UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Dissertação de Mestrado

MONITORAMENTO DE PROCESSOS EM BATELADAS ATRAVÉS DE CARTAS DE CONTROLE MULTIVARIADAS UTILIZANDO ANÁLISE DE COMPONENTES PRINCIPAIS MULTIDIRECIONAIS

Mestrando: Danilo Marcondes Filho

Orientador: Flávio S. Fogliatto

Porto Alegre, janeiro de 2001

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ESCOLA DE ENGENHARIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

MONITORAMENTO DE PROCESSOS EM BATELADAS ATRAVÉS DE CARTAS DE CONTROLE MULTIVARIADAS UTILIZANDO ANÁLISE DE COMPONENTES PRINCIPAIS MULTIDIRECIONAIS

Danilo Marcondes Filho

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção como requisito parcial à obtenção do título de Mestre em Engenharia de Produção

Área de concentração: Qualidade

Porto Alegre, janeiro de 2001

MONITORAMENTO DE PROCESSOS EM BATELADAS ATRAVÉS DE CARTAS DE CONTROLE MULTIVARIADAS UTILIZANDO ANÁLISE DE COMPONENTES PRINCIPAIS MULTIDIRECIONAIS

Danilo Marcondes

Esta dissertação foi julgada adequada para obtenção do título de MESTRE EM ENGENHARIA e aprovada em sua forma final pelo orientador e pela banca examinadora de curso de Pós-Graduação em Engenharia de Produção.

Orientador:_____
Flávio Sanson Fogliatto

Banca Examinadora:

- **Prof. Doutor Adelmo Luis Cechin**
- **Prof. Doutor João Riboldi**
- **Prof. Doutor Luiz Paulo Luna de Oliveira**

Coordenador:_____
Luiz Afonso dos Santos Senna

Dedico esta dissertação à minha amada companheira, Rose Lanius, pelo grande estímulo e incansável compreensão.

AGRADECIMENTOS

Ao professor Dr. Flávio Sanson Fogliatto, pela sua orientação extremamente criteriosa e incentivadora, contribuindo, assim, de forma relevante para a realização deste trabalho.

Ao professores Drs. Adelmo Luis Cechin e Luiz Paulo Luna de Oliveira, por terem proporcionado discussões frutíferas sobre o tema desenvolvido.

Ao professor Dr. Luiz Paulo Luna de Oliveira, pelos ensinamentos recebidos.

À minha querida amiga Prof. Regina Lanius, pela revisão pacienciosa do texto.

Ao meu grande amigo José Cláudio Bernardes, pela impressão da dissertação.

À Prof. Maria Adélia, pelo impagável apoio computacional.

Aos professores, funcionários e alunos do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul, pelos apoios recebidos.

À minha querida mãe, pela grande força que sempre me deu.

RESUMO

Através da utilização de sistemas de controle informatizados, indústrias buscam uma caracterização mais completa de seus processos, gerando, freqüentemente, um grande volume de dados acerca de inúmeras variáveis de desempenho dos processos. Entretanto essa grande massa de dados é subaproveitada, visto que as cartas de controle multivariadas tradicionais, tais como as cartas de Hotelling, MCUSUM e MEWMA não são suficientemente robustas para tratar com um grande número de variáveis correlacionadas, pois foram concebidas para monitorar um número reduzido de variáveis de qualidade do produto final. Diante disso, ferramentas multivariadas mais robustas, que incorporam essas informações, vêm sendo desenvolvidas; dentre estas, destaca-se a *Multiway Principal Component Analysis* (Análise de Componentes Principais Multidirecionais; ACPM).

Neste trabalho, apresenta-se a base teórica para a elaboração de uma carta de controle a partir da estratégia ACPM, para monitoramento *on-line* de processos em bateladas. Nesses processos, aferições freqüentes de variáveis de processo são disponibilizadas durante a realização de cada batelada. As cartas de controle baseadas em ACPM permitem o tratamento de variáveis autocorrelacionadas, com médias que descrevem trajetórias não lineares se observadas em momentos seqüenciados no tempo. Tais variáveis aparecem com freqüência em processos por bateladas. Uma aplicação das cartas de controle baseadas em ACPM em um sub-processo de produção de borracha para cobertura de pneus ilustra a utilização da ferramenta.

ABSTRACT

Using online control systems, industry aims at a more thorough characterization of processes; as a result, a large amount of data on several process performance variables in generated on a frequent basis. However, such wealth of data is not utilized in its full potential, as traditional Multivariate Control Charts, such as the Hotelling, MCUSUM and MEWMA Control Charts, are not appropriate to deal with a great number of correlated variables, since they were devised to monitor a reduced number of quality variables in the final product. Therefore, more thorough multivariate tools have been developed to make better use of online process data. Among these tools the *Multiway Principal Component Analysis* (MPCA) is our object of interest.

This paper presents the theoretical basis for the elaboration of a Control Chart based on the MPCA strategy for monitoring of online batch processes. In these processes, frequent checks of process variables are made available from each batch. Control Charts based on MPCA enable the treatment of correlated variables, with averages that describe non-linear trajectories when observed at sequenced periods of time; such variables are common in batch processes. One application of the Control Charts based on MPCA in a particular stage of a rubber production process illustrates the utilization of the tool.

LISTAS DE FIGURAS

CAPÍTULO 2

1 Carta de controle $\overline{X}$ 12

2 Carta de controle Qui-Quadrado para monitorar a resistência à tensão e o diâmetro da fibra têxtil 23

3 Carta de controle de Hotelling para monitorar a resistência à tensão e o diâmetro da fibra têxtil (Primeira fase do controle)..... 28

4 Mudança súbita na média..... 32

5 Mudança gradativa na média..... 32

6 Carta de controle |**S**| para monitorar a variabilidade do processo..... 36

CAPÍTULO 3

7 Diagrama de dispersão dos dados das análises químicas..... 75

8 Representação Geométrica dos Componentes Principais..... 77

9 Carta de Hotelling em relação às amostras apresentadas na Tabela 3; significância de 5% (α=5%)..... 82

10 Carta de Shewhart para monitorar os escores padronizados referentes ao primeiro Componente Principal; significância de 5% (α=5%). Exemplo proposto por Jackson (1980)..... 83

11 Carta de Shewhart para monitorar os escores padronizados referentes ao segundo Componente Principal; significância de 5% (α=5%). Exemplo proposto por Jackson (1980)..... 84

12 Carta de Shewhart para monitorar os escores padronizados referentes aos resultados obtidos através do primeiro método; significância de 5% (α=5%). Exemplo proposto por Jackson (1980)..... 85

13 Carta de Shewhart para monitorar os escores padronizados referentes aos resultados obtidos através do segundo método; significância de 5% (α=5%). . Exemplo proposto por Jackson (1980)..... 86

14 Carta de Hotelling em relação às 30 amostras preliminares; LSc obtido com significância de 5% (α=5%)..... 100

15 Projeção das 30 observações preliminares no plano definido pelos três CPs retidos..... 100

16 Carta *Q* em relação as 30 amostras preliminares; LSC obtido com significância de 5% (α=5%)..... 101

17 Carta de Hotelling em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); e nova amostra retirada do processo (amostra *26*)..... 102

18 Carta *Q* em relação as 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); e nova amostra retirada do processo (amostra *26*)..... 103

19 Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*)..... 104

20 Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*)..... 104

21 Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados

com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*)..... 105

22 Carta de Hotelling em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); nova amostra retirada do processo (amostra *27*)..... 106

23 Carta *Q* em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); nova amostra retirada do processo (amostra *26*)..... 106

24 Contribuição das 8 variáveis de processo para o escore y_1 (CP_1); amostra *26*..... 109

25 Contribuição das 8 variáveis de processo para o valor *Q*; amostra *27*..... 110

CAPÍTULO 4

26 Esquematização de uma batelada..... 112

27 Trajetórias de três variáveis de processo observadas em 20 instantes, durante 30 bateladas consecutivas. Exemplo adaptado de Nomikos & MacGregor, 1995..... 114

28 Representação dos dados num processo em bateladas, organizados na matriz $\underline{\mathbf{X}}$ tridimensional..... 115

