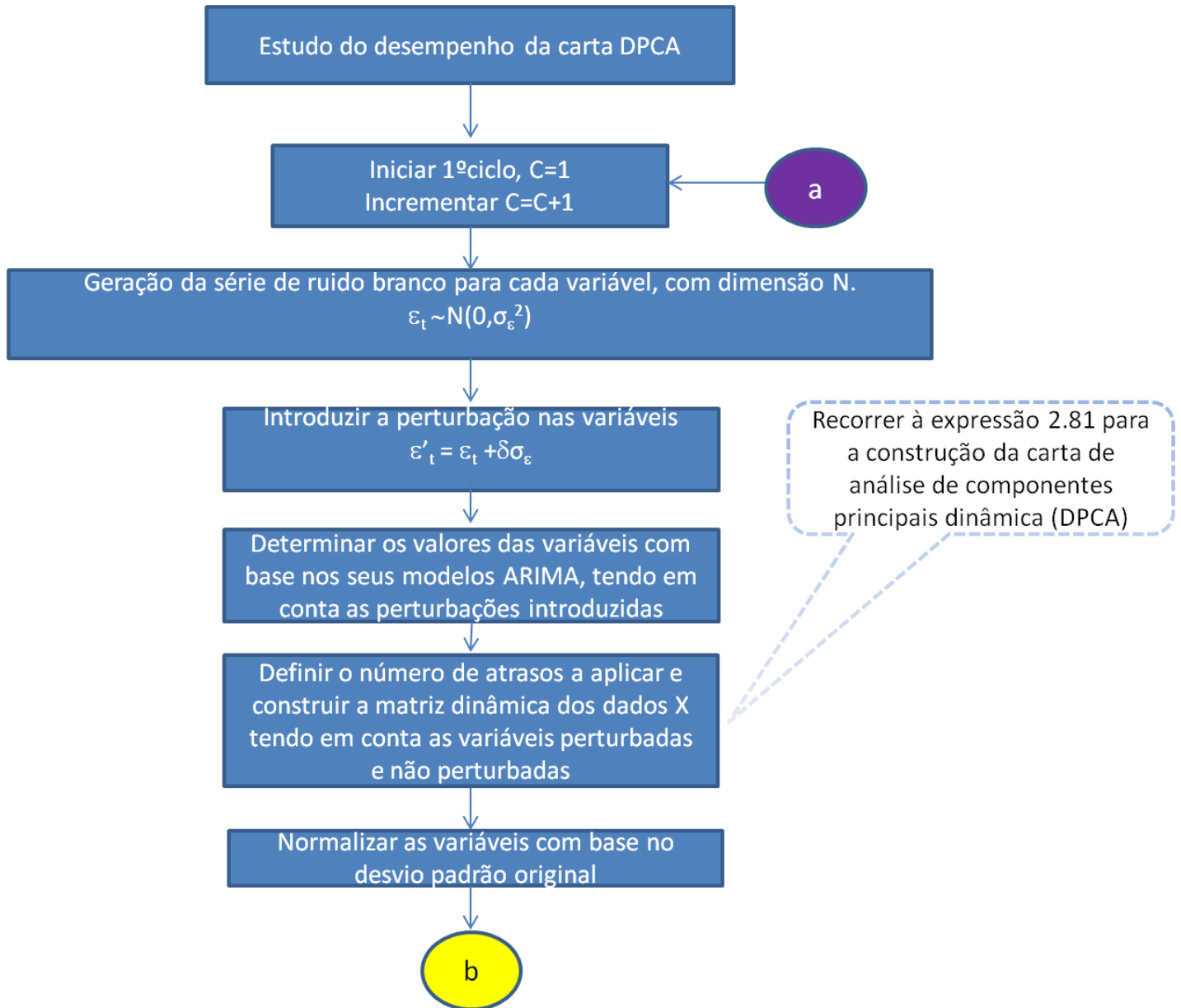


Figura 3.1 - Construção da carta DPCA e determinação do  $ARL_{EC}$  e  $SDRL_{EC}$



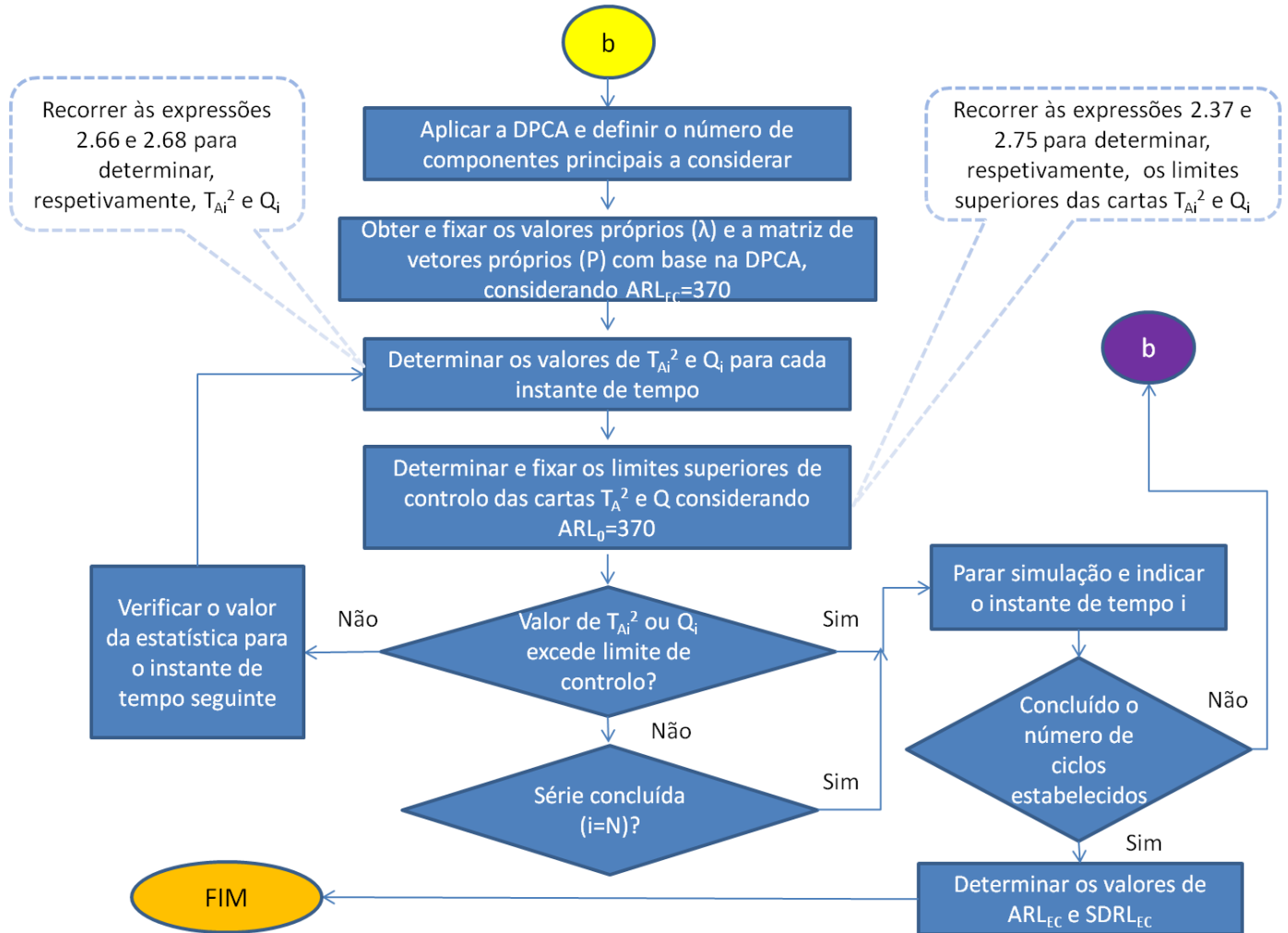
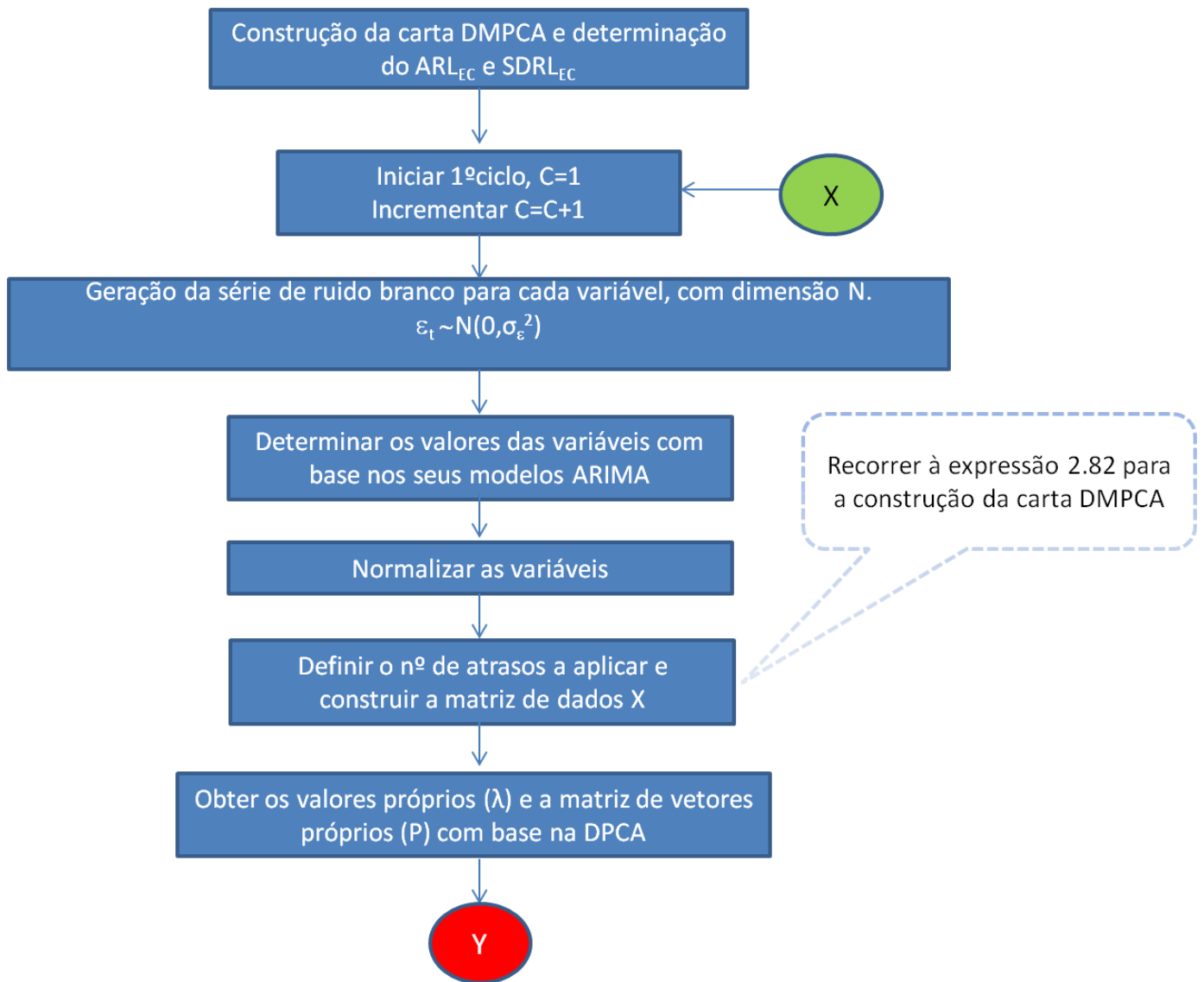