29 Representação dos dados num processo em bateladas, organizados na matriz **X** bidimensional..... 116

30 Esquema comparativo das estratégias de monitoramento *off-line* e *on-line* via ACPM..... 125

CAPÍTULO 5

31 Trajetórias da variável X_1 observadas em 15 instantes, durante 22 bateladas preliminares..... 136

32 Trajetórias da variável X_2 observadas em 15 instantes, durante 22 bateladas preliminares..... 137

33 Carta de Hotelling em relação às 22 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 138

34 Carta *Q* em relação às 22 bateladas preliminares; LSC obtido com significância de 5% (α=5%).....139

35 Carta de Hotelling em relação às 17 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 140

36 Carta *Q* em relação às 17 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 141

37 Carta de Hotelling em relação às 16 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 141

38 Carta *Q* em relação às 16 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 142

39 Carta de Hotelling em relação às 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 142

40 Carta *Q* em relação às 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 143

41 Trajetórias padrão da variável X_1 observadas em 15 instantes, durante 15 bateladas preliminares..... 144

42 Trajetórias padrão da variável X_2 observadas em 15 instantes, durante 15 bateladas preliminares..... 144

43 Carta de Hotelling construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); nova baletada observada no processo (batelada *16*)..... 146

44 Carta *Q* construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); nova baletada observada no processo (batelada *16*)..... 146

45 Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 147

46 Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 147

47 Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 148

48 Carta de Shewhart para o quarto Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 148

49 Contribuições conjuntas das 30 variáveis de processo para os escores y_1 (CP_1) e y_4 (CP_4); batelada *16*..... 149

50 Contribuição das 30 variáveis de processo para o valor *Q*; batelada *16*..... 150

51 Carta de Hotelling construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*)..... 151

52 Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 152

53 Carta de Shewhart para o segundo Componente Principal; LSC e LIC

calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 153

54 Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 153

55 Carta de Shewhart para o quarto Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 155

56 Contribuições conjuntas das variáveis X_1 e X_2, no instante 4 da batelada *16*, para os escores parciais $y_{12,16}$ e $y_{42,16}$..... 155

57 Carta Q construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*)..... 156

58 Contribuições conjuntas das variáveis X_1 e X_2, no instante 2 da batelada *16*, para o resíduo parcial $Q_{2,16}$..... 157

LISTAS DE TABELAS

CAPÍTULO 2

1 Médias amostrais $\bar{x}_{1,k}$ e $\bar{x}_{2,k}$ com respectivos valores qui-quadrado. Exemplo adaptado de Montgomery (1996)..... 21

2 Médias, desvios padrões, covariâncias e os respectivos valores de Hotelling. Exemplo proposto por Montgomery (1996)..... 26

CAPÍTULO 3

3 Resultado das 15 amostras preliminares $\mathbf{x}'_k$, escores $c_{i,k}$ (i=1,2) obtidos através da expressão (32) e referentes escores padronizados (ver seção 3.3). As amostras *16*, *17* e *18* representam observações futuras retiradas deste processo. Exemplo adaptado proposto por Jackson (1980)..... 73

ANEXO 1

4 Glossário..... 164

ANEXO 2

5 Dados simulados, relativos à 30 amostras de 8 variáveis de processo, utilizados para o exemplo apresentado no Capítulo 3..... 165

ANEXO 3

6 Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas..... 166

7 Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas..... 167

SUMÁRIO

LISTA DE FIGURAS..... x

LISTA DE TABELAS..... xvi

RESUMO..... xiii

ABSTRACT..... xix

1. INTRODUÇÃO..... 1

1.1 JUSTIFICATIVA..... 3

1.2 OBJETIVOS..... 4

1.3 O MÉTODO DE TRABALHO..... 5

1.4 ESTRUTURA DA DISSERTAÇÃO..... 6

1.5 LIMITAÇÕES DO TRABALHO..... 7

2. REVISÃO BIBLIOGRÁFICA..... 9

2.1 INTRODUÇÃO..... 9

2.1.1 CARTAS DE CONTROLE UNIVARIADAS..... 10

2.1.2 NOTAÇÃO E DEFINIÇÕES..... 13

2.1.3 PRINCIPAIS LINHAS DE PESQUISA EM CARTAS DE CONTROLE DE PROCESSO..... 14

2.2 CARTAS DE CONTROLE MULTIVARIADAS TRADICIONAIS..... 16

2.2.1 Carta de controle Qui-Quadrado para monitoramento de médias..... 17

2.2.2 Carta de controle T^2 de Hotelling para monitorar as médias do processo..... 24

2.2.2.1 Cálculo dos limites de controle quando $n = 1$..... 30

2.2.3 Cartas de controle multivariadas para monitorar a variabilidade do processo..... 33

2.2.4 Carta de controle multivariada para a soma acumulada (MCUSUM) para monitoramento de médias de processo..... 37

2.2.5 Carta de controle multivariada para a média móvel exponencialmente ponderada (MEWMA)..... 46

2.2.6 Procedimentos de controle para dados autocorrelacionados..... 49

2.2.7 Procedimentos a posteriori para identificar variáveis fora de controle em cartas multivariadas..... 50

2.2.7.1 Métodos de diagnóstico via cartas univariadas..... 50

2.2.7.2 Métodos de diagnóstico via decomposição da estatística T^2 de Hotelling..... 52

2.3 CARTAS DE CONTROLE MULTIVARIADAS BASEADAS EM MÉTODOS DE PROJEÇÃO..... 58

2.3.1 Carta de controle baseada na projeção de dados via ACP..... 60

2.3.2 Carta de controle baseada na projeção de dados via PEL..... 64

2.3.2.1 Projeção de dados via PEL usando múltiplos blocos..... 67

2.3.3 Métodos não lineares de projeção de dados..... 67

2.4 CONSIDERAÇÕES FINAIS..... 68

3. COMPONENTES PRINCIPAIS..... 70

3.1 BASE TEÓRICA..... 70

3.2 INTERPRETAÇÃO GEOMÉTRICA DOS COMPONENTES PRINCIPAIS..... 76

3.3 CONTROLE MULTIVARIADO VIA CARTA DE HOTELLING UTILIZANDO CPs..... 79

3.4 CARTAS DE CONTROLE UTILIZANDO PROJEÇÃO DE DADOS VIA CPs..... 87

3.4.1 Padronização dos dados preliminares do processo..... 92

3.4.2 Métodos de escolha do número q de CPs necessários no modelo..... 95

3.4.2.1 A regra da linha de corte..... 96

3.4.2.2 A regra da validação cruzada..... 97

3.4.3 Exemplo..... 98

3.4.4 Diagnósticos via Gráficos de Contribuição..... 107

4. CARTAS DE CONTROLE MULTIVARIADAS VIA ANÁLISE DE COMPONENTES PRINCIPAIS MULTIDIRECIONAIS (ACPM)..... 111

4.1 PROCESSOS EM BATELADAS..... 111

4.2 ACPM..... 114

4.3 MONITORAMENTO *ON-LINE* VIA ACPM..... 122

4.3.1 Preenchimento antecipado de uma observação $\mathbf{x}'_f$ 127

4.3.2 Limites de controle para os escores parciais $c_{ql,f}$ *e para os resíduos* $Q_{l,f}$..... 129

4.3.3 Controle do erro tipo I (probabilidade de alarme falso) no monitoramento on-line.....133

5. ESTUDO DE CASO..... 135

5.1 APRESENTAÇÃO DO EXEMPLO..... 135

5.2 CONSTRUÇÃO DO MODELO ACPM DE REFERÊNCIA..... 137

5.3 MONITORAMENTO *OFF-LINE* DE UMA BATELADA FUTURA..... 145

5.3 MONITORAMENTO *ON-LINE* DE UMA BATELADA FUTURA..... 150

5.4 CONSIDERAÇÕES FINAIS..... 157

6. CONCLUSÃO..... 160

6.1 SUGESTÕES PARA PESQUISAS FUTURAS..... 162

ANEXO 1..... 164

ANEXO 2..... 165

ANEXO 3..... 166

REFERÊNCIAS BIBLIOGRÁFICAS..... 168

LISTAS DE FIGURAS

CAPÍTULO 2

1 Carta de controle $\overline{X}$ 12

2 Carta de controle Qui-Quadrado para monitorar a resistência à tensão e o diâmetro da fibra têxtil 23

3 Carta de controle de Hotelling para monitorar a resistência à tensão e o diâmetro da fibra têxtil (Primeira fase do controle)..... 28

4 Mudança súbita na média..... 32

5 Mudança gradativa na média..... 32

6 Carta de controle |**S**| para monitorar a variabilidade do processo..... 36

CAPÍTULO 3

7 Diagrama de dispersão dos dados das análises químicas..... 75

8 Representação Geométrica dos Componentes Principais..... 77

9 Carta de Hotelling em relação às amostras apresentadas na Tabela 3; significância de 5% (α=5%)..... 82

10 Carta de Shewhart para monitorar os escores padronizados referentes ao primeiro Componente Principal; significância de 5% (α=5%). Exemplo proposto por Jackson (1980)..... 83

11 Carta de Shewhart para monitorar os escores padronizados referentes ao segundo Componente Principal; significância de 5% (α=5%). Exemplo proposto por Jackson (1980)..... 84

12 Carta de Shewhart para monitorar os escores padronizados referentes aos resultados obtidos através do primeiro método; significância de 5% (α=5%). Exemplo proposto por Jackson (1980)..... 85

13 Carta de Shewhart para monitorar os escores padronizados referentes aos resultados obtidos através do segundo método; significância de 5% (α=5%). . Exemplo proposto por Jackson (1980)..... 86

14 Carta de Hotelling em relação às 30 amostras preliminares; LSc obtido com significância de 5% (α=5%)..... 100

15 Projeção das 30 observações preliminares no plano definido pelos três CPs retidos..... 100

16 Carta *Q* em relação as 30 amostras preliminares; LSC obtido com significância de 5% (α=5%)..... 101

17 Carta de Hotelling em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); e nova amostra retirada do processo (amostra *26*)..... 102

18 Carta *Q* em relação as 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); e nova amostra retirada do processo (amostra *26*)..... 103

19 Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*)..... 104

20 Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*)..... 104

21 Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados

com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*)..... 105

22 Carta de Hotelling em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); nova amostra retirada do processo (amostra *27*)..... 106

23 Carta *Q* em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); nova amostra retirada do processo (amostra *26*)..... 106

24 Contribuição das 8 variáveis de processo para o escore y_1 (CP_1); amostra *26*..... 109

25 Contribuição das 8 variáveis de processo para o valor *Q*; amostra *27*..... 110

CAPÍTULO 4

26 Esquematização de uma batelada..... 112

27 Trajetórias de três variáveis de processo observadas em 20 instantes, durante 30 bateladas consecutivas. Exemplo adaptado de Nomikos & MacGregor, 1995..... 114

28 Representação dos dados num processo em bateladas, organizados na matriz $\underline{\mathbf{X}}$ tridimensional..... 115

29 Representação dos dados num processo em bateladas, organizados na matriz **X** bidimensional..... 116

30 Esquema comparativo das estratégias de monitoramento *off-line* e *on-line* via ACPM..... 125

CAPÍTULO 5

31 Trajetórias da variável X_1 observadas em 15 instantes, durante 22 bateladas preliminares..... 136

32 Trajetórias da variável X_2 observadas em 15 instantes, durante 22 bateladas preliminares..... 137

33 Carta de Hotelling em relação às 22 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 138

34 Carta *Q* em relação às 22 bateladas preliminares; LSC obtido com significância de 5% (α=5%).....139

35 Carta de Hotelling em relação às 17 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 140

36 Carta *Q* em relação às 17 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 141

37 Carta de Hotelling em relação às 16 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 141

38 Carta *Q* em relação às 16 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 142

39 Carta de Hotelling em relação às 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 142

40 Carta *Q* em relação às 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%)..... 143

41 Trajetórias padrão da variável X_1 observadas em 15 instantes, durante 15 bateladas preliminares..... 144

42 Trajetórias padrão da variável X_2 observadas em 15 instantes, durante 15 bateladas preliminares..... 144

43 Carta de Hotelling construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); nova baletada observada no processo (batelada *16*)..... 146

44 Carta *Q* construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); nova baletada observada no processo (batelada *16*)..... 146

45 Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 147

46 Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 147

47 Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 148

48 Carta de Shewhart para o quarto Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*)..... 148

49 Contribuições conjuntas das 30 variáveis de processo para os escores y_1 (CP_1) e y_4 (CP_4); batelada *16*..... 149

50 Contribuição das 30 variáveis de processo para o valor *Q*; batelada *16*..... 150

51 Carta de Hotelling construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*)..... 151

52 Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 152

53 Carta de Shewhart para o segundo Componente Principal; LSC e LIC

calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 153

54 Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 153

55 Carta de Shewhart para o quarto Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em15 instantes da nova baletada observada no processo (batelada *16*)..... 155

56 Contribuições conjuntas das variáveis X_1 e X_2, no instante 4 da batelada *16*, para os escores parciais $y_{12,16}$ e $y_{42,16}$..... 155

57 Carta Q construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*)..... 156

58 Contribuições conjuntas das variáveis X_1 e X_2, no instante 2 da batelada *16*, para o resíduo parcial $Q_{2,16}$..... 157

LISTAS DE TABELAS

CAPÍTULO 2

1 Médias amostrais $\bar{x}_{1,k}$ e $\bar{x}_{2,k}$ com respectivos valores qui-quadrado. Exemplo adaptado de Montgomery (1996)..... 21

2 Médias, desvios padrões, covariâncias e os respectivos valores de Hotelling. Exemplo proposto por Montgomery (1996)..... 26

CAPÍTULO 3

3 Resultado das 15 amostras preliminares $\mathbf{x}'_k$, escores $c_{i,k}$ (i=1,2) obtidos através da expressão (32) e referentes escores padronizados (ver seção 3.3). As amostras *16*, *17* e *18* representam observações futuras retiradas deste processo. Exemplo adaptado proposto por Jackson (1980)..... 73

ANEXO 1

4 Glossário..... 164

ANEXO 2

5 Dados simulados, relativos à 30 amostras de 8 variáveis de processo, utilizados para o exemplo apresentado no Capítulo 3..... 165

ANEXO 3

6 Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas..... 166

7 Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas..... 167

CAPÍTULO 1

1- INTRODUÇÃO

Atualmente, existe um interesse crescente em exercer um controle mais eficaz da qualidade dos processos, isto é, busca-se ferramentas mais robustas que sejam capazes de detectar rapidamente possíveis desvios significativos em relação ao padrão de qualidade desejado para os produtos.

O controle estatístico do processo (CEP) tradicional já vem sendo utilizado há algum tempo para monitorar a qualidade dos produtos em muitas indústrias. Na maioria das aplicações, isto se dá através da utilização de cartas de controle individuais para as variáveis de qualidade do produto final. No entanto, em muitos processos, as variáveis de interesse apresentam-se fortemente correlacionadas. Neste caso, a abordagem tradicional (através das cartas univariadas) constitui-se numa ferramenta pouco eficiente para monitorar tais processos, pois não considera a estrutura de correlação das variáveis na montagem do seu padrão de referência. Como

conseqüência, as cartas univariadas podem não detectar desvios significativos em relação ao padrão desejado, gerando sinalizações distorcidas sobre o andamento do processo e conclusões errôneas sobre o mesmo.

Muitos processos industriais só podem ser eficientemente controlados através da utilização de ferramentas estatísticas multivariadas. As cartas de controle multivariadas são ferramentas de controle que incorporam a estrutura de correlação das variáveis de interesse na estratégia de monitoramento dos processos, contudo ainda são pouco usadas nas indústrias. Quando utilizadas, a aplicação das cartas de controle multivariadas restringe-se ao monitoramento de variáveis de qualidade do produto final (características de qualidade medidas após o término do processo) ao invés das variáveis de processo. Entretanto a abordagem multivariada pode ser aplicada com sucesso em processos onde existe uma série de variáveis de processo fortemente correlacionadas que precisam ser monitoradas em tempo contínuo ou por bateladas.

O foco principal deste trabalho são cartas de controle multivariadas aplicadas no controle de processos em batelada e como exemplo pode-se citar a produção de polímeros, produtos farmacêuticos e bioquímicos. Basicamente, um processo de produção por batelada segue as seguintes etapas: as matérias-primas são colocadas em um recipiente e processadas durante um determinado tempo (normalmente fixo) até que o produto seja obtido. Durante esse tempo de processamento, a trajetória de uma série de variáveis de processo, tais como temperaturas e pressões, podem ser medidas. Após o fim de cada batelada, o produto é analisado quanto às suas principais características de qualidade para verificar se está dentro do padrão de qualidade desejado.