Figura 3.2 - Estudo do desempenho da carta DPCA a alterações no parâmetro médio



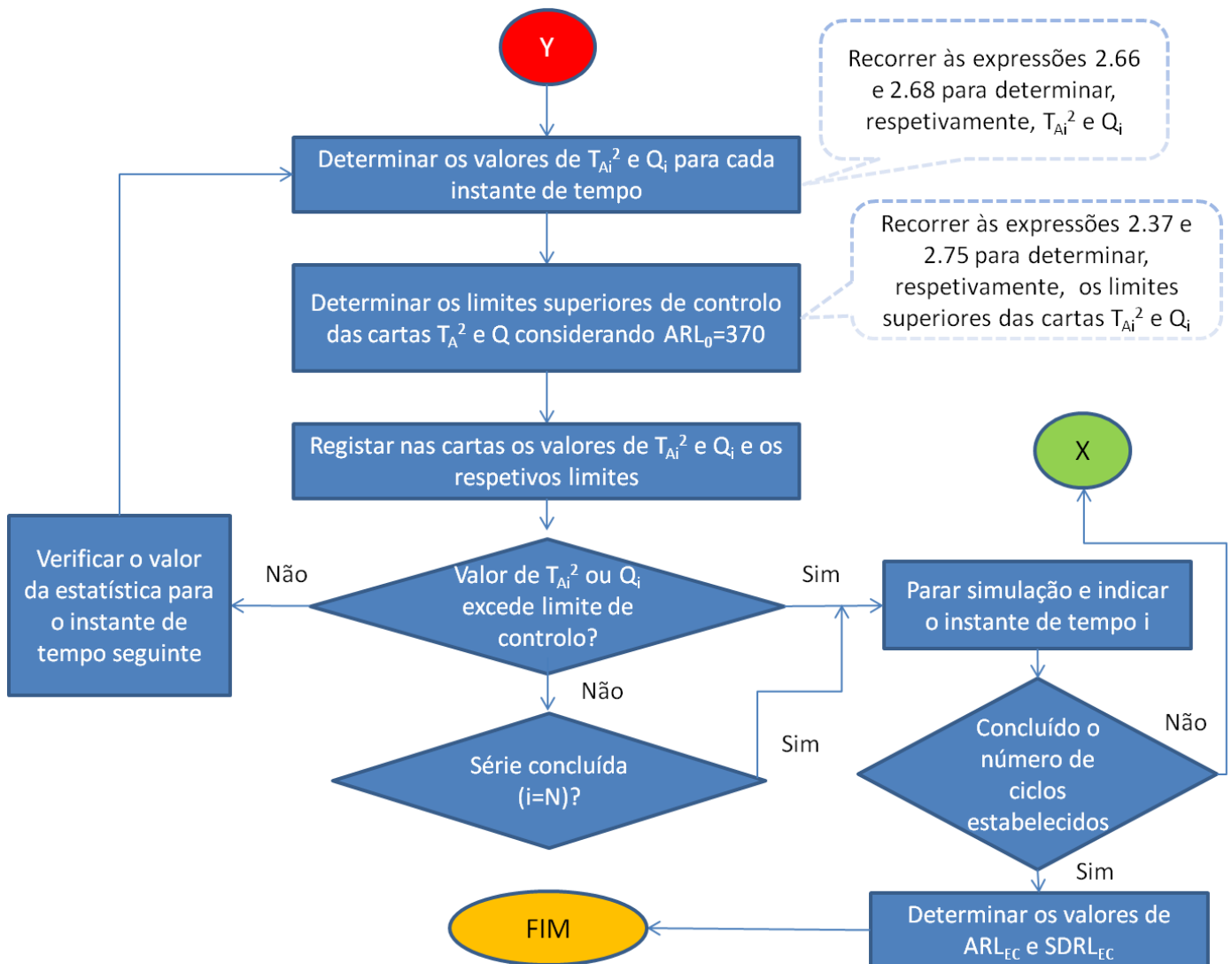
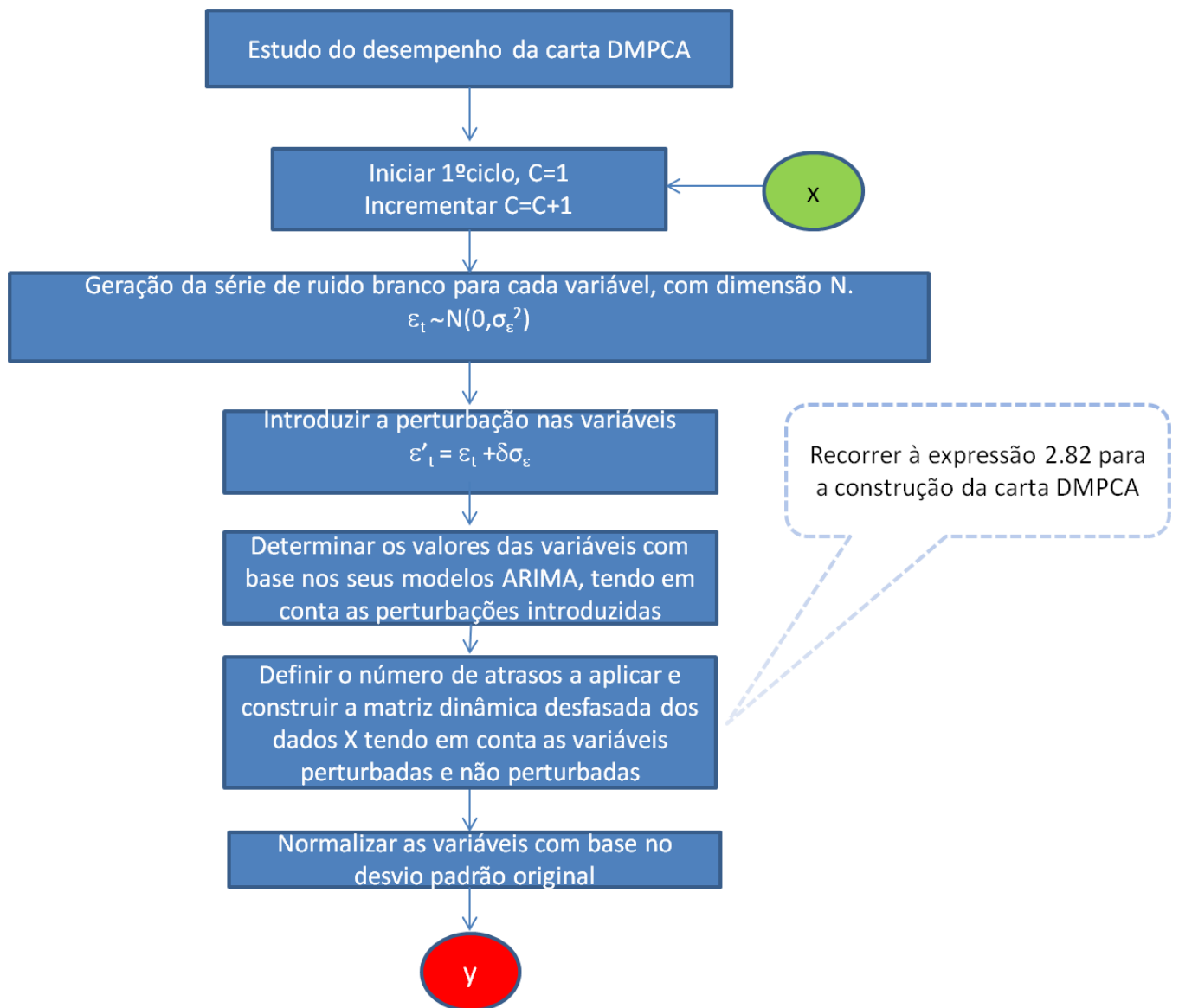


Figura 3.3 - Construção da carta DMPCA e determinação do  $ARL_{EC}$  e  $SDRL_{EC}$



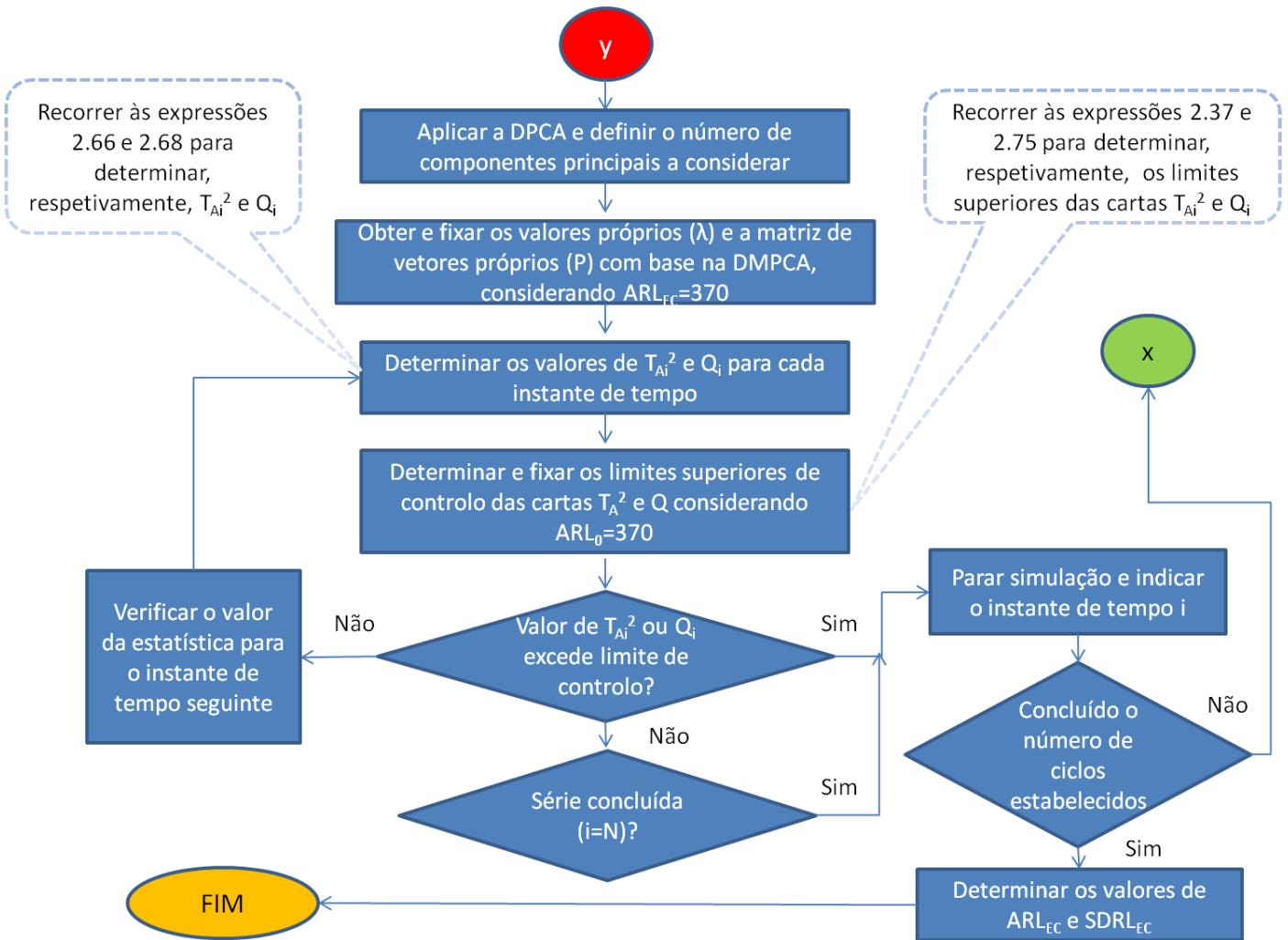


Figura 3.4 - Estudo do desempenho da carta DMPCA a alterações no parâmetro médio

### 3.3. COMPARAÇÃO DO DESEMPENHO DAS CARTAS DE CONTROLO

Tal como foi sendo referido ao longo do estudo, as medidas de desempenho que serão utilizadas como base de análise comparativa entre as várias cartas de controlo multivariadas são o *ARL* e o seu respetivo desvio padrão, *SDRL*. Por norma, na maioria dos estudos, a análise dos valores de *SDRL* tende a ser ignorada, no entanto, a sua interpretação, juntamente com o valor de *ARL*, assume um papel extremamente importante. Uma razão para não se descurar a análise do *SDRL* é por exemplo, os casos em que o *ARL* apresenta valores baixos mas o seu valor de *SDRL* comparativamente com as outras cartas é elevado, isto pode significar que a carta em causa apresenta um comportamento instável, assim percebe-se que o valor *SDRL* permite auxiliar a tomada de decisão de qual a carta mais adequada para o caso. A conclusão supracitada, será de extrema importância pois irá apoiar a tomada de decisão de qual a melhor carta a adotar, quando o processo é sujeito a um tipo de perturbação numa ou mais variáveis em simultâneo.

### 3.4. INTERPRETAÇÃO DAS CARTAS DE CONTROLO MULTIVARIADAS

Na presente dissertação, é essencial realizar uma boa interpretação das cartas de controlo multivariadas para posteriormente se realizar uma boa análise comparativa. Para tal, é necessário que seja fácil a deteção de falhas, deteção de causas especiais, por exemplo pontos fora dos limites de controlo. No capítulo 4 e nos seus subcapítulos respetivos, são apresentadas figuras e tabelas que mostram a forma como foram interpretadas as cartas de controlo multivariadas e como foi feita a comparação das três cartas. Desta forma, identificaram-se como pontos essenciais para a interpretação das cartas de controlo multivariadas, os seguintes:

- Determinação dos Limites de Controlo – Cálculo do valor do Limite Superior de Controlo (*LSC*) para um  $ARL_{EC} = 370$ , para cada uma das cartas tendo em conta os cinco cenários e os parâmetros de autocorrelação definidos para cada um dos processos, *AR*(1) e *ARMA*(1,1);

- Comparação do desempenho das cartas de controlo multivariadas – Cálculo o valor de cada medida de desempenho,  $ARL$  e  $SDRL$ , para cada uma das cartas tendo em conta os cinco cenários, os vários níveis do tamanho de alteração da média ( $\delta$ ) definidos e os parâmetros de autocorrelação para os dois processos,  $AR(1)$  e  $ARMA(1,1)$ . Através destes dados realizou-se a análise de comparação e desempenho das três cartas;
- Validação dos pressupostos das cartas de controlo multivariadas – Apresenta as figuras da Função de Autocorrelação ( $FAC$ ) e da Função de Autocorrelação Parcial ( $FACP$ ) para analisar se a nova carta proposta ( $DMPCA$ ) elimina a presença de autocorrelação na estatística  $T^2$  e se esta assume violação ou não o pressuposto de independência no tempo. Para ajudar nesta análise e para retirar conclusões para diversos ambientes, elaboraram-se as figuras das funções  $FAC$  e  $FACP$ , para os vários cenários definidos, tanto para a estatística  $T^2$  e  $Q$  para cada uma das cartas baseadas na análise de componentes principais.
- Ganhos da carta  $DMPCA$  – Apresenta de forma gráfica, os ganhos que a carta  $DMPCA$  tem em relação à carta  $T^2$  e à carta  $DPCA$ , no que concerne às medidas de desempenho utilizadas neste estudo,  $ARL$  e  $SDRL$ .

### 3.5. APLICAÇÃO PRÁTICA

A aplicação prática corresponde à última fase da metodologia proposta. Esta fase tem como principal objetivo identificar as vantagens e desvantagens das cartas de controlo na ótica do utilizador que irá aplicar a carta.

## 4. ANÁLISE DOS RESULTADOS

Este capítulo tem como principal objetivo analisar o desempenho das três cartas de controlo multivariadas com dados autocorrelacionados. No primeiro ponto apresentam-se os limites de controlo que se determinaram para os cinco cenários propostos. No segundo ponto, realiza-se uma análise de sensibilidade entre as três cartas, nomeadamente a carta *DMPCA* e a carta *DPCA*, através das medidas de desempenhos utilizadas, *ARL* e *SDRL*, para os dois processos em causa (*AR(1)* e *ARMA(1,1)*). No terceiro, apresenta-se um subcapítulo dedicado à validação dos pressupostos das cartas de controlo multivariadas. No quarto ponto são apresentados os ganhos da nova carta proposta, *DMPCA*, relativamente às outras cartas de controlo multivariadas. Por último, no quinto ponto apresentam-se as vantagens e desvantagens das três cartas.

### 4.1. DETERMINAÇÃO DOS LIMITES DE CONTROLO

Para o estudo das cartas multivariadas com dados autocorrelacionados, nomeadamente para a carta *DPCA* e para a nova carta proposta *DMPCA*, é essencial que se calcule os limites de controlo para ambas as cartas.

A determinação dos limites de controlo foi feita através de simulação no *software* MATLAB, com base nas séries e número de observações definidos no capítulo anterior, substituindo os parâmetros autorregressivos desde a autocorrelação muito forte até autocorrelação fraca, isto para todos os cinco cenários estabelecidos. De salientar que apenas se estudou dois processos de autocorrelação, *AR(1)* e *ARMA(1,1)*, onde para o primeiro foram considerados quatro cenários para diferentes níveis de autocorrelação e para o segundo apenas foi considerado o cenário sobre forte presença de autocorrelação.

A Tabela 4.1 apresenta cada um dos cinco cenários considerados e o seu modelo de série temporal correspondente, parâmetros autoregressivos e de média móvel, bem como o correspondente *LSC* para a estatística *Q*.

**Tabela 4.1 – Configuração do Processo de Simulação e o Limite Superior de Controle (LSC) para a estatística  $Q$  ( $ARL_{Em\ Controlo} = 370$ )**

Cenários	Processos	Parâmetros do Processo	Q Limite Superior de Controlo	
			DPCA	DMPCA
1	AR(1)	$\phi_1=\phi_2=\phi_3=0,8$ (autocorrelação muito forte)	3,17	3,58
2		$\phi_1=0,7, \phi_2=\phi_3=0,5$ (autocorrelação forte)	7,42	8,80
3		$\phi_1=0,7, \phi_2=0,5, \phi_3=0,2$ (autocorrelação forte)	10,42	12,32
4		$\phi_1=0,5, \phi_2=\phi_3=0,2$ (autocorrelação fraca)	14,42	15,32
5	ARMA(1,1)	$\phi_1=0,7, \phi_2=\phi_3=0,5$ (autocorrelação forte) $\theta_1=0,6, \theta_2=0,4, \theta_3=0,6$	13,46	16,15

#### 4.2. COMPARAÇÃO DO DESEMPENHO DAS CARTAS DE CONTROLO MULTIVARIADAS

Neste presente subcapítulo, apresenta-se de forma sintética, através de um quadro, um resumo da comparação realizada entre as três cartas de controlo, carta de controlo  $T^2$  de Hotelling, carta *DPCA* e a nova carta de controlo proposta *DMPCA*. A construção da tabela 6 foi realizada através do estudo dos vários cenários apresentados no subcapítulo 4.1 com base em simulação através do *software* MATLAB. É importante realçar que para este estudo, os pressupostos de simulação foram idênticos para cada uma das cartas utilizadas, no entanto, foi necessário a realização de vários programas de simulação diferenciados por cada tipo de carta. As diferenças entre os vários programas concentraram-se na estrutura de construção das cartas, tal como supracitado e ilustrado no subcapítulo 3.2 e referido ao longo do capítulo 2. Para além da construção das cartas e dos programas de simulação na tabela 6, apresentam-se os resultados obtidos através da simulação para os cinco cenários de autocorrelação para os dois processos em causa (*AR(1)* e *ARMA(1,1)*), tomando como medidas de desempenho o *ARL* e o correspondente *SDRL* para determinar deslocamentos médios. O parâmetro  $\delta$  é o tamanho da alteração da média, medida em relação ao desvio padrão (nova média =  $\mu + \delta\sigma$ ) e com incrementos de 0,5, para apresentar o desenvolvimento do comportamento das cartas com pequenas variações da média.

A Tabela 4.2 mostra claramente que a carta  $T^2$  de Hotelling apresenta os piores resultados no que concerne às medidas de desempenho, ou seja, para o *ARL* e *SDRL*, para pequenas e grandes alterações na média.

**Tabela 4.2 - Comparação das medidas de desempenho, ARL e SDRL, considerando os cinco cenários escolhidos com vários graus de autocorrelação**

Cenários	Processo	$\phi_1$	$\phi_2$	$\phi_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\delta$	T <sup>2</sup> Hotelling		DPCA		DMPCA	
									ARL	SDRL	ARL	SDRL	ARL	SDRL
1	AR(1)	0,8	0,8	0,8				0,0	370,5	369,7	369,4	363,3	370,7	364,1
								0,5	235,5	232,5	67,5	63,5	45,0	41,0
								1,0	95,3	94,2	13,9	9,9	8,8	5,4
								1,5	38,6	38	6,1	3,4	4,4	1,8
								2,0	13,6	13,1	3,8	1,8	3,1	1,0
								2,5	6,6	6,1	2,7	1,2	2,5	0,7
								3,0	3,7	3,1	2,1	0,9	2,1	0,7
2	AR(1)	0,8	0,5	0,5				0,0	370,1	369,4	371,1	369	369,8	363,9
								0,5	229,6	228,5	62,5	58,1	42,6	38,7
								1,0	88,8	88,3	13,4	9,4	8,6	5,4
								1,5	33,1	32,5	6,1	3,3	4,4	1,7
								2,0	13,8	13,3	3,8	1,7	3,2	0,9
								2,5	6,6	6,1	2,7	1,1	2,6	0,6
								3,0	3,7	3,2	2,2	0,8	2,3	0,5
3	AR(1)	0,8	0,5	0,2				0,0	370,5	369,7	370,1	366,8	370,4	365,1
								0,5	231,8	227,2	62,8	57,8	41,9	38,1
								1,0	90,4	89,8	13,2	9,2	8,5	5,3
								1,5	34,7	34,1	6,0	3,2	4,4	1,8
								2,0	14,7	14,1	3,8	1,7	3,2	0,9
								2,5	7,3	6,8	2,7	1,1	2,6	0,6
								3,0	4,0	3,5	2,1	0,8	2,3	0,5
4	AR(1)	0,5	0,2	0,2				0,0	370,3	364,1	369,3	362,4	370,4	365,9
								0,5	229,6	227,7	110,0	109,0	89,7	86,5
								1,0	86,8	86,8	21,6	19,9	16,0	14,1
								1,5	31,2	30,7	7,1	5,7	5,3	3,5
								2,0	12,4	11,9	3,3	2,2	3,1	1,4
								2,5	5,8	5,3	2,1	1,2	2,3	0,7
								3,0	3,2	2,6	1,5	0,7	2,1	0,4
5	ARMA(1,1)	0,7	0,5	0,5	0,6	0,4	0,6	0,0	370,5	369,7	370,6	363,8	370	367
								0,5	235,5	232,5	83,4	79,7	84,9	80,8
								1,0	95,3	94,2	16,6	13,1	13	9,7
								1,5	38,6	38	6,4	4,4	5,2	2,7
								2,0	17,6	17,1	3,4	2,2	3,2	1,3
								2,5	9,1	8,6	2,1	1,3	2,4	0,7
								3,0	5,4	4,8	1,5	0,8	2,1	0,5

Em relação ao *ARL*, quando se compara a carta *DMPCA* com a carta *DPCA*, o desempenho de ambas são semelhantes para deslocamentos médios de magnitudes médias e grandes (1,5 a 3,0), no entanto para deslocamentos médios pequenos (0,5 a 1,0) o desempenho de *DMPCA* mostra ser superior. No que diz respeito ao *SDRL*, a carta *DMPCA* apresenta os menores valores, para todas as situações fora de controlo, e todos os cenários, o que torna esta carta mais fiável. Quando comparados os resultados dos cinco cenários, para as cartas *DMPCA* e *DPCA*, é visível que, com a diminuição do grau de autocorrelação, os valores de *ARL* fora do controlo aumentam, fazendo com que a nova carta proposta seja mais atraente para os processos com um grau de autocorrelação dos dados médio alto.

Os resultados da simulação para o último cenário, cenário 5, modelo de média móvel autorregressivos com três variáveis,  $ARMA(1,1)$ , são mostrados na Tabela 4.1 e na figura 4.5, para  $ARL$  e  $SDRL$ . As figuras 4.1, 4.2, 4.3, 4.4 e 4.5 (a), representativas dos cinco cenários estudados, mostram que até a um valor de 0,5 para o tamanho da alteração da média, ambas as cartas de controle,  $DPCA$  e  $DMPCA$  mostram um comportamento semelhante, porém quando este tamanho aumenta a carta de controle proposta  $DMPCA$  pode conseguir menores valores de  $ARL$ , o que se traduz em um melhor desempenho. Esta conclusão torna-se ainda mais importante quando são analisadas as figuras 4.1, 4.2, 4.3, 4.4 e 4.5 (b), onde a carta  $DMPCA$  também apresenta os menores valores de  $SDRL$ , conduzindo a resultados mais assertivos (menor dispersão).

De seguida, são apresentadas as curvas  $ARL$  e  $SDRL$  para os cinco cenários estabelecidos na tabela 4.1.

- **Cenário 1**

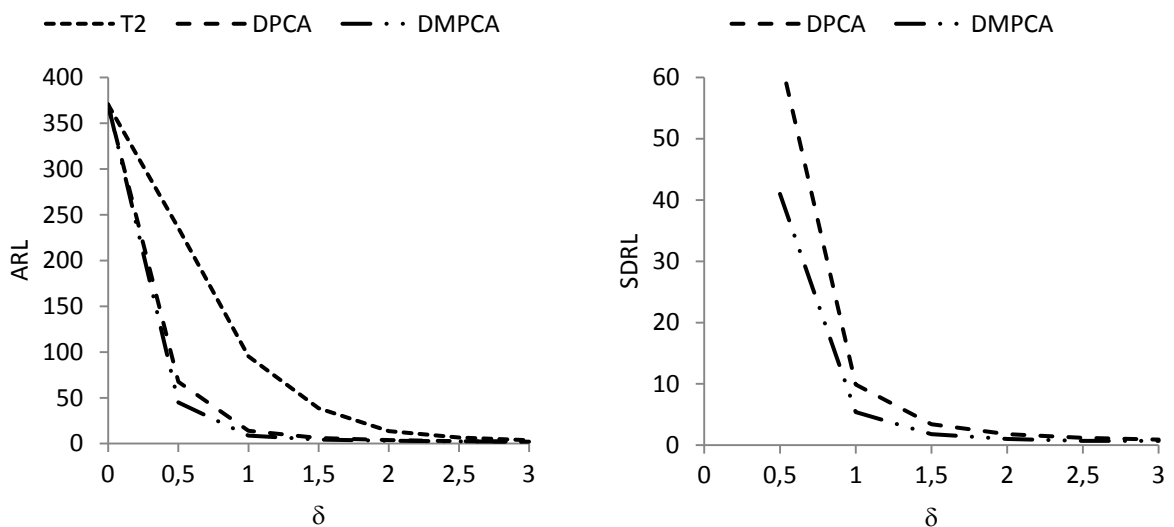


Figura 4.1 – Cenário 1: (a) curvas de  $ARL$  para a carta  $T^2$  de Hotelling,  $DPCA$  e  $DMPCA$ . (b) curvas  $SDRL$  para  $DPCA$  e  $DMPCA$ , considerando uma alteração na média acima de  $0,5 \delta$ .

- **Cenário 2**

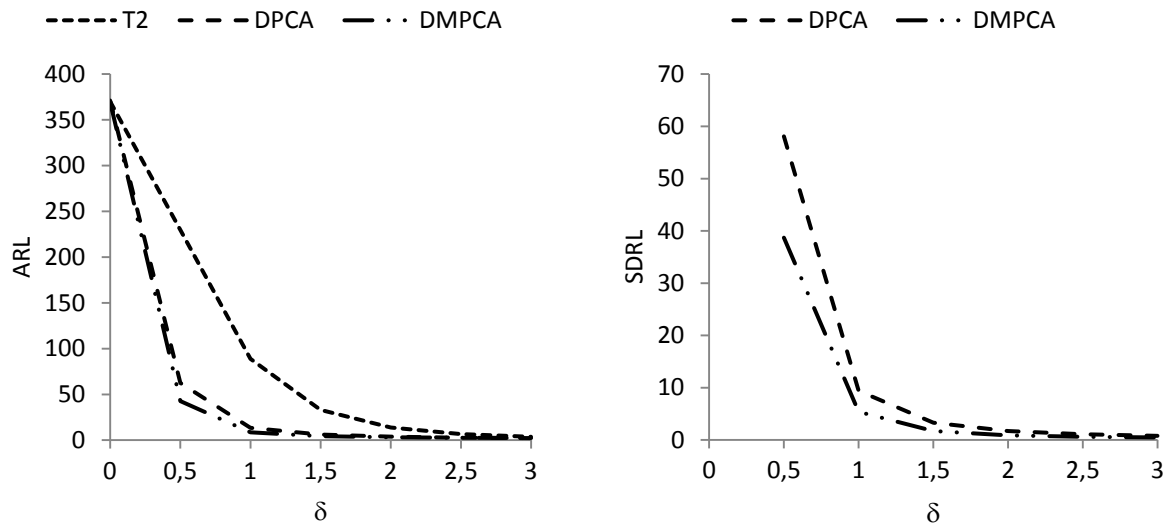


Figura 4.2 – Cenário 2: (a) curvas de  $ARL$  para a carta  $T^2$  de Hotelling,  $DPCA$  e  $DMPCA$ . (b) curvas  $SDRL$  para  $DPCA$  e  $DMPCA$ , considerando uma alteração na média acima de  $0,5 \delta$ .

- **Cenário 3**

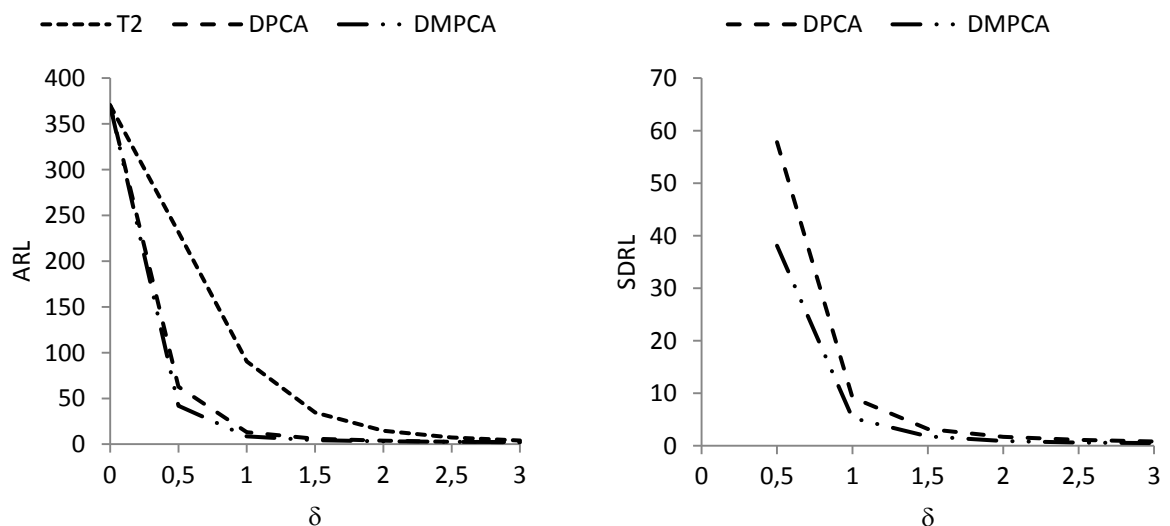


Figura 4.3 – Cenário 3: (a) curvas de  $ARL$  para a carta  $T^2$  de Hotelling,  $DPCA$  e  $DMPCA$ . (b) curvas  $SDRL$  para  $DPCA$  e  $DMPCA$ , considerando uma alteração na média acima de  $0,5 \delta$ .

- **Cenário 4**

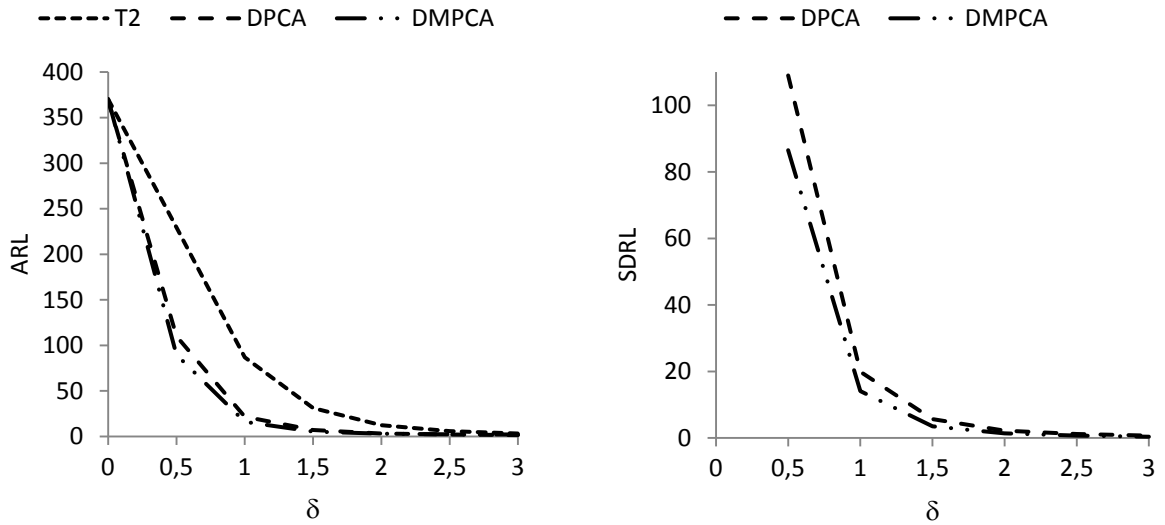


Figura 4.4 – Cenário 4: (a) curvas de ARL para a carta  $T^2$  de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de  $0,5 \delta$ .