Processos em batelada geralmente exibem variações entre as diversas bateladas de produção. Essas variações podem ser provocadas por pequenos desvios nas trajetórias das variáveis do processo em relação às suas trajetórias-padrão. Porém se uma causa indesejada estiver atuando sobre o processo durante uma batelada (levando o processo a um estado fora de controle), desvios significativos em relação à trajetória-padrão de uma ou mais variáveis do processo deverão ocorrer, podendo gerar um produto com baixa qualidade ou fora do padrão de qualidade desejado. Dessa forma, torna-se necessário construir uma carta de controle que busque, através de várias bateladas passadas bem sucedidas, uma distribuição de referência (ou padrão de

comportamento) para as trajetórias dessas variáveis. Assim, pode-se exercer um monitoramento *on-line* do processo, corrigindo em tempo real eventuais distúrbios que venham a ocorrer.

Neste trabalho, apresenta-se uma carta de controle multivariada para monitoramento *on-line* de processos em batelada, baseada em uma abordagem proposta por Nomikos & MacGregor (1994). Como habilitadora dessas cartas, apresenta-se uma ferramenta estatística adequada para monitorar processos em batelada chamada *Multiway Principal Components Analysis* (Análise de Componentes Principais Multidirecionais).

1.1 - Justificativa

No Brasil, muitas indústrias possuem processos cujos produtos são produzidos em bateladas. Nssas indústrias, o monitoramento da produção é totalmente informatizado. Através de computadores conectados *on-line* com a produção, quantidades expressivas de dados sobre as variáveis de processo, tais como temperaturas, pressões e fluxos, são coletados rotineiramente. Durante uma batelada, chega-se a coletar informações, segundo a segundo, acerca de todas as variáveis de processo. No final de cada batelada, obtém-se um histórico considerável de informações sobre o processo.

Entretanto pode-se notar, no meio industrial, uma subutilização da grande massa de informações disponíveis acerca do desempenho dos processos produtivos. A maioria das estratégias de controle de qualidade restringe-se ao monitoramento das variáveis de qualidade final do produto. Ao final de cada batelada, analisa-se o que foi produzido e, se o produto estiver dentro dos padrões de qualidade desejado, é aproveitado, caso contrário, toda a batelada de produção é descartada.

Uma batelada de produção pode gerar um produto com qualidade insatisfatória devido à má qualidade do lote de matérias-primas utilizado. Um controle de qualidade realizado sobre essas matérias primas pode, assim, minimizar o problema. Por outro lado, matérias-primas podem

ser de ótima qualidade e, mesmo assim, no final da batelada, o que foi produzido pode estar fora dos padrões estabelecidos para o produto. Isso acontece, na quase totalidade das vezes, porque, em algum momento durante o curso da batelada, alguma causa indesejada atuou sobre o processo, refletindo nas variáveis envolvidas. Assim, a trajetória de alguma variável do processo deve ter sido fortemente alterada em relação a um padrão.

Dada à grande quantidade de informações coletadas *on-line* durante o processo (no caso em foco, ao longo das bateladas), torna-se desejável, nestas indústrias, a adoção de um controle estatístico efetivo sobre as variáveis do processo, construindo uma distribuição de referência para as trajetórias dessas variáveis.

A abordagem usual de controle dos processos em batelada, sem dúvida, não é a mais indicada, pois acarreta prejuízo para as indústrias. Se uma batelada é descartada, não se sabe efetivamente o que acarretou essa não-conformidade. Não existe nenhum tipo de sinalização vinda do processo sobre a possível causa do problema. Assim, até que se ajuste o processo novamente, as bateladas seguintes, provavelmente, também produzirão produtos abaixo dos padrões de qualidade.

É neste contexto que o tema abordado nesta dissertação encontra sua justificativa.

1.2 - Objetivos

O primeiro objetivo deste trabalho é a apresentação detalhada de uma ferramenta estatística de controle que permite o monitoramento *on-line* de processos em batelada, proposta por Nomikos & MacGregor (1994), chamada Análise de Componentes Principais Multidirecionais. Para tanto, será descrita a teoria para a confecção de uma carta de controle multivariada para monitorar processos deste tipo, utilizando a abordagem proposta por esses dois autores.

O segundo objetivo deste trabalho é apresentar uma aplicação da carta de controle proposta por Nomikos & MacGregor (1994), utilizando dados obtidos de um processo de produção de borracha para cobertura de pneus. A partir da ferramenta estatística de controle apresentada, será elaborada uma carta de controle para monitorar uma operação que compõe este processo e que ocorre em batelada.

Como ultimo objetivo, apresenta-se uma revisão bibliográfica detalhada das principais estratégias de controle multivariado da qualidade disponíveis na literatura. Cartas de controle multivariadas constituem-se em desenvolvimento teórico relativamente recente, ainda pouco explorado sob diversos aspectos. Neste sentido, esta dissertação traz um texto introdutório bastante detalhado sobre cartas de controle multivariadas, suas peculiaridades, aplicações e desenvolvimento recentes. Um texto desta natureza pode suprir uma lacuna quanto à existência de referências em português.

1.3 - O método de trabalho

O método de pesquisa adotado nesta dissertação é a pesquisa exploratória, aliada a um estudo de caso, conforme proposto por Yin (1994).

Inicialmente, apresenta-se uma revisão bibliográfica sobre as abordagens usuais utilizadas na construção de cartas de controle multivariadas tradicionais.

A seguir, será fornecido todo o referencial teórico para a elaboração das cartas de controle multivariadas, utilizando a Análise de Componentes Principais (ACP).

A ACP é uma técnica estatística muito usada no tratamento de problemas onde existem informações sobre um número muito grande de variáveis correlacionadas (os processos em batelada são um bom exemplo deste tipo de problema). Essa técnica permite reduzir as variáveis analisadas para um número menor de variáveis; estas últimas são chamadas de Componentes Principais e trazem grande parte da informação contida nas variáveis iniciais.

Mostra-se que a carta de controle em relação para os Componentes Principais é uma abordagem mais informativa em relação às cartas de controle multivariadas usuais. As informações contidas nas variáveis de processo são projetadas sobre um espaço de dimensões menores, tornando, assim, as cartas de controle mais simples de serem apresentadas e mais fáceis de serem interpretadas.

Numa segunda etapa, apresenta-se em detalhes a teoria que embasa as cartas de controle multivariadas para monitoramento *on-line* de processos em batelada, utilizando a abordagem sugerida por Nomikos & MacGregor (1994).

Finalmente, apresenta-se um estudo de caso. Nesse estudo, é elaborada uma carta de controle, utilizando a abordagem proposta, para monitorar um processo em batelada de produção de borracha para cobertura de pneus.

1.4 - Estrutura da Dissertação

Capítulo 1.

Neste capítulo, expôs-se a importância da utilização de uma nova abordagem para monitorar processos em bateladas; apresentou-se também os objetivos a serem alcançados, os métodos empregados, bem como as limitações deste trabalho.

Capítulo 2.

No segundo capítulo, é feita uma revisão bibliográfica que respalda todo o trabalho e são apresentadas as principais abordagens de controle multivariado utilizadas no monitoramento da qualidade dos processos.

Capítulo 3.

No terceiro capítulo, é apresentada toda teoria necessária para o entendimento de uma técnica estatística multivariada usada para reduzir dados. Essa ferramenta estatística chama-se Análise de Componentes Principais.

Capítulo 4.

Neste capítulo, é apresentada com detalhes a abordagem proposta neste trabalho para o monitoramento *on-line* de processos em batelada. É fornecido todo o embasamento teórico para a elaboração de uma carta de controle através da Análise de Componentes Principais Multidirecionais.

Capítulo 5.

Neste capítulo, apresenta-se um estudo de caso, utilizando a abordagem proposta no capítulo 4, e elabora-se uma carta de controle para monitorar um processo de produção em bateladas de borrachas para cobertura de pneus.

Capítulo 6.

O último capítulo apresenta as considerações finais sobre o tema proposto, bem como sugestões para pesquisas futuras para extensão deste trabalho.

1.5 - Limitações do Trabalho

O corpo deste trabalho constitui-se na apresentação de uma ferramenta estatística robusta para monitorar processos em batelada. A força dessa ferramenta pode ser evidenciada através do estudo de caso apresentado, onde elabora-se, com êxito, uma carta de controle que nos permite monitorar *on-line* a qualidade de um processo de produção em bateladas.