- **Cenário 5**

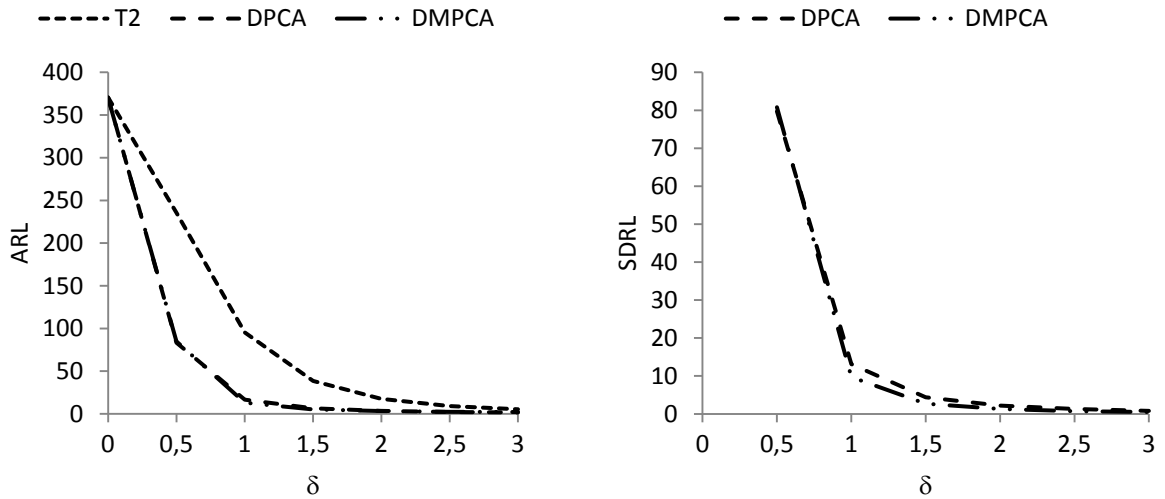


Figura 4.5 – Cenário 5: (a) curvas de ARL para a carta  $T^2$  de Hotelling, DPCA e DMPCA. (b) curvas SDRL para DPCA e DMPCA, considerando uma alteração na média acima de  $0,5 \delta$ .

Com base na interpretação gráfica dos vários cenários apresentados e analisando ao detalhe os dados apresentados na Tabela 6, verifica-se que a carta DMPCA é a mais eficaz pois apresenta

o melhor valor de  $ARL$ , independentemente das perturbações a que é submetida e seja qual for o cenário em causa, desde a autocorrelação muito forte até à autocorrelação fraca no modelo  $AR(1)$  e também no modelo  $ARMA(1,1)$  com autocorrelação muito forte. Com estes resultados, conclui-se que a carta  $DMPCA$  quando comparada com a carta de Hotelling  $T^2$  e a carta  $DPCA$ , é mais rápida na deteção de alterações na média do processo, mesmo quando estas são pequenas (0,5), médias (1 a 1,5) ou elevadas (2 a 3). Por outro lado, esta carta também é a que apresenta um valor de  $SDRL$  mais baixo para cada um dos cenários. Esta conclusão também é de extrema importância, uma vez que esta medida de desempenho tem um papel fundamental no apoio da tomada de decisão de qual a melhor carta a adotar quando se está perante um processo é sujeito a um tipo de perturbação numa ou mais variáveis em simultâneo, o que neste caso em específico, demonstra que esta carta quando comparada com as outras, apresenta um comportamento mais estável, ou seja, com menos discrepância. De salientar que nesta análise, a carta de Hotelling  $T^2$  não aparece nos gráficos acima apresentados de  $SDRL$  porque apresenta valores substancialmente mais elevados, não permitindo efetuar uma boa comparação entre as cartas  $DPCA$  e  $DMPCA$ , concluindo-se assim que o comportamento desta carta é bastante mais instável do que as restantes.

Os resultados presentes nas figuras acima indicadas em suma, mostram que, para turnos de 0,5, ambas as cartas de controlo,  $DPCA$  e  $DMPCA$  apresentam comportamentos semelhantes, porém, quando a magnitude da alteração da média aumenta, a carta de controlo proposta,  $DMPCA$  é capaz de atingir menores valores de  $ARL$ , o que se traduz em um melhor desempenho. Esta conclusão torna-se ainda mais evidente após a análise da medida de desempenho  $SDRL$ , onde a carta  $DMPCA$  também apresenta os menores valores  $SDRL$ , conduzindo a resultados mais assertivos (menor dispersão).

#### 4.3. VALIDAÇÃO DOS PRESSUPOSTOS DAS CARTAS DE CONTROLO MULTIVARIADAS

Este subcapítulo tem a intenção de provar que a nova carta proposta  $DMPCA$  com a sua estrutura matricial pode eliminar a presença de autocorrelação nas estatísticas  $T^2$ , ao contrário da carta  $DPCA$  que induz autocorrelação nas estatísticas  $T^2$  (mesmo quando as variáveis do processo não apresentam autocorrelação nem correlação cruzada). As figuras abaixo apresentadas (figuras 4.6 a 4.9 e figuras I.1 a VIII.2) e, para cada uma das cartas  $DPCA$  e

*DMPCA*, obtidas a partir de uma simulação com 1000 (*DPCA*) e 2001 (*DMPCA*) observações, onde foi aplicado cada um dos cenários (1, 2, 3, 4 e 5), para todos os tipos de comportamento dinâmico autorregressivo.

Nas figuras abaixo apresentadas (representativas das *FAC* e *FACP* da carta *DPCA*), para a maioria dos cenários propostos, é evidente a presença de autocorrelação principalmente na estatística  $T^2$ , calculada a partir dos primeiros três componentes principais, onde se consegue mostrar a violação do pressuposto de independência.

Em contraste, nas figuras representativas das *FAC* e *FACP* da carta *DMPCA*, mostra claramente que a nova proposta com a sua estrutura de matriz desfasada pode eliminar a autocorrelação das estatísticas  $T^2$  e, conseqüentemente, permite assumir a condição de independência. No entanto, para a carta *DMPCA*, os gráficos da Função de Autocorrelação (*FAC*) e Função de Autocorrelação Parcial (*FACP*) mostraram, igualmente, o ruído branco pesado nas estatísticas  $T^2$  de alguns dos cenários apresentados.

- **Cenário 1**

*DPCA* – Estatística  $T^2$

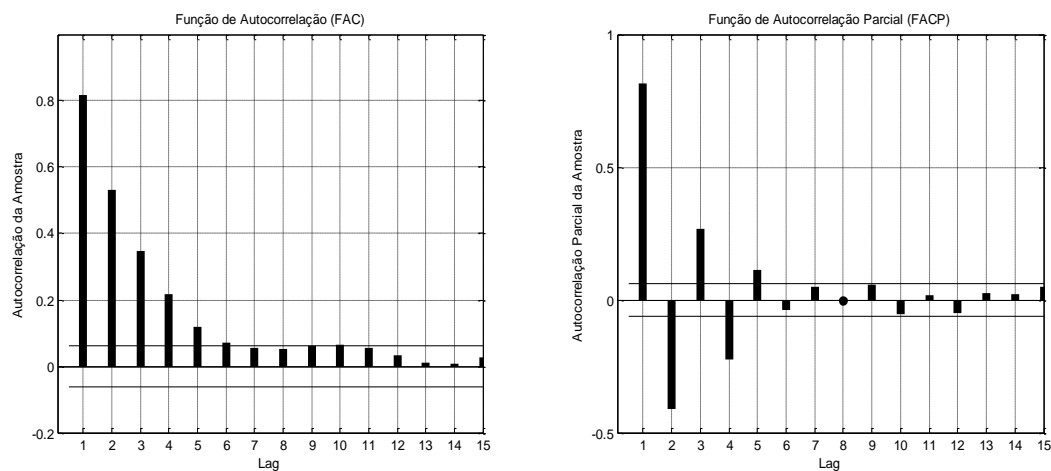


Figura 4.6 - FAC e FACP da carta DPCA para a estatística  $T^2$  (Cenário 1)

DPCA – Estatística  $Q$

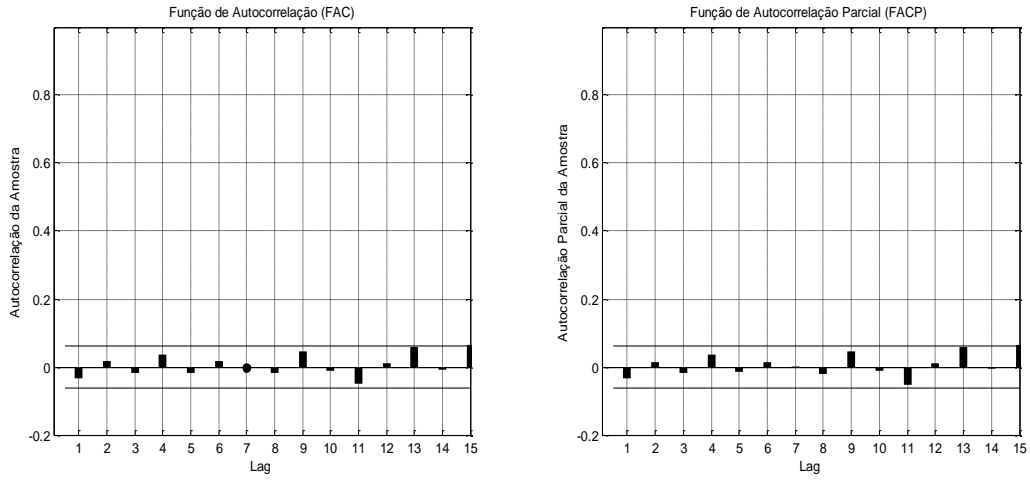


Figura 4.7 - FAC e FACP da carta DPCA para a estatística  $Q$  (Cenário 1)

DMPCA - Estatística  $T^2$

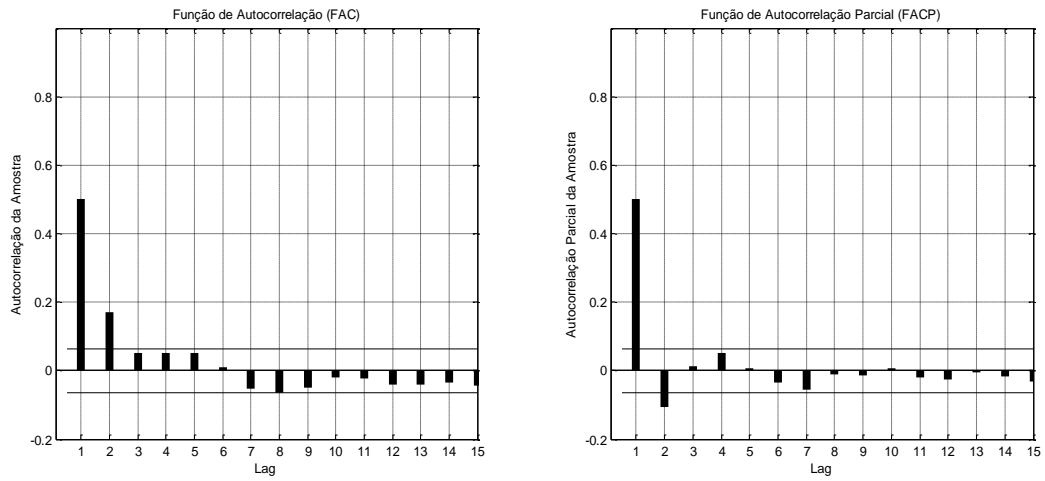
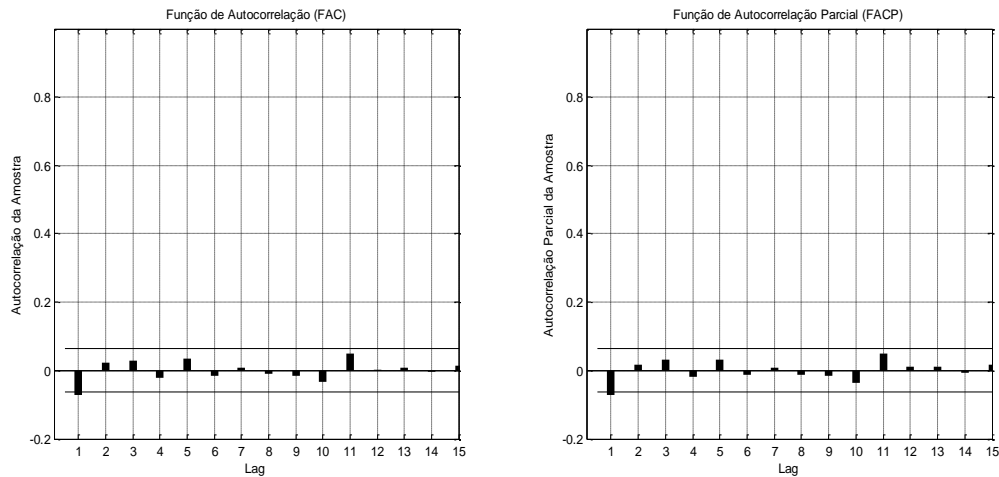


Figura 4.8 - FAC e FACP da carta DMPCA para a estatística  $T^2$  (Cenário 1)

*DMPCA* - Estatística  $Q$ 

**Figura 4.9 - FAC e FACP da carta DMPCA para a estatística  $Q$  (Cenário 1)**

A pressuposto de independência foi verificada para as cartas de controlo *DPCA* e *DMPCA* através da Função de Autocorrelação (*FAC*) e a Função de Autocorrelação Parcial (*FACP*) para as estatísticas  $T^2$ . Para todos os cenários, as estatísticas  $T^2$  obtidas pela carta de controlo *DPCA* falharam o pressuposto de independência, enquanto que para a nova carta proposta *DMPCA*, com sua nova estrutura da matriz desfasada a condição de independência pode ser assumida.

Em anexo, apresentam-se os resultados para os restantes cenários considerados para o estudo comparativo das cartas de controlo multivariadas com dados autocorrelacionados em análise, carta *DPCA* e *DMPCA*.

Os resultados para os restantes cenários, 2, 3, 4 e 5, refletem as mesmas conclusões já retirada para o cenário 1, a carta de controlo *DPCA* viola o pressuposto da independência das variáveis em estudo, principalmente para a estatística  $T^2$ , apresentando-se sempre fora dos limites da Função de Autocorrelação (*FAC*) e também da Função de Autocorrelação Parcial (*FACP*).

#### 4.4. GANHOS DA NOVA CARTA PROPOSTA, *DMPCA*, RELATIVAMENTE ÀS OUTRAS CARTAS DE CONTROLO MULTIVARIADAS

Com as análises realizadas nos subcapítulos anteriores, é notório que a nova carta proposta, *DMPCA*, apresenta muito melhores resultados no que concerne à deteção e eliminação da autocorrelação nas estatísticas  $T^2$  e  $Q$  relativamente às cartas  $T^2$  e carta *DPCA*.

Por outro lado, em relação às medidas de desempenho utilizadas para o estudo comparativo em causa, o *ARL* e o seu respetivo *SDRL*, mas mais especificamente em relação ao *ARL* é bastante evidente através dos gráficos apresentados abaixo (figuras 4.10 a 4.14), o ganho que a nova carta proposta, *DMPCA*, tem no que concerne ao *ARL* versus as outras duas cartas,  $T^2$  e *DPCA*. No entanto, há que salientar que para alguns cenários a partir de um certo grau de perturbação no parâmetro da média ( $\delta$ ) a carta *DPCA* apresenta melhores níveis de desempenho do que a nova carta de proposta *DMPCA*. Mas em todos os cenários e na amostra dos vários graus de alteração da média, a nova carta *DMPCA* apresenta ganhos muito significativos de *ARL* em comparação com a carta  $T^2$  e a carta *DPCA*.

- **Cenário 1**

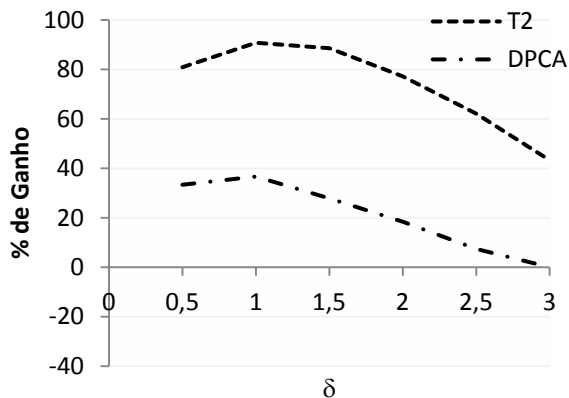


Figura 4.10 - Cenário 1: Os ganhos obtidos com a nova carta proposta *DMPCA* quando comparada com a carta  $T^2$  de Hotelling e a carta *DPCA*

- **Cenário 2**

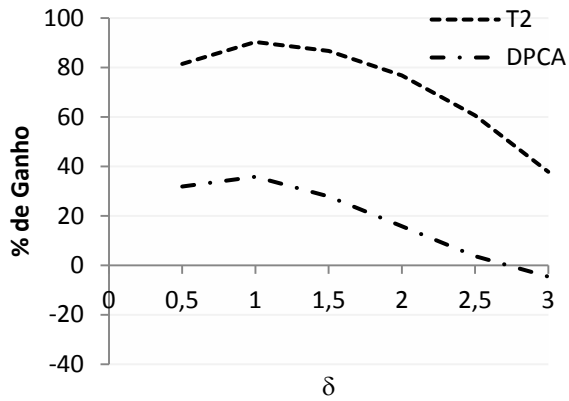


Figura 4.11 - Cenário 2: Os ganhos obtidos com a nova carta proposta *DMPCA* quando comparada com a carta  $T^2$  de Hotelling e a carta *DPCA*

- **Cenário 3**

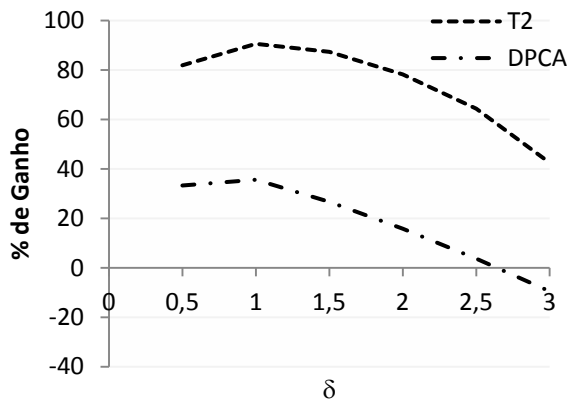


Figura 4.12 - Cenário 3: Os ganhos obtidos com a nova carta proposta *DMPCA* quando comparada com a carta  $T^2$  de Hotelling e a carta *DPCA*

- **Cenário 4**

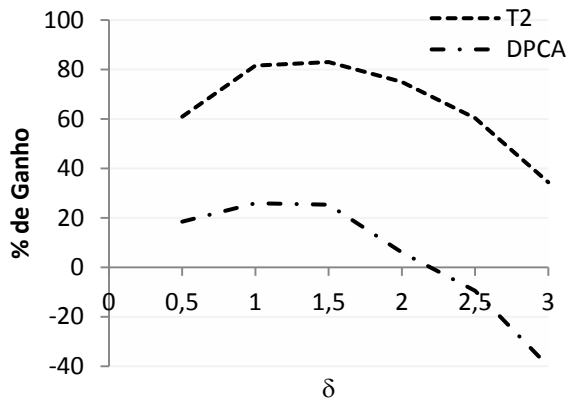


Figura 4.13 - Cenário 4: Os ganhos obtidos com a nova carta proposta *DMPCA* quando comparada com a carta  $T^2$  de Hotelling e a carta *DPCA*

- **Cenário 5**

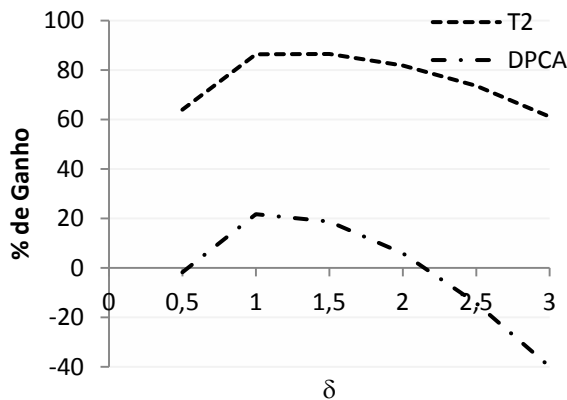


Figura 4.14 - Cenário 5: Os ganhos obtidos com a nova carta proposta *DMPCA* quando comparada com a carta  $T^2$  de Hotelling e a carta *DPCA*

É notório o enorme percentual de ganho da nova carta de controlo proposta *DMPCA* sobre a carta de controlo  $T^2$  de Hotelling para todos os níveis de alteração no tamanho da alteração na média ( $\delta$ ) para qualquer um dos cenários apresentados, como já era esperado.

No que respeita à comparação com a carta de controlo *DPCA*, a percentagem de ganhos nas medidas de desempenho assumidas para o estudo, *ARL* e *SDRL*, continua a ser superior mas mais ligeiro principalmente no que diz respeito a alterações de magnitudes pequenas e moderadas.

Considerando alterações na média de magnitude média e grande, como um grau de diminuição da autocorrelação, os ganhos nas médias de desempenho consideradas tornam-se menores para a nova carta de controlo, *DMPCA*, sendo até negativos para os últimos dois cenários, correspondendo a uma mudança de posições.

No entanto, em todos os cenários estudados, a carta de controlo *DPCA* violou o princípio da independência da estatística  $T^2$ .

#### 4.5. VANTAGENS E DESVANTAGENS DAS CARTAS MULTIVARIADAS

Depois de toda a análise realizada tanto na perspectiva de comparação como na de performance das três cartas de controlo multivariadas com dados autocorrelacionado ( $T^2$ , *DPCA* e *DMPCA*), o objetivo deste subcapítulo é identificar as vantagens e desvantagens através da análise teórica e prática feita ao longo deste estudo, mas na ótica do ambiente industrial, ou seja, na visão do processo produtivo e de aplicação na industria.

##### Vantagens da carta $T^2$

- Facilidade na aplicação;
- Possibilita simultaneamente a monitorização de duas ou mais características da qualidade relacionadas – carta de controlo multivariada;
- Não é necessário recorrer a simulação, para se obter os limites de controlo, considerando um  $ARL_{EC}$  específico;
- É possível aplicar para casos em que os dados sejam autocorrelacionados;

- É eficaz quando se pretende controlar um pequeno número de características da qualidade (variáveis);

### **Desvantagens da carta $T^2$**

- Tem como um dos pressuposto de utilização as variáveis serem independentes entre si e normalmente distribuídas, para a qual é necessário proceder à modelação dos processos ou transformação dos dados, quando estamos perante a violação deste princípio base, o que é muito frequente no ambiente industrial da era moderna;
- Não é muito eficaz na deteção de pequenas e moderadas alterações nos parâmetros do processo, porque toma sempre como base a informação presente na última amostra analisada;
- Perante um número elevado de características da qualidade (variáveis) a controlar, por norma, perde eficácia.

### **Vantagens da carta DPCA**

- Carta muito eficiente na monitorização de processos muito complexos e com grande número de variáveis para controlar;
- Possibilita simultaneamente o controlo de um grande número de variáveis;
- Apresenta as ferramentas perfeitas para identificar as variáveis mais importantes no espaço dos componentes principais;
- Muito sensível à deteção de alterações na média do processo;

- Permite a redução do número de variáveis a controlar, sem perda significativa de informação, simplificando a análise dos dados;
- Como este modelo não tem como pressuposto que as variáveis sejam normalmente distribuídas, esta carta permite eliminar a autocorrelação sem ser necessário recorrer à modelação nos casos em que os dados exibem autocorrelação.

### **Desvantagens da carta DPCA**

- É necessário a utilização de duas cartas e ambas têm de ser interpretadas em conjunto;
- Para determinar os limites de controlo para um  $ARL_{EC}$  específico, é necessário recorrer à construção de modelos de simulação;
- A construção desta carta é de execução complexa, é necessário grande conhecimento científico, o que a torna uma carta de difícil aplicabilidade para a maioria das indústrias, apesar de existir muitas soluções que permitem a análise de componentes principais;
- Viola o pressuposto da independência dos dados na estatística  $T^2$ .

### **Vantagens da carta DMPCA**

- Permite o controlo simultâneo de um grande número de variáveis;
- Carta muito eficiente na monitorização de processos muito complexos e com grande número de variáveis para controlar;
- Apresenta uma abordagem semelhante à da carta PCA dinâmica (DPCA);

- Nova estrutura matricial desfasada que permite a eliminação da autocorrelação dos valores presentes na estatística  $T^2$ ;
- Evidência desempenhos superiores às outras duas cartas para todas as alterações nas magnitudes da média (com menores valores de  $ARL$  e  $SDRL$  fora de controlo), particularmente para pequenas e médias alterações;
- Simplicidade da construção da Matriz desfasada;
- Possibilidade de utilizar as variáveis de dados originais sem necessidade de modelação;
- Processo de cálculo baseado no método de  $PCA$  no que respeita a delinear a construção e interpretação de ocorrências fora de controlo;
- A carta  $DMPCA$  apresenta grande percentagem de ganhos em relação às medidas de desempenho consideradas,  $ARL$  e  $SDRL$ , quando comparada com as outras duas cartas, carta  $T^2$  de Hotteling e carta  $DPCA$ .
- Não viola o pressuposto da independência dos dados na estatística  $T^2$ , quando a autocorrelação apresentada é moderada.

### **Desvantagens da carta $DMPCA$**

- Para determinar os limites de controlo para um  $ARL_{EC}$  específico, é necessário recorrer à construção de modelos de simulação;
- Tal como para a carta  $DPCA$ , uma vez que esta nova carta proposta é baseada na carta  $DPCA$  mas com uma matriz diferente, a construção desta carta é de execução complexa, e é necessário grande conhecimento científico, o que a torna uma carta de

difícil aplicabilidade para a maioria das indústrias, apesar de existir muitas soluções que permitem a análise de componentes principais.

## 5. CONCLUSÃO E RECOMENDAÇÕES

O principal objetivo do presente estudo passa por desenvolver uma carta de controlo multivariada (nova), designada de carta *DMPCA* (*Deployed Matrix Principal Components Analysis*) que permita corrigir uma violação dos pressupostos subjacentes às cartas de controlo, conduzida pelas cartas *DPCA* (*Dynamic Principal Component Analysis*). Para além desta primeira parte, pretende-se também avaliar o comportamento da nova carta *DMPCA* e estabelecer uma comparação do seu desempenho com as cartas  $T^2$  de Hotelling e *DPCA*, tendo por base o mesmo modelo matemático e as mesmas variáveis, de forma a identificar qual a carta que melhor permite detetar alterações no processo.

O desenvolvimento deste estudo foi feito segundo duas perspetivas, a teórica e a prática. A primeira parte, “Parte Teórica”, onde estão presentes os fundamentos teóricos necessários, que se utilizaram como base ao desenvolvimento da metodologia proposta, esta foi realizada com recurso a livros, publicações e vários artigos científicos referentes às várias temáticas ligadas à qualidade, mais focadas à construção e análise de cartas de controlo e monitorização dos processos industriais. A segunda parte, “Parte Prática” procurou, de certa forma, dar validade aos desenvolvimentos propostos seguindo como base todo o estudo feito *à priori* referente ao estudo do estado da arte da investigação das cartas de controlo multivariadas para dados autocorrelacionados, com grande enfoque nas cartas com base na análise de componentes principais.

O capítulo dedicado ao estudo das cartas de controlo teve, como objetivo, identificar e analisar a melhor carta de controlo e os métodos que devem ser utilizados na deteção de causas especiais. O primeiro estudo efetuado teve, como principal objetivo, estabelecer uma comparação entre o desempenho das três cartas de controlo, quando o processo é sujeito a um tipo de perturbação numa ou mais variáveis em simultâneo. Estas cartas foram construídas tendo por base o mesmo modelo matemático, *AR*(1) e *ARMA*(1, 1) e as mesmas variáveis. Este estudo foi realizado através de simulação, fazendo a análise e comparação de cada uma das cartas individualmente entre elas, concluiu-se numa primeira instância, que a carta baseada na análise de componentes principais dinâmicas (*DPCA*) é melhor que a carta  $T^2$  no que concerne à medida de desempenho de *ARL*, ou seja, esta carta é muito mais rápida na deteção de alterações reduzidas, moderadas ou elevadas no parâmetro da média do processo e também é a que apresenta um valor de *SDRL* mais baixo.

O segundo estudo, apresenta como principal objetivo a validação dos pressupostos das cartas de controlo multivariadas. A intenção neste ponto é provar que a nova carta proposta *DMPCA*, com a sua estrutura matricial desfasada pode eliminar a presença de autocorrelação nas estatísticas  $T^2$ , ao contrário da carta *DPCA* que induz autocorrelação nas estatísticas  $T^2$ , mesmo quando as variáveis do processo não são nem auto nem correlacionadas cruzadamente. Este estudo foi comprovado através de simulação, tal como indicado no Capítulo 3- Metodologia e mais concretamente analisado no Capítulo 4 – Análise dos Resultados, onde é notório através da Função de Autocorrelação (*FAC*) e da Função de Autocorrelação Parical (*FACP*) que a carta *DMPCA* apresenta um melhor desempenho no que concerne à eliminação da autocorrelação para a estatísticas  $T^2$  e  $Q$ , para cada um dos cenários estabelecidos onde se variou o grau de correlação e os níveis de tamanho da alteração na média,  $\delta$ , para cada um dos modelos matemáticos (*AR(1)* e *ARMA(1, 1)*).

O terceiro estudo, apresenta um estudo de extrema importância com base nos anteriormente referidos. Trata-se de uma análise de ganhos da nova carta proposta, *DMPCA*, relativamente às outras cartas de controlo multivariadas com dados autocorrelacionados, mais concretamente, carta  $T^2$  e carta *DPCA*. Os ganhos são apresentados em percentagem (%) e mostram deste modo, os ganhos da carta *DMPCA* no que diz respeito ao *ARL* e ao *SDRL* em relação às outras duas. Mais uma vez, esta nova proposta demonstra ter performances excecionais quando comparado com outras cartas.