No entanto é importante destacar duas importantes limitações existentes na abordagem proposta:

Ressalta-se, inicialmente, que a abordagem de controle proposta só pode ser implantada em processos em batelada totalmente informatizados, pois, durante a realização de cada batelada, deve-se dispor de medições *on-line* das variáveis do processo. Caso contrário, não se pode utilizar esse tipo de carta de controle, pois são exatamente essas informações do processo que serão usadas na construção de uma distribuição de referência robusta para as variáveis de processo.

Por último, é importante destacar que, através da abordagem proposta neste trabalho, será elaborada uma carta de controle que permite o monitoramento *on-line* de um processo em bateladas descrito no capítulo 5. Entretanto, para viabilizar a efetiva execução desse controle, necessita-se de *software* que incorpore a teoria proposta. São necessárias ferramentas computacionais que forneçam aos operadores, em tempo real, informações sobre o processo através da geração de cartas e gráficos de controle. O desenvolvimento dessas ferramentas computacionais transcende o escopo desta dissertação.

CAPÍTULO 2

2- REVISÃO BIBLIOGRÁFICA

2.1- Introdução

Neste capítulo, será feita uma introdução às abordagens tradicionais de controle estatístico multivariado de processos. Uma revisão das cartas de controle multivariadas mais difundidas na literatura será apresentada, além de uma breve descrição dos procedimentos de controle da qualidade de maior utilização prática. A estrutura desta revisão está baseada no trabalho de Lowry & Montgomery (1995), sendo dividida em três grandes seções, contendo (*i*) uma introdução às cartas de controle univariadas, (*ii*) uma revisão sobre cartas de controle multivariadas tradicionais e (*iii*) uma revisão sobre cartas de controle multivariadas baseadas em métodos de projeção.

2.1.1- Cartas de controle univariadas

A maior parte das indústrias tem como objetivo comum a busca da manutenção e da melhoria contínua da qualidade dos seus produtos, face a um atual mercado extremamente competitivo. Dentro desse contexto existe um interesse renovado no estudo dos métodos estatísticos para controle dos processos. O *Controle Estatístico do Processo* (CEP) constitui-se num grupo de ferramentas desenvolvidas para monitorar o desempenho de um processo, sendo as *cartas de controle* (CCs) possivelmente a ferramenta mais sofisticada; ver Montgomery (1996).

As cartas de controle foram idealizadas na década de vinte por Walter A. Shewhart, a partir do seu interesse em entender as causas que acarretam variabilidade nos processos. Esse autor classificou a variabilidade como sendo provocada por *causas comuns* (pequenas variações aleatórias inerentes ao processo), e por *causas especiais* ou *assinaláveis* (presença de eventos anômalos no processo que prejudicam a qualidade do produto). Através das CCs monitora-se a variabilidade existente nos processos, buscando detectar a possível presença de causas especiais. Causas especiais devem ser identificadas e corrigidas para que se restabeleça o padrão de desempenho esperado. As CCs também permitem a redução sistemática da variabilidade do processo, melhorando, dessa forma, a qualidade, a produtividade, a confiabilidade e o custo do que está sendo produzido.

Deve-se implementar CCs para monitorar as variáveis que influenciam a qualidade dos itens produzidos. Uma determinada variável de qualidade ou característica de qualidade (CQ) é monitorada através de sucessivas amostras espaçadas no tempo; as medições obtidas em cada amostra são representadas em ordem cronológica num gráfico e comparadas a determinados limites de controle obtidos considerando apenas causas comuns de variação presentes no processo. Se alguma medição ultrapassar os limites de controle, o processo é então considerado fora de controle estatístico. O processo também será considerado fora de controle, mesmo que nenhuma medição ultrapasse os limites, se uma seqüência de medições apresentar algum comportamento sistemático não aleatório.

As cartas de controle são divididas em *CCs para variáveis* e *CCs para atributos*. A primeira deve ser empregada quando as CQs de interesse são medidas e expressas numa escala

contínua de valores; por exemplo, temperatura, diâmetro, peso, etc. A segunda é utilizada para monitorar CQs que não são expressas numa escala contínua (número de defeitos ou não conformidades em uma unidade do produto), ou não são quantificáveis, mas sim resultado de uma classificação do tipo "conforme" ou "não conforme" (número de unidades do produto não conformes ou defeituosas).

Em aplicações práticas, as CCs para variáveis mais utilizadas são as cartas $\overline{X}$ e R (Montgomery, 1996). Através da carta $\overline{X}$ monitora-se, a partir de médias amostrais, a localização (média ou valor alvo de uma variável) do processo. A carta R é utilizada para monitorar, a partir de amplitudes amostrais, a variabilidade do processo. A localização do processo também pode ser monitorada através da *carta para a soma acumulada* (carta CUSUM; ver Lucas, 1976) e da *carta para a média móvel ponderada exponencialmente* (carta EWMA; ver Crowder, 1989). A variabilidade do processo também pode ser monitorada através da carta S, que utiliza o desvio-padrão amostral no lugar da amplitude (Montgomery, 1996).

A Fig.1 apresenta uma carta $\overline{X}$. Cada ponto representa a média $\bar{x} = \frac{1}{n}\sum_{l=1}^{n} x_l$ obtida através de amostras com n medições consecutivas ($x_1,...,x_n$) acerca de alguma CQ de interesse. O intervalo de tempo entre amostras deve ser suficiente para que se possa assumir independência entre as médias obtidas em cada amostra. Dessa forma, dado que o processo está sob controle estatístico (ou seja, apenas causas de variação inerentes ao processo estão presentes), considera-se que o comportamento da variável em questão esteja sendo bem descrito por uma distribuição gaussiana (conhecida como distribuição Normal) com média μ_0 e desvio-padrão σ_0, ambos conhecidos. Determinam-se os limites de controle para a carta $\overline{X}$ da seguinte forma:

$$\text{LSC} = \mu_0 + \text{L}\ \sigma_0 / \sqrt{n}$$

$$\text{LC} = \mu_0$$

$$\text{LIC} = \mu_0 - \text{L}\ \sigma_0 / \sqrt{n}\ ,$$

onde LSC, LC e LIC representam, respectivamente, o limite superior de controle, a linha central da carta e o limite inferior de controle, e L representa a distância máxima tolerada, em termos de

desvios-padrão da média ($\sigma_{\bar{x}} = \sigma_0 / \sqrt{n}$), entre a média $\bar{x}$ de cada amostra e a média μ_0 do processo sob controle (conhecida como distância padronizada).

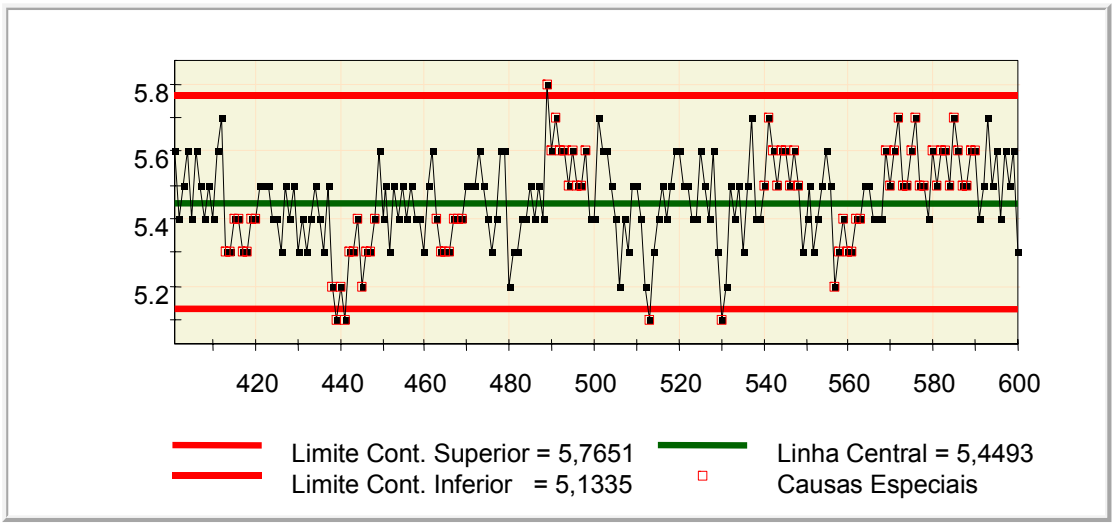

Figura 1. Carta de controle $\overline{X}$.