Para além destas conclusões, também se retiraram algumas conclusões globais referentes à investigação científica realizada neste trabalho. Uma das conclusões é que os métodos de decomposição em componentes ortogonais e os gráficos de contribuição são os métodos mais apropriados quando se pretende identificar variáveis responsáveis por uma situação fora de controlo, para as cartas  $T^2$  e carta *DPCA* respetivamente. Já o método de decomposição em componentes ortogonais revela-se numa ferramenta melhor que o método da estatística  $d$  uma vez que permite verificar se a causa especial de variação é devido a alteração de uma variável e/ou a alteração da correlação entre variáveis. É também de extrema importância, a verificação da existência de autocorrelação antes de iniciar o controlo estatístico do processo, uma vez que permite evitar o uso de cartas de controlo incorretas, pois como se pode constatar, a autocorrelação tem um impacto acentuado no desempenho das cartas de controlo. Neste tipo de estudos, o desenvolvimento de metodologias utilizando fluxogramas, é muito útil para realização do que se pretende aplicar na prática. Na aplicação por simulação

destacam-se os aspetos positivos e negativos da utilização das várias cartas. Durante os vários estudos realizados, conclui-se particularmente no estudo do *ARL*, para a carta *DPCA* que este apresenta-se como uma mais-valia uma vez que na literatura encontram-se poucos estudos desta natureza. A carta *DPCA* demonstrou-se numa excelente ferramenta estatística para monitorizar a qualidade de um processo, quando se pretende detetar alterações no parâmetro médio do processo. Uma vez que a carta *DPCA* apresenta um melhor desempenho que a carta  $T^2$ , recomenda-se o uso desta carta como alternativa à carta  $T^2$ . A construção da carta *DPCA* pode ser mais complicada que a da carta  $T^2$ , no entanto, a melhoria substancial na redução do *ARL* pode e deve justificar a sua aplicação. Embora já exista no mercado *software* que permite efetuar a análise de componentes principais e, também a aplicação de cartas multivariadas baseadas em *PCA*, seria possível adaptar estes programas a uma carta *DPCA* e à nova proposta *DMPCA*.

No que diz respeito à nova carta de controlo, *DMPCA*, que segue uma abordagem semelhante à da carta *PCA* dinâmica (*DPCA*), esta propõe uma nova estrutura de matriz desfasada que permite a eliminação a autocorrelação dos valores de monitorização presentes na estatística  $T^2$ . A nova carta *DMPCA* com a sua nova matriz desfasada, evidência desempenhos superiores para todos as alterações nas magnitudes da média (com menores valores de *ARL* e *SDRL* fora de controlo). Este benefício é particularmente relevante na presença de pequenas e médias mudanças quando comparado com a tradicional  $T^2$  de Hotelling assim como com a carta *DPCA*. Estas conclusões foram retiradas de um estudo de simulação, onde foram considerados vários cenários construídos a partir de três variáveis autocorrelacionados para vários níveis de autocorrelação, de pequenos a moderados e fortes, seguindo dois diferentes modelos de séries temporais (*AR* (1) e *ARMA* (1,1)) e não ter nenhuma variável de correlação cruzada entre os cenários.

Considerando o atual estágio de desenvolvimento desta nova carta proposta, *DMPCA*, pode-se concluir que o estudo da aplicação desta carta de controlo deve ser continuado, alargado e desenvolvido, tal como as evidências da matriz recém-implementada ser ineficaz sempre que as variáveis do processo são fortemente autocorrelacionados ou interligadas.

Tomando em consideração as atuais limitações da carta *DMPCA*, há, no entanto, bastantes benefícios significativos para serem tomados em conta a partir desta nova proposta, a saber, a simplicidade de construção da matriz, a possibilidade de utilizar as variáveis de dados originais sem a necessidade de modelação e também o facto do processo de cálculo basear-se

no método de *PCA* no que respeita a traçar a construção e interpretação de ocorrências fora de controlo.

Por fim, e como foi possível verificar ao longo deste trabalho, os resultados foram extremamente satisfatórios, alcançando-se assim todos os objetivos pretendidos.

## REFERÊNCIAS

- Alwan, L. C., e Roberts, H. V. (1988). Time-Series Modeling for Statistical Process Control. *Journal of Business & Economic Statistics* , pp. 87-95.
- Aparisi, F., Champ, C. W., e Garcia-Diaz, J. C. (2004). A Performance Analysis of Hotelling X<sup>2</sup> Control Chart with Supplementary Run Rules. *Quality Eng.* , pp. 359-368.
- Bersimis, S., Psarakis, S., e Panaretos, J. (2006). Multivariate Statistical Process Control Charts: An Overview. *Quality and Reliability Engineering International* , pp. 517-543.
- Box, G. E., Jenkins, G. M., e Reinsel, G. C. (2008). Time Series Analysis: Forecasting and Control (4<sup>a</sup> ed.). *Wiley* .
- Burr, I. W. (1967). "The Effect of Non-Normality on Constants for X and R Charts". Vol. 23, pp. 563-569.
- Burr, I. W. (1976). *Statistical Quality Control Methods*. M. Dekker.
- Crosier, R. B. (1988). Multivariate Generalizations of Cumulative Sum Quality Control Schemes. *Technometrics* , pp. 291-303.
- Doty, L. A. (1996). *Statistical Process Control, 2ª Edição*. New York: Industrial Press Inc.
- Duncan, A. J. (1986). *Quality Control Industrial Statistics, 5.ª Edição*. Irwin, Homewood.
- English, J. R., Lee, C. S., Martin, T. W., e Tilmon, C. (2000). Detecting Changes in Autoregressive Processes with X-Bar and EWMA Charts. *IIE Transactions* , pp. 1103-1113.
- Ferrer, A. (2007). Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. *Quality Engineering* , pp. 311-325.
- Ghute, V. B., e Shirke, D. T. (2008). A Multivariate Synthetic Control Chart for Monitoring Process Mean Vector. *Communications in Statistics -Theory and Methods* , pp. 2136–2148.
- Harris, T. J., e Ross, W. H. (1991). Statistical Process Control Procedures for Correlated Observations. *The Canadian Journal of Chemical Engineering* , pp. 48-57.
- Houshmand, A. A., & Javaheri, A. (1998). Multivariate Ridge Residual Charts. *Quality Engeneering* , 617-624.

- Jackson, J. E. (1985). Multivariate Quality Control. *Communications in Statistics - Theory and Methods* , pp. 2657-2688.
- Javaheri, A., e Houshmand, A. (2001). Average Run Length Comparison of Multivariate Control Charts. *Journal of Statistical Computation and Simulation* , pp. 125 - 140.
- Juran, J. M., Gryna, F. M. (1993). *Quality Control Handbook, 3ª Edição*. New York: McGraw-Hill.
- Kano, M., Hasebe, S., Hashimoto, I., e Ohno, H. (2001). A New Multivariate Statistical Process Monitoring Method Using Principal Component Analysis. *Computers and Chemical Engineering* , pp. 1103-1113.
- Karim, A., Karim, S., Jafari, M.R., Mosallaei, M. (2007). Soft sensor based on dynamic principal component analysis and radial basis function neural network for distillation column. *Proceedings of the World Congress on Engineering and Computer Science* , pp. 560-564.
- Kourti, T. (2005). Application of Latent Variable Methods to Process Control and Multivariate Statistical Process Control in Industry. *International Journal of Adaptive Control and Signal Processing* .
- Kourti, T., e MacGregor, F. (1995). Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods. *Chemometrics and Intelligent Laboratory Systems* , pp. 3-21.
- Kruger U., Zhou, Y., Irwin, G.W. . (2004). Improved principal component monitoring of large scale processes. *Journal of Process Control* , pp. 879-888.
- Ku, W., Storer, R. H., e Georgakis, C. (1995). Disturbance Detection and Isolation by Dynamic. *Chemometrics and Intelligent Laboratory Systems* , pp. 179-196.
- Ledolter, J. e Burrill, C. W. (1999). *Statistical Quality Control: Strategies and Tools for Continual Improvement*. New York: John Wiley & Sons, Inc. .
- Lee, C., Choi, S.W., Lee, In-B. . (2004). Sensor fault identification based on time-lagged PCA in dynamic processes. *Chemometrics and Intelligent Laboratory System* , pp. 165-178.
- Lee, J.-M., Yoo, C., Lee, I.-B. (2004). Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science* , pp. 2995-3006.

- Lin, W. S., e Adams, B. M. (1996). Combined Control Chart for Forecast-Based Monitoring Schemes. *Journal of Quality Technology* , pp. 289-301.
- Liu, B., Makis, V. (2008). Gearbox failure diagnosis based on vector autorregressive modelling of vibration data and dynamic principal component analysis. *Journal Management Mathematics* , pp. 39-50.
- Lowry, C. A., e Montgomery, D. C. (1995). A Review of Multivariate Control Charts. *IIE Transactions* , pp. 800-810.
- Lowry, C. A., Woodall, W. H., Champ, C. W., e Rigdon, S. E. (1992). Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* , pp. 46-53.
- Lowry, C. A., Woodall, W. H., Champ, C. W., e Rigdon, S. E. (1992). Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* , pp. 46-53.
- Lu, C. W., e Reynolds, M. R. (1999a). Control Chart for Monitoring the Mean and Variance of Autocorrelated Processes. *Journal of Quality Technology* , pp. 259-274.
- Lu, C. W., e Reynolds, M. R. (2001). CUSUM Chart for Monitoring an Autocorrelated Process. *Journal fo Quality Technology* , pp. 316-334.
- Lu, C. W., e Reynolds, M. R. (1999b). EWMA Control Chart for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology* , pp. 166-188.
- Mason, R. L., Tracy, N. D., e Young, J. C. (1995). Decomposition of T<sup>2</sup> for Multivariate Control Chart Interpretation. *Journal of Quality Technology* , pp. 99-108.
- Matos, A. S. (2006). Engenharia de Controlo do Processo e Controlo Estatístico da Qualidade: Metodologia de Integração Aplicada na Indústria da Pasta de Papel. *Tese de Doutoramento. Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa* .
- Miranda, L. M. (2002). *Qualidade do Processo Produtivo da Pasta de Papel*. Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
- Montgomery, D. C. (2005). *Introduction to Statistical Quality Control (5ª ed.)*. Jonh Wiley & Sons.
- Montgomery, D. C., e Mastrangelo, C. M. (1991). Some Statistical Process Control Methods of Autocorrelated Data. *Journal of Quality Technology* , pp. 179-193.

- Montgomery, D. C., e Runger, G. C. (2006). *Applied Statistics and Probability for Engineers*, 4ª Edição. New York: John Wiley & Sons.
- Montgomery, D. C., Johnson, L. A., e Gardiner, J. S. (1990). *Forecasting & Time Series Analysis*, 2ª Edição. New York: McGraw-Hill.
- Montgomery, D. (2008a). *Design and Analysis of Experiments*, 7ª Edição. New York: John Wiley & Sons.
- Montgomery, D. (2008b). *Introduction to Statistical Quality Control*, 6ª Edição. New York: John Wiley & Sons.
- Murphy, B.J. (1987). Selecting out of Control Variables with the T<sup>2</sup> Multivariate Quality Control Procedures. *The Statistician* , pp. 571-583.
- Oakland, J. (2008). *Statistical Process Control*, 6ª Edição. Oxford: Butterworth-Heinemann.
- Ostyn, B., Darius, P., Baerdemaeker, J., e Ketelaere, B. (2007). Statistical Monitoring of a Sealing Process by Means of Multivariate Accelerometer Data. *Quality Engineering* , pp.299-310.
- Palm, A. C. (2000). "Discussion: Controversies and Contradictions in Statistical Process Control". *Journal of Quality Technology* , Vol.32, pp. 356-360.
- Pereira, Z. L., e Requeijo, J. G. (2008). *Planeamento e Controlo Estatístico de Processos*. Caparica: FCT- Fundação da Faculdade de Ciências e Tecnologia da Universidade.
- Pitt, Hy. (1994). *SPC for the Rest of Us: A Personal Path to Statistical Process Control*. Massachusetts: Addison-Wesley Publishing Company.
- Quesenberry, Charles P. (1997). *SPC Methods for Quality Improvement*. New York: John Wiley & Sons, Inc.
- Reynolds, M. R., e Lu, C. W. (1997). Control Chart for Monitoring Processes With Autocorrelated Data. *NonLinear Analysis, Theory, Methods Applications* , pp. 4059-4067.
- Runger, G. C., Alt, F. B., & Montgomery, D. C. (1996). Contributors to a Multivariate Statistical Process Control Signal. *Communications in Statistics - Theory and Methods* , pp. 2203-2213.

- Runger, G. C., e Willemain, T. R. (1995). Model-Based and Model-Free Control of Autocorrelated Processes. *Journal of Quality Technology* , pp. 283-292.
- Russell, E. L., Chiang, L. H., e Braatz, R. D. (2000). Fault Detection in Industrial Processes Using Canonical Variate Analysis and Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* , pp. 81-93.
- Shewhart, Walter A. (1931). *Economic control of quality of manufactured product*. D. Van Nostrand Company.
- Spedding, T. A. e Rawlings, P. L. (1994). Non-Normality in Statistical Process Control Measurements. *International Journal of Quality & Reliability Management* , Vol. 11(6), pp 27-37.
- Tracy, N. D., Young, J. C., e Mason, R. L. (1992). Multivariate Control Charts for Individual Observations. *Journal of Quality Technology* , pp. 88-95.
- Treasure, R.J., Kruger, U., Cooper, J.E. (2004). Dynamic multivariate statistical process control using subspace identification. *Journal of Process Control* , pp. 279-292.
- Tsung, F. (2000). Statistical Monitoring and Diagnosis of Automatic Controlled Processes Using Dynamic PCA. *International Journal of Production Research* , pp. 625-637.
- Vander Wiel, S. A. (1996a). Monitoring Processes That Wander Using Intregated Moving Average Models. *Technometrics* , pp. 139-151.
- Vasilopoulos, A. V., e Stamboulis, A. P. (1978). Modification of Control Chart Limits in the Presence of Data Correlation. *Journal of Quality Technology* , pp. 20-30.
- Wardell, D., Moskowitz, H., e Plante, R. (1994). Run-Length Distributions of Special-Cause Control Charts for Correlated Processes. *Technometrics* , pp. 3-27.
- Wheeler, D. J. (1995). *Advanced Topics in Statistical Process Control, 2ª Edição*. Knoxville, Tennessee: SPC Press.
- Xie, L., Zhang, J., Wang, S. (2006). Investigation of dynamic multivariate chemical process monitoring. *Chinese Journal of Chemical Engineering* , pp. 559-568.
- Yang, J., e Hancock, W. M. (1990). Statistical Quality Control for Correlated Samples. *International Journal of Production Process* , pp. 595-608.