A escolha do valor de L deve levar em conta as probabilidades de ocorrência dos erros tipo I e tipo II. Espera-se que grande parte das médias amostrais estejam contidas dentro dos limites de controle, se estas estão sendo obtidas a partir do processo considerado sob controle. No entanto, mesmo com a presença de causas comuns apenas, existe a probabilidade de que uma determinada amostra forneça uma média além dos limites de controle, gerando, assim, um alarme falso (sinalização incorreta sobre a possível presença de uma causa especial); comete-se então o erro tipo I. Quanto maior for o valor de L, menor será a probabilidade de cometer o erro tipo I. Em contrapartida, quanto mais distantes da linha central estiverem os limites de controle, menor será a probabilidade da carta detectar uma efetiva mudança na média do processo (presença de uma causa especial de variação). Nesse caso, pode-se incorrer no erro tipo II. Quanto menor for o valor de L, menor será a probabilidade de cometer o erro tipo II. O valor L = 3 constitui-se numa escolha bastante razoável (Montgomery 1992), pois obtém-se uma carta relativamente sensível à presença de causas especiais e com uma boa proteção diante da possibilidade de alarme falso.

A carta $\overline{X}$, exemplificada na Fig.1, apresenta cinco pontos assinalados além dos limites de controle (quatro pontos abaixo de LIC e um acima de LSC), indicando a possível presença de

uma causa especial de variação no processo. A presença de um evento anômalo no processo também pode ser identificada observando-se algumas seqüências não aleatórias de pontos assinaladas. Apesar de estar dentro dos limites de controle, esses pontos apresentam um comportamento sistemático; por exemplo, vários pontos consecutivos do mesmo lado da carta.

2.1.2- Notação e definições

A seguinte notação e definições são utilizadas neste trabalho: Letras maiúsculas em negrito representam matrizes; letras maiúsculas em itálico representam variáveis aleatórias; letras maiúsculas em negrito e itálico representam vetores de variáveis aleatórias; letras minúsculas em itálico designam observações univariadas, escalares ou escores; letras minúsculas em negrito representam vetores de observações ou escores. O inverso de uma matriz **S** é representado por $\mathbf{S}^{-1}$ e sua transposta por $\mathbf{S}'$; da mesma forma, $\mathbf{s}'$ representa o transposto do vetor **s**. Uma matriz identidade de dimensão p por p é representada por $\mathbf{I}_p$. O traço de uma matriz **S** é simbolizado por $tr(\mathbf{S})$ e representa a soma dos elementos de sua diagonal principal. Uma matriz diagonal é definida como possuindo todos os elementos fora da diagonal principal iguais a zero, sendo representada por $\mathbf{S}=\text{diag}[s_{11},s_{22},...,s_{pp}]$. O determinante de uma matriz **S** é representado por $|\mathbf{S}|$. Uma matriz **S** é dita não singular se possui determinante diferente de zero; simboliza-se $|\mathbf{S}| \neq 0$. Uma matriz **S**, dita simétrica, possui $s_{ij} = s_{ji}$. Um vetor $\mathbf{u}_i$ de dimensão $p \times 1$ é dito ortonormal a um vetor $\mathbf{u}_j$, de mesma dimensão, se $\mathbf{u}_i'\mathbf{u}_j = 0$ (isto é, $\mathbf{u}_i$ ortogonal a $\mathbf{u}_j$), para $i \neq j$, e $\mathbf{u}_i'\mathbf{u}_i = 1$ ($\mathbf{u}_i$ é normalizado, isto é, possui comprimento unitário). Uma matriz $\mathbf{U}=[\mathbf{u}_1,\mathbf{u}_2,...,\mathbf{u}_p]$ de dimensão $p \times p$ é dita ortonormal se apresenta em suas colunas p vetores ortonormais.

Um grupo de dados multivariados obtidos a partir de m amostras preliminares independentes, acerca de p variáveis, serão arranjados em uma matriz **X**, de dimensão $m \times p$, onde

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \cdots \\ \mathbf{x}'_m \end{bmatrix}$$

apresenta em cada linha um vetor transposto representando uma observação $\mathbf{x}'_k$, de dimensão $1 \times p$, com $k = 1,...,m$.

Numa matriz $\mathbf{X}$ de dados ditos *centrados*, cada observação $\mathbf{x}'_k = (x_{1k},...,x_{pk})'$ é transformada em $\mathbf{x}'_{c,k} = (x_{c,1k},...,x_{c,pk})'$, com $x_{c,ik} = (x_{ik} - \bar{x}_i)$, para $i = 1,...,p$, onde $\bar{x}_i$ representa a média amostral da $i^{ésima}$ variável, estimada a partir de m amostras preliminares.

Numa matriz $\mathbf{X}$ de dados ditos *padronizados*, cada observação $\mathbf{x}'_k = (x_{1k},...,x_{pk})'$ é transformada em $\mathbf{x}'_{d,k} = (x_{d,1k},...,x_{d,pk})'$, com $x_{d,ik} = (x_{ik} - \bar{x}_i)/s_i$, para $i = 1,...,p$, onde $\bar{x}_i$ e s_i representam, respectivamente, a média e o desvio-padrão amostrais da $i^{ésima}$ variável, estimados a partir de m amostras preliminares.

2.1.3- Principais linhas de pesquisa em cartas de controle de processo

Passados mais de 70 anos após a construção da primeira carta de controle por W. A. Shewhart, pode-se identificar três tendências na pesquisa sobre cartas de controle de processo. Essas tendências são apresentadas a seguir (Palm *et al.*, 1997).

A primeira linha de pesquisa investiga a efetiva aplicação das cartas de controle, utilizando abordagens usuais, em diferentes áreas dentro de uma empresa. Esta linha de pesquisa é motivada pela constante quebra de paradigmas em busca de uma maior competitividade nas indústrias, estando inserida dentro de um contexto onde programas de melhoria da qualidade em larga escala estão sendo aplicados, orientados pela filosofia do "pensamento estatístico" de Deming. Apesar de não oferecerem contribuições teóricas significativas para o desenvolvimento

de cartas de controle, trabalhos de viés aplicado têm promovido a difusão da filosofia das cartas de controle em diversos setores da atividade econômica, buscando novas fronteiras para sua aplicação. Um exemplo disso é dado pelo crescente interesse na aplicação de cartas de controle na área da saúde (ver Palm *et al*., 1997).

A segunda linha de pesquisa investiga o desenvolvimento e aplicação de cartas de controle para processos industriais avançados, balizados por duas características essenciais: (*i*) dados *on-line* são obtidos em grande quantidade; dessa forma, as cartas de controle devem permitir o tratamento de grande massas de dados multivariados gerados pelo processo; e (*ii*) é difícil realizar uma ação corretiva rapidamente, quando o processo está fora de controle, devido à sua complexidade. Relativamente à característica (*ii*) um aspecto central diz respeito à escolha das variáveis do processo que realmente sejam relevantes para a qualidade final do produto. Assim, além do desenvolvimento de ferramentas mais robustas, que considerem de maneira eficiente a grande quantidade de informações vindas do processo, é necessário obter um grupo adequado de variáveis de processo que forneçam uma boa caracterização do desempenho do produto final. Cartas de controle avançadas demandam um excelente conhecimento sobre controle estatístico de processos e *expertise* técnica acerca dos processos industriais focalizados. As aplicações dessas cartas demandam um ambiente industrial caracterizado por processos gerenciados por sistemas informatizados de alta tecnologia. Indústrias de processos químicos e de semicondutores são exemplos de ambientes com este perfil. O controle de processos em indústrias químicas, em particular, constitui-se no principal foco de interesse desta dissertação.

Além das linhas descritas acima, uma terceira linha investiga aspectos estritamente teóricos relacionados às cartas de controle. Tal linha é desenvolvida por pesquisadores em universidades, engenheiros de produção em indústrias e departamentos dentro de escolas de administração que estudam, com viés essencialmente estatístico, as diversas abordagens para as cartas de controle, buscando continuamente alternativas para sua melhoria. Parece claro que os estudiosos desta linha de pesquisa devem dar suporte estatístico às novas cartas desenvolvidas pelos membros da segunda linha de pesquisa, oferecendo a estes alternativas para a resolução dos problemas decorrentes das novas exigências de controle advindas dos processos atuais.