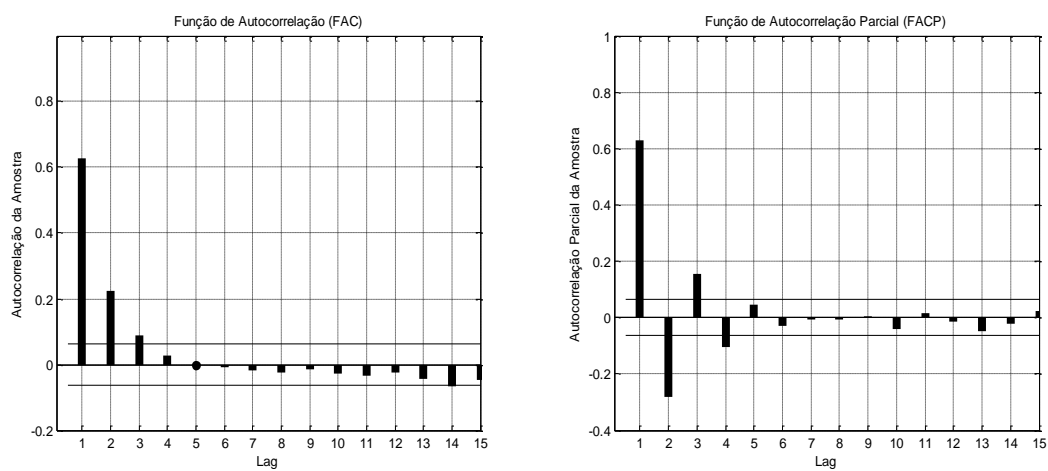
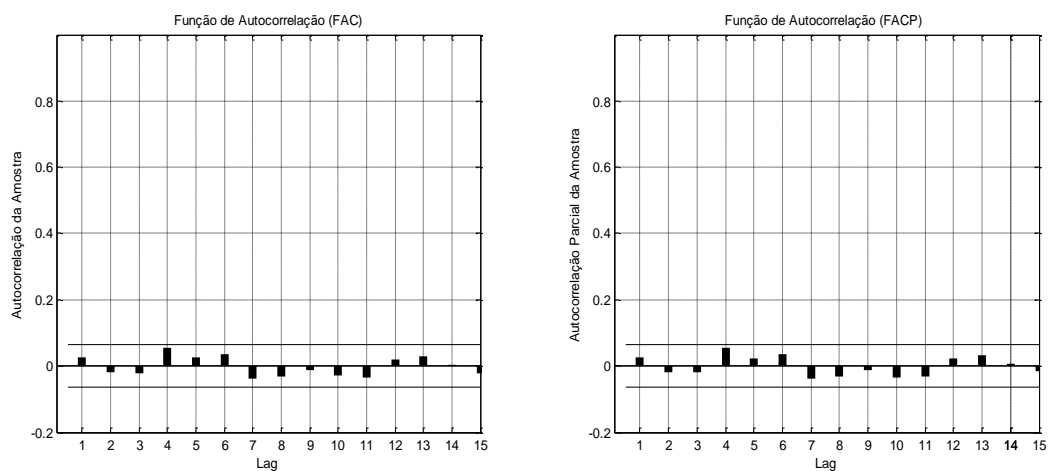
Yashchin, E. (1993). Performance of CUSUM Control Schemes for Serially Correlated Observations. *Technometrics* , pp. 37-52.

Zang, N. F. (1998). A Statistical Control Chart for Stationary Process Data. *Technometrics* , pp. 24-28.

## ANEXOS

## Anexo I – Estudo do Controlo Estatístico do Processo com Dados Autocorrelacionados

## I. Estudo da Autocorrelação da carta DPCA – Cenário 2

• Cenário 2DPCA – Estatística  $T^2$ Figura I.1 - FAC e FACP da carta DPCA para a estatística  $T^2$  (Cenário 2)DPCA – Estatística  $Q$ Figura I.2 - FAC e FACP da carta DPCA para a estatística  $Q$  (Cenário 2)

## II. Estudo da Autocorrelação da carta DMPCA – Cenário 2

### DMPCA – Estatística $T^2$

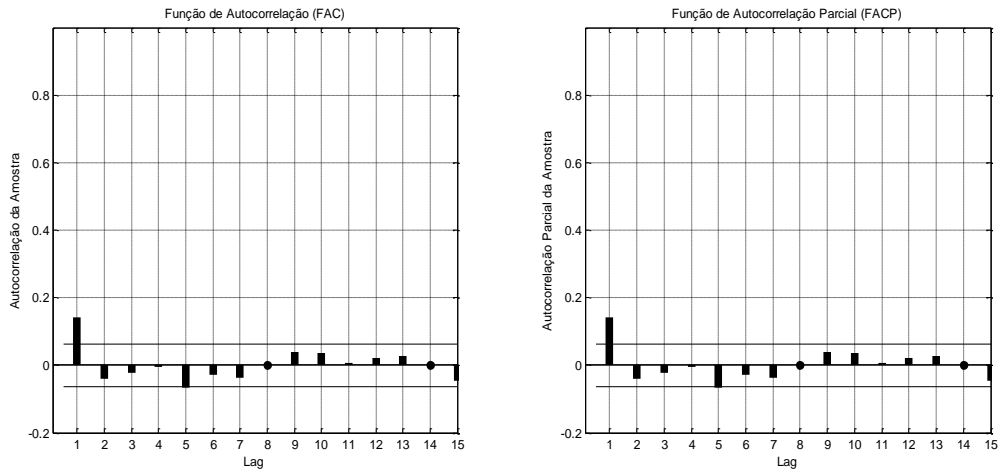


Figura II.1 - FAC e FACP da carta DMPCA para a estatística  $T^2$  (Cenário 2)

### DMPCA – Estatística $Q$

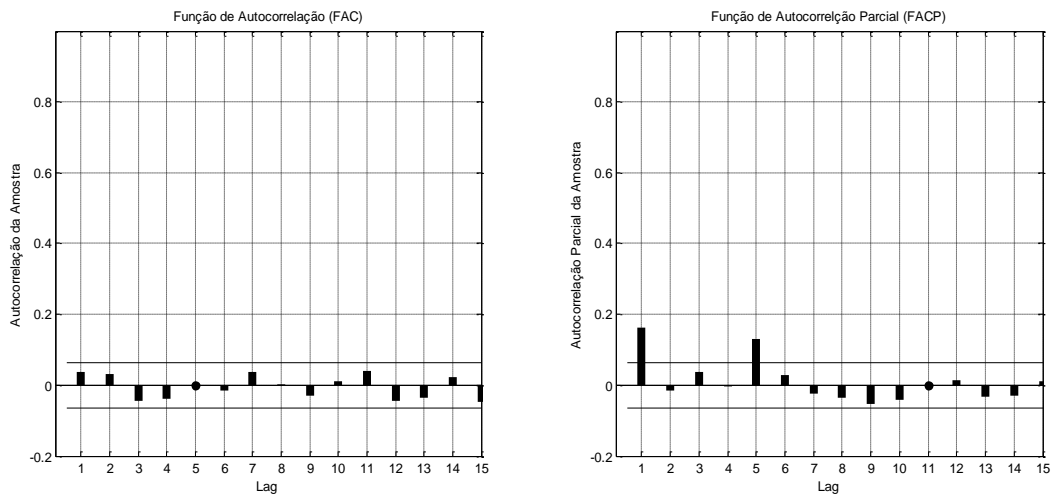


Figura II.2 - FAC e FACP da carta DMPCA para a estatística  $Q$  (Cenário 2)

### III. Estudo da Autocorrelação da carta DPCA – Cenário 3

- **Cenário 3**

DPCA - Estatística  $T^2$

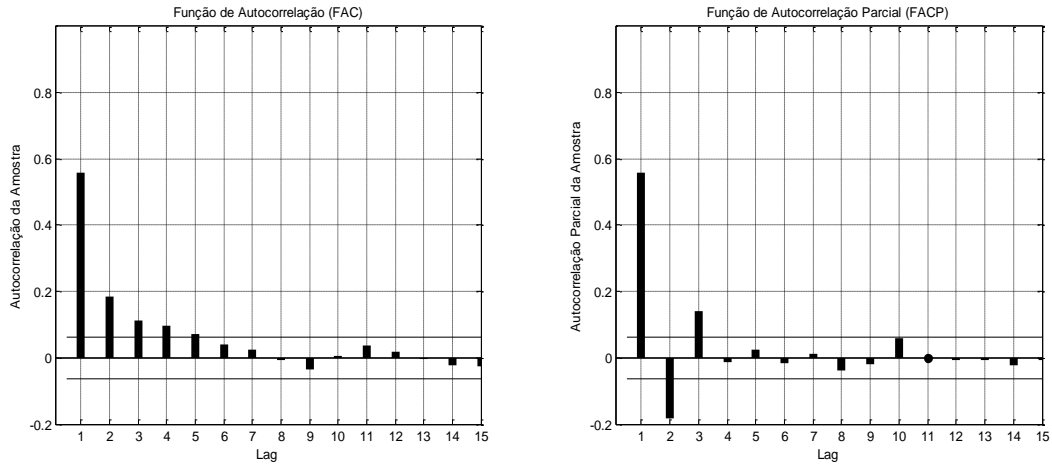


Figura III.1 - FAC e FACP da carta DPCA para a estatística  $T^2$  (Cenário 3)

DPCA – Estatística  $Q$

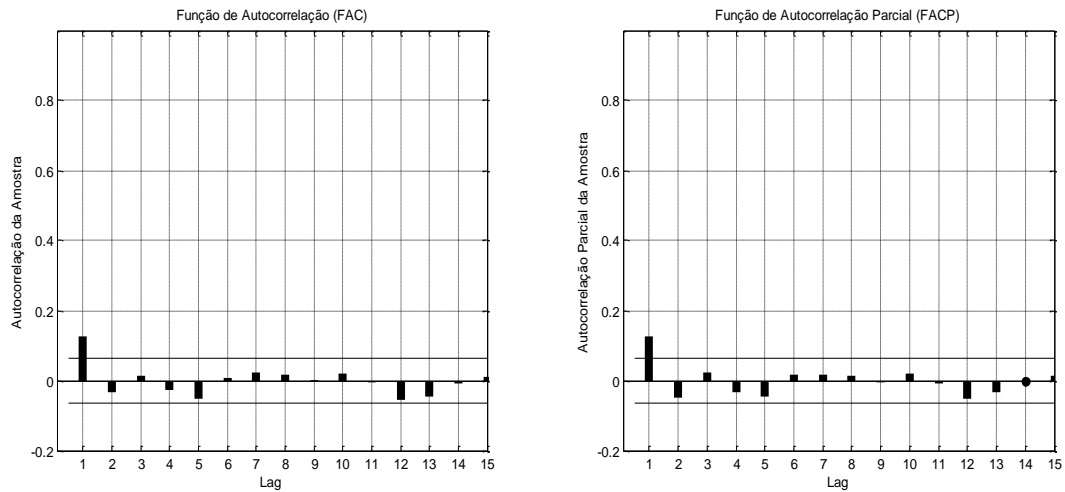


Figura III.2 - FAC e FACP da carta DPCA para a estatística  $Q$  (Cenário 3)

#### IV. Estudo da Autocorrelação da carta DMPCA – Cenário 3

##### DMPCA – Estatística $T^2$

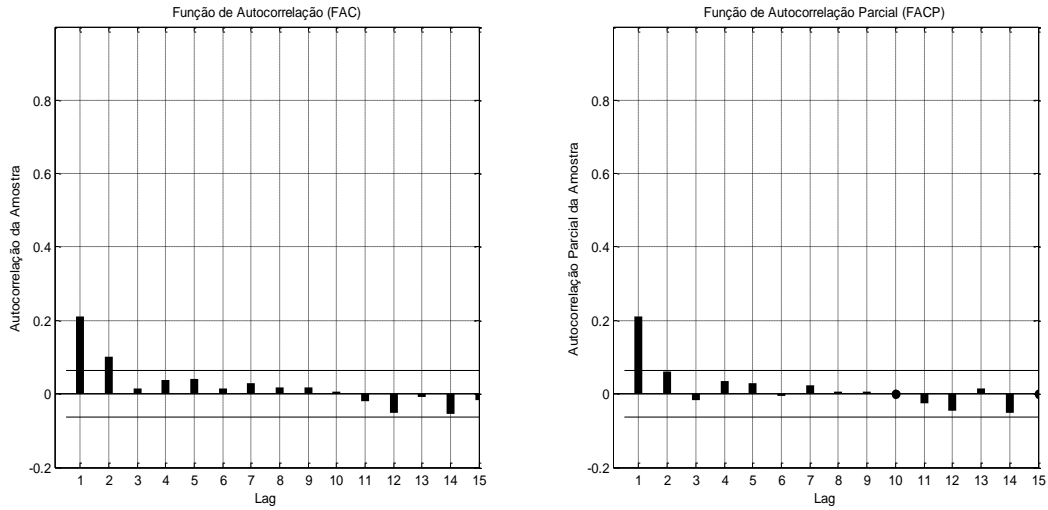


Figura IV.1 - FAC e FACP da carta DMPCA para a estatística  $T^2$  (Cenário 3)

##### DMPCA – Estatística $Q$

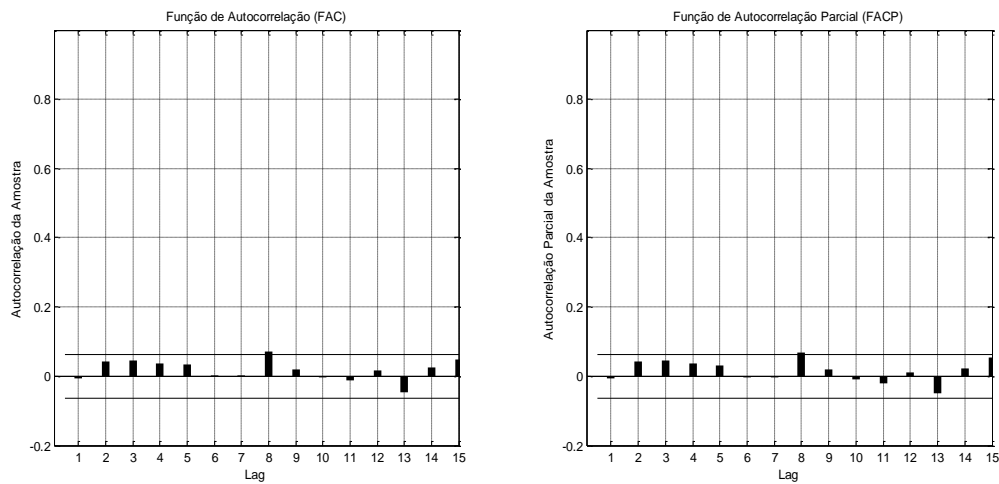


Figura IV.2 - FAC e FACP da carta DMPCA para a estatística  $Q$  (Cenário 3)