O tema abordado nesta dissertação pertence à segunda linha de pesquisa apresentada acima. Neste trabalho, serão estudadas cartas de controle para monitoramento de processos

informatizados, onde uma grande quantidade de informações multivariadas do processo, oriundas de variáveis de processo mensuradas em tempo real, encontram-se disponíveis. Mais especificamente, será apresentada uma ferramenta estatística de controle, originalmente proposta por Nomikos & MacGregor (1994), que incorpora tais informações, fornecendo sinalizações mais precisas sobre o estado do processo, a cada instante. O objetivo é a obtenção de uma carta de controle mais sensível à possível presença de causas especiais de variação sobre o processo.

2.2- Cartas de controle multivariadas tradicionais

O uso de cartas de controle multivariadas para monitorar processos industriais está se tornando cada vez mais difundido (Mason *et al.*, 1995). Isso pode ser explicado pela atual tecnologia de aquisição *on-line* de dados existente. Computadores realizam, em tempo real, várias mensurações acerca das variáveis de processo e de qualidade final do produto, tornando disponíveis inúmeras informações sobre o processo. Diante dessa realidade, existe um crescente interesse em desenvolver ferramentas mais robustas, que incorporem essa grande massa de dados no monitoramento desses processos. Tais ferramentas devem permitir o monitoramento conjunto de variáveis de interesse (tanto variáveis de processo, quanto variáveis de qualidade do produto final), considerando sua estrutura de correlação.

Hotelling foi pioneiro na pesquisa sobre controle multivariado de qualidade. Este autor utilizou uma abordagem multivariada de controle em dados contendo informações sobre localizações de bombardeios, durante a Segunda Guerra Mundial. Os desenvolvimentos teóricos propostos por Hotelling, bem como desdobramentos posteriores, são apresentados nas próximas seções.

2.2.1- Carta de controle Qui-Quadrado para monitoramento de médias

Primeiramente, admite-se que exista apenas uma variável de interesse, designada por *X*, a ser monitorada em um determinado processo. Amostras periódicas independentes, compostas por *n* observações da variável ($x_1,\ldots,x_n$) são obtidas do processo. Seja $\bar{x}$ a média amostral das *n* observações. Suponha que a variável *X* siga uma distribuição Normal, com média $\mu=\mu_0$ e desvio-padrão $\sigma=\sigma_0$ conhecidos e obtidos do processo sob controle estatístico; ou seja, $X \sim N(\mu_0, \sigma_0)$. Assim, o valor

$$Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma_0} \qquad (1)$$

é uma variável que segue uma distribuição Normal com média 0 e desvio-padrão unitário; ou seja, $Z \sim N(0, 1)$. Monitora-se então as médias amostrais através da variável *Z*. O valor de *Z* representa a distância padronizada entre uma média amostral $\bar{x}$ e a média do processo μ_0. Se o processo está sob controle esperam-se apenas pequenos desvios aleatórios em relação à média do processo. Mediante tal cenário, diz-se que apenas causas comuns de variação estão presentes e que o processo está sob controle estatístico.

O valor *Z* [obtido através da expressão (1)], correspondente à média da $k^{ésima}$ amostra retirada do processo (com k=1,2,...), não deve ultrapassar certos limites, se o processo for considerado sob controle estatístico. Quando utiliza-se a variável *Z*, reescreve-se os limites de controle apresentados na seção 2.1.1 da seguinte forma:

$$\text{LSC} = \text{L} = Z_{\frac{\alpha}{2}} ,$$

$$\text{LC} = 0 ,$$

$$\text{LIC} = -\text{L} = -Z_{\frac{\alpha}{2}} ,$$

onde $Z_{\frac{\alpha}{2}}$ é o percentil da distribuição Normal, obtido a partir da probabilidade α de alarme falso adotada (erro tipo I).

Os americanos geralmente adotam α= 0,1%, gerando, assim, o intervalo de tolerância $\mu_0 \pm 3{,}3\sigma_0/\sqrt{n}$ ($|Z_{\frac{0{,}001}{2}}| = 3{,}3$), com $Z \in [-3{,}3;+3{,}3]$, ou seja, $|L| = 3{,}3$); enquanto que os europeus adotam $\mu_0 \pm 3\sigma_0/\sqrt{n}$ ($Z \in [-3,+3]$, ou seja, $|L| = 3$) como intervalo de tolerância para as médias amostrais, obtidas a partir de um processo sob controle estatístico (Montgomery, 1996). Quando $Z_{\frac{\alpha}{2}} = 3$, como $Z \sim \boldsymbol{N}(0, 1)$, a probabilidade de alarme falso é 0,27% (α=0,27%).

Sempre que a distância padronizada entre uma média amostral $\bar{x}$ e a média histórica do processo μ_0 for superior a L ou inferior a –L, existirão fortes indícios de que a média do processo sofreu uma alteração (isto é, $\mu \neq \mu_0$). Neste caso, diz-se que o processo está fora de controle estatístico, com incidência de causas especiais de variação. Este procedimento de controle, criado por Walter A. Shewhart, é conhecido na literatura como *Carta de Controle de Shewhart.*

Considere-se, agora, o quadrado da expressão (1), designado por χ^2 e dado por:

$$\chi_1^2 = \frac{(\overline{X} - \mu_0)^2 n}{\sigma_0^2}$$

ou

$$\chi_1^2 = n(\overline{X} - \mu_0)(\sigma_0^2)^{-1}(\overline{X} - \mu_0)\,. \qquad (2)$$

A variável χ_1^2 segue uma distribuição do Qui-Quadrado com um grau de liberdade e representa a distância quadrada padronizada entre a média $\bar{x}$ obtida em cada amostra e a média do processo μ_0. Assim, pode-se monitorar as médias amostrais $\overline{X}$ através do valor χ^2 [obtido através da expressão (2)]. Se L = 3, então:

LSC = 9 e

LIC = 0 .

Através da expressão (2), observa-se que o valor mínimo dessa estatística é zero (LIC=0), pois esta representa uma medida de distância quadrada. Assim, se a distância quadrada padronizada entre a média amostral $\bar{x}$ e a média do processo μ_0 ultrapassar 9 [$(\pm 3)^2$], existirão fortes indícios de que o processo esteja fora de controle estatístico ($\mu \neq \mu_0$). Suponha-se, agora, que existam *p* variáveis de interesse sendo monitoradas em um determinado processo e que estas variáveis estejam correlacionadas. Seja $\boldsymbol{X} = [X_1, X_2, ..., X_p]'$ um vetor contendo *p* variáveis aleatórias, tal que X_i, $i = 1,\ldots, p$, seja a $i^{ésima}$ variável de interesse, e $x_{i1},\ldots, x_{in}$ suas *n* observações. Para cada amostra periódica independente de *n* observações dessas *p* variáveis obtém-se o vetor de médias $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, ..., \bar{x}_p]'$. Suponha-se que as *p* variáveis sigam uma distribuição Normal *p*-variada, com vetor de médias $\boldsymbol{\mu} = \boldsymbol{\mu_0}$ e matriz de covariâncias $\boldsymbol{\Sigma} = \boldsymbol{\Sigma_0}$ conhecidos e dados por:

$$\boldsymbol{\mu_0} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \ldots\ldots \\ \mu_{0p} \end{bmatrix} \qquad \text{e} \qquad \boldsymbol{\Sigma_0} = \begin{bmatrix} \sigma^2_{0,1} & \sigma_{0,12} & \ldots\ldots & \sigma_{0,1p} \\ \sigma_{0,21} & \sigma^2_{0,2} & \ldots\ldots & \sigma_{0,2p} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \sigma_{0,p1} & \sigma_{0,p2} & \ldots\ldots & \sigma^2_{0,pp} \end{bmatrix},$$

onde μ_{0i} e $\sigma^2_{0,i}$ para $i,j = 1,\ldots,p$, correspondem, respectivamente, à média e à variância da $i^{ésima}$ variável, e $\sigma_{0,ij}$ representa a covariância entre variáveis dependentes *i* e *j*. Assim, diz-se que $\boldsymbol{X} \sim N(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$.