## V. Estudo da Autocorrelação da carta DPCA – Cenário 4

- Cenário 4

### DPCA – Estatística $T^2$

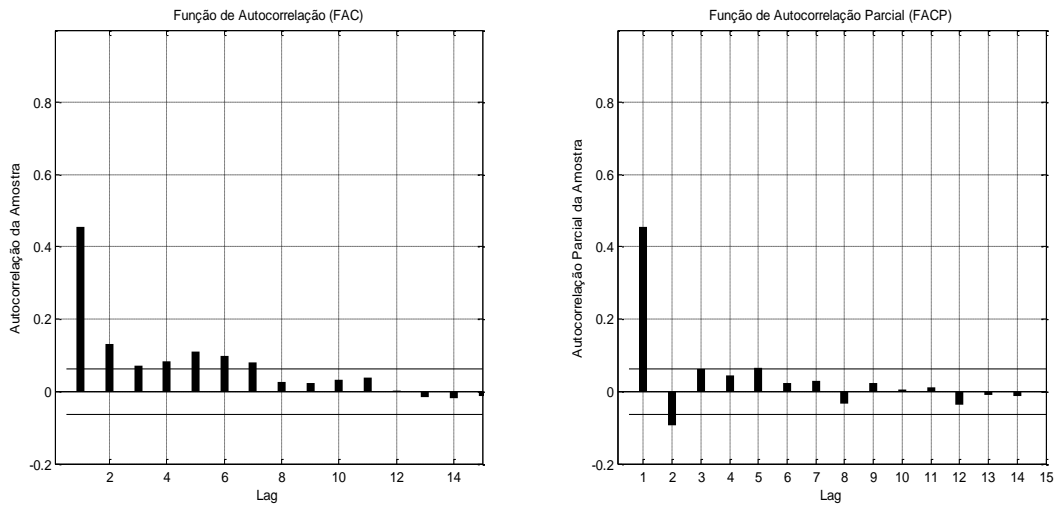


Figura V.1 - FAC e FACP da carta DPCA para a estatística  $T^2$  (Cenário 4)

### DPCA – Estatística $Q$

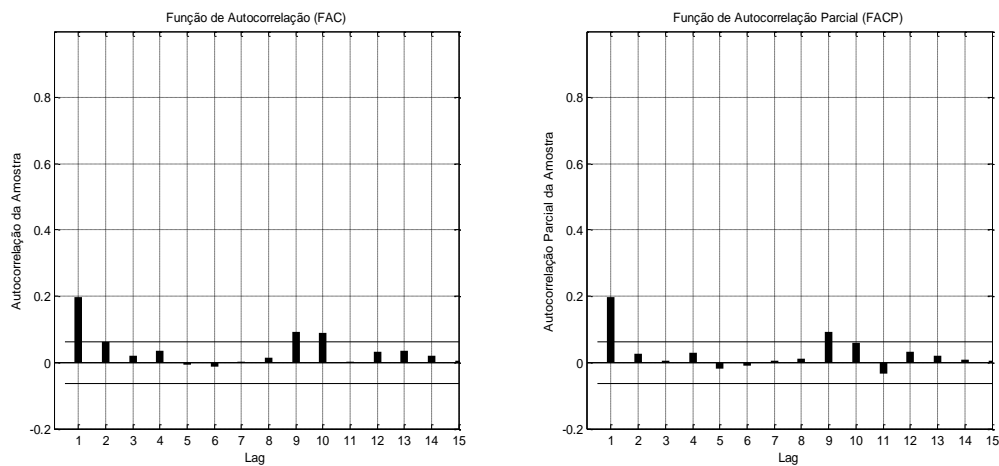


Figura V.2 - FAC e FACP da carta DPCA para a estatística  $Q$  (Cenário 4)

## VI. Estudo da Autocorrelação da carta DMPCA – Cenário 4

### DMPCA – Estatística $T^2$

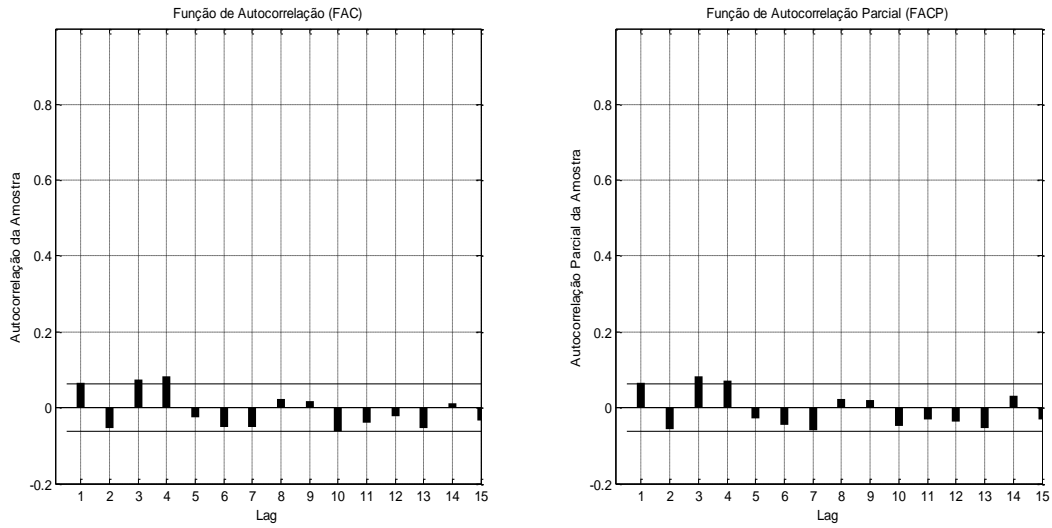


Figura VI.1 - FAC e FACP da carta DMPCA para a estatística  $T^2$  (Cenário 4)

### DMPCA – Estatística $Q$

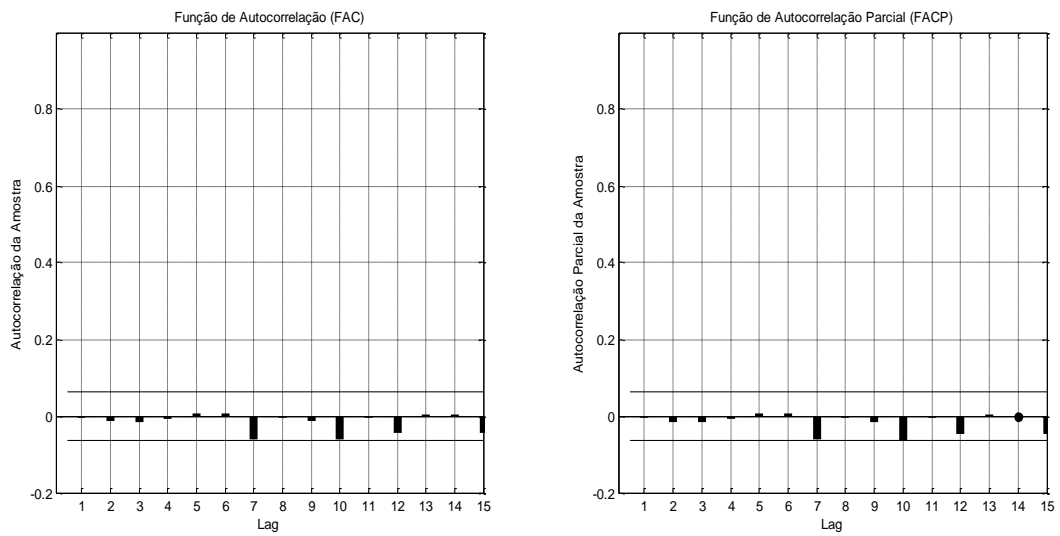


Figura VI.2 - FAC e FACP da carta DMPCA para a estatística  $Q$  (Cenário 4)

## VII. Estudo da Autocorrelação da carta DPCA – Cenário 5

- **Cenário 5**

*DPCA - Estatística  $T^2$*

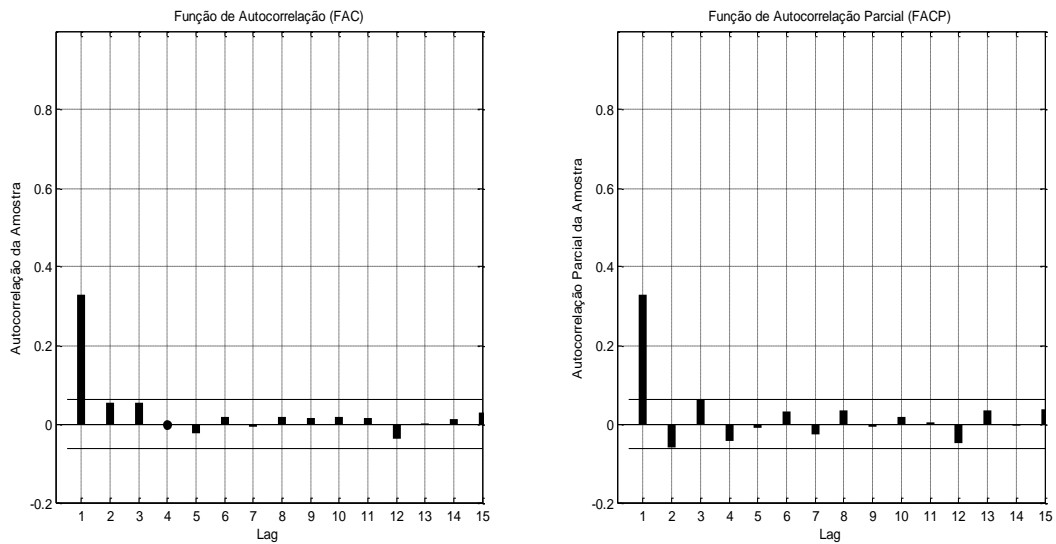


Figura VII.1 - FAC e FACP da carta DPCA para a estatística  $T^2$  (Cenário 5)

*DPCA - Estatística  $Q$*

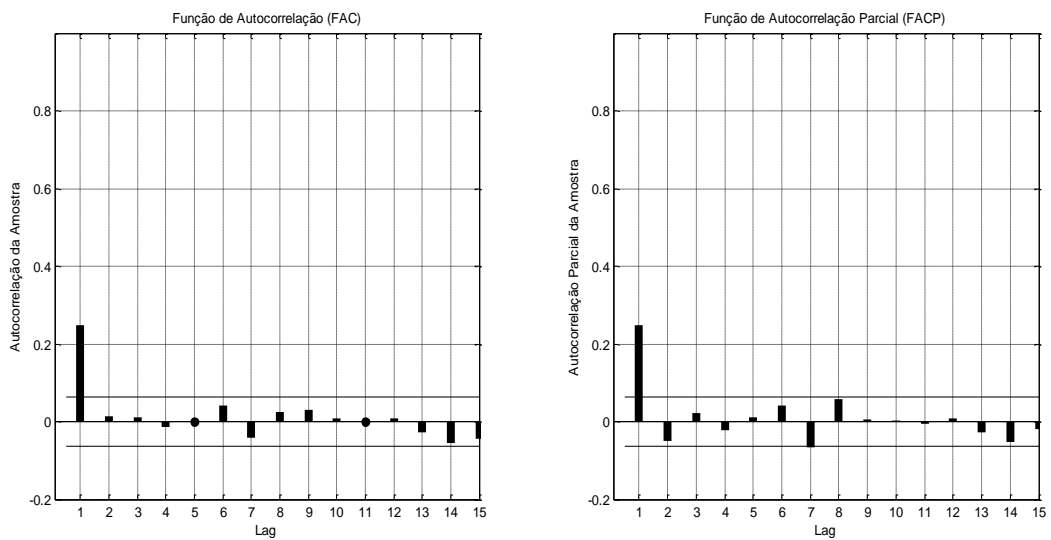


Figura VII.2 - FAC e FACP da carta DPCA para a estatística  $Q$  (Cenário 5)

## VIII. Estudo da Autocorrelação da carta DMPCA – Cenário 5

### DMPCA - Estatística $T^2$

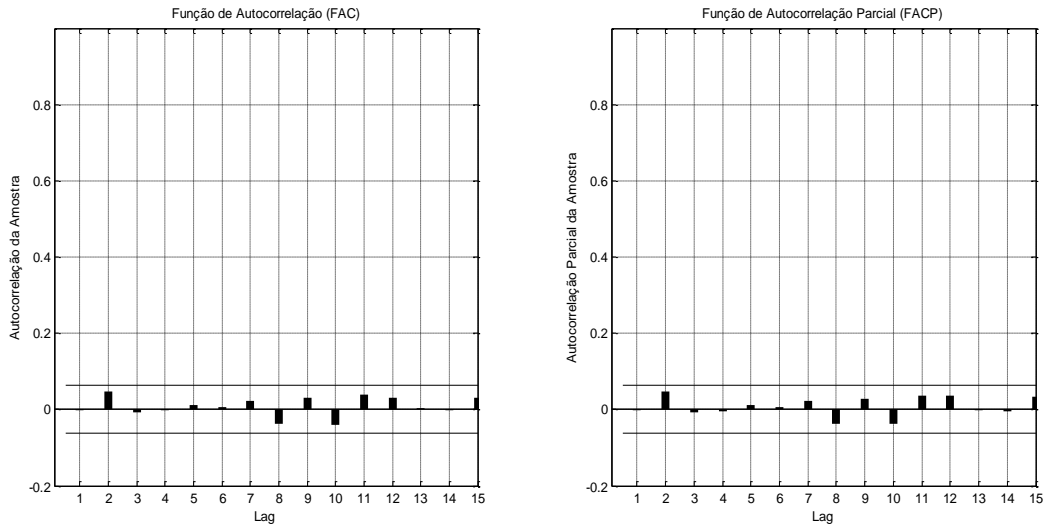


Figura VIII.1 - FAC e FACP da carta DMPCA para a estatística  $T^2$  (Cenário 5)

### DMPCA - Estatística $Q$

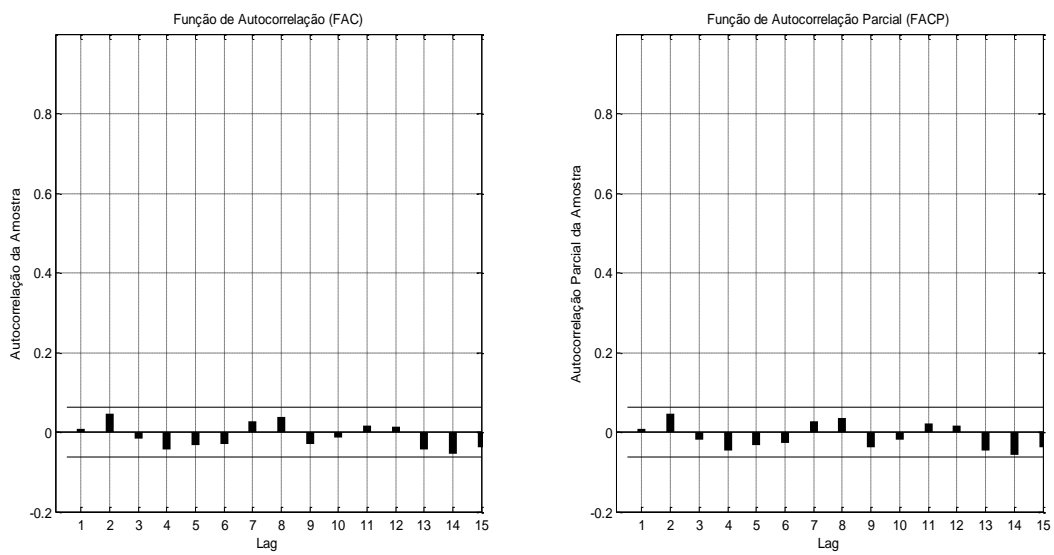


Figura VIII.2 - FAC e FACP da carta DMPCA para a estatística  $Q$  (Cenário 5)