Finalmente, suponha-se que o vetor de médias e a matriz de covariâncias do processo caracterizem um processo em estado de controle estatístico. A expressão (2) pode ser generalizada para o caso onde *p* variáveis são amostradas simultaneamente (Lowry & Montgomery, 1995). Assim tem-se a estatística

$$\chi^2_p = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu_0})' \boldsymbol{\Sigma_0^{-1}} (\overline{\boldsymbol{X}} - \boldsymbol{\mu_0}), \qquad (3)$$

onde χ^2_p segue uma distribuição do Qui-Quadrado com *p* graus de liberdade e representa a distância quadrada padronizada, *p*-dimensional, entre um vetor de médias $\bar{\mathbf{x}}$ e o vetor de médias do processo $\boldsymbol{\mu_0}$. A raiz quadrada χ_p dessa quantidade é conhecida como *"distância de Mahalanobis*; Rencher, 1995. Para que o processo gerador dos dados seja considerado sob controle estatístico, o valor de χ^2_p não deve ultrapassar certos limites. Estes limites de controle são obtidos da seguinte forma:

$$\text{LSC} = \chi^2_{p,\alpha} \quad \text{e}$$

$$\text{LIC} = 0,$$

onde $\chi^2_{p,\alpha}$ representa o percentil da distribuição do Qui-Quadrado com *p* graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada.

Assim, estando o processo sob controle estatístico, o valor $\chi^2_{p,\alpha}$ representa a distância quadrada padronizada máxima tolerada (em *p* dimensões) entre o vetor $\bar{\mathbf{x}}$, correspondente às médias da $k^{ésima}$ amostra (com *k*=1,2,...), e $\boldsymbol{\mu_0}$. Se $\chi^2_{p,k} > \chi^2_{p,\alpha}$, há indícios de que a média de pelo menos uma das variáveis no processo tenha-se alterado, podendo, assim, caracterizar uma situação onde o processo está fora de controle estatístico ($\boldsymbol{\mu} \neq \boldsymbol{\mu_0}$, ou seja, $\mu_i \neq \mu_{0i}$ para pelo menos uma das *p* variáveis). Este procedimento de verificação caracteriza as chamadas *Cartas de Controle Qui-Quadrado* ou *de Hotelling* (Lowry & Montgomery, 1995).

Considere-se o seguinte exemplo, adaptado de Montgomery (1996). Num determinado processo existem duas variáveis sendo monitoradas conjuntamente ($p = 2$); são elas: a resistência à tensão (X_1) e o diâmetro (X_2) de uma fibra têxtil. Um engenheiro de qualidade decidiu monitorar estas duas variáveis selecionando amostras periódicas de tamanho 10 (n=10) do processo. O vetor de médias e a matriz de covariâncias do processo são supostamente conhecidos, sendo compostos pelos seguintes elementos: μ_{01}= 115,5 psi (libras por polegada quadrada), μ_{02}= 1,06 (x 10^{-2}) *in* (polegadas), σ^2_{01}=1,20 , σ^2_{02}= 0,82 e σ_{012} = 0,80. Em notação matricial, tem-se:

$$\boldsymbol{\mu_0} = \begin{bmatrix} 115,5 \\ 1,06 \end{bmatrix} \qquad \text{e} \qquad \boldsymbol{\Sigma_0} = \begin{bmatrix} 1,20 & 0,80 \\ 0,80 & 0,82 \end{bmatrix}$$

A Tab. 1 apresenta o resultado de 20 amostras periódicas consecutivas retiradas desse processo . Adotando $\alpha = 0,001$ ($\alpha = 0,1\%$), obtém-se LSC =13,816. Assim, se o processo estiver sob controle estatístico, a distância máxima tolerável, no plano bidimensional, entre o vetor de médias amostrais e o vetor das médias do processo ($\boldsymbol{\mu_0}$) é de 13,816.

Tabela 1. Médias amostrais $\bar{x}_{1,k}$ e $\bar{x}_{2,k}$ com respectivos valores Qui-Quadrado. Exemplo adaptado de Montgomery (1996).

AMOSTRA	MÉDIAS AMOSTRAIS		VALOR CALCULADO
(k)	RESISTÊNCIA À TENSÃO (X_1)	DIÂMETRO (X_2)	$\chi^2_{p,k}$
1	115,25	1,04	1,70843
2	115,91	1,06	4,007035
3	115,05	1,09	4,167733
4	116,21	1,05	11,68262
5	115,90	1,07	3,996512
6	115,55	1,06	0,059593
7	114,98	1,05	6,683953
8	115,25	1,10	0,968895
9	116,15	1,09	10,9468
10	115,92	1,05	4,006047
11	115,75	0,99	0,504942
12	114,90	1,06	8,581395
13	116,01	1,05	5,95936
14	115,83	1,07	2,745872

Tabela 1. Médias amostrais $\bar{x}_{1,k}$ e $\bar{x}_{2,k}$ com respectivos valores Qui-Quadrado. Exemplo adaptado de Montgomery (1996). (Continuação).

AMOSTRA	MÉDIAS AMOSTRAIS		VALOR CALCULADO
(k)	RESISTÊNCIA À TENSÃO (X_1)	DIÂMETRO (X_2)	$\chi^2_{p,k}$
15	115,29	1,11	0,47564
16	115,63	1,04	0,267965
17	115,47	1,03	0,031919
18	115,58	1,05	0,11186
19	115,72	1,06	1,153721
20	115,40	1,04	0,317442

Comparando em cada amostra o referente valor $\chi^2_{2,k}$ obtido [calculado através da expressão (3)] em relação à LSC, verifica-se que nenhum é superior a 13,816. Conclui-se, então, que o processo está sob controle estatístico, ou seja, existem apenas desvios aleatórios dos vetores de médias amostrais em relação ao vetor de médias do processo, provocados por causas comuns de variação. A Fig.2 mostra os escores gerados nas 20 amostras, apresentados em seqüência. As amostras 4, 9 e 12 apresentaram escores próximos ao limite superior tolerado, representado pela linha horizontal na parte superior do gráfico. No entanto, nos dois casos, observa-se que nas amostras seguintes os escores não permaneceram num mesmo patamar, apresentando valores mais baixos. Esse fato ratifica a conclusão de que não houve mudança no vetor das médias do processo, pois não existe um desvio sistemático em relação ao padrão desejado, mas sim apenas variações aleatórias (causas comuns de variação) em relação a essas médias.

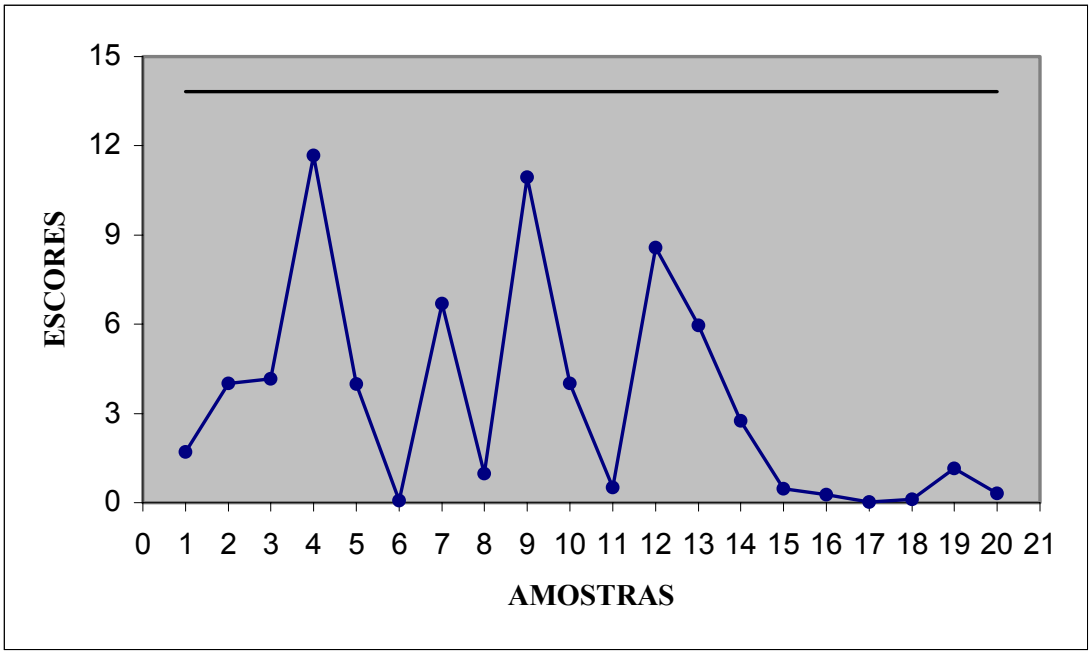

Figura 2. Carta de controle Qui-Quadrado para monitorar a resistência à tensão e o diâmetro da fibra têxtil.

Uma questão importante no projeto de cartas de controle é a definição do tamanho das amostras e a freqüência de amostragem. Esses parâmetros devem ser selecionados de forma a tornar a carta de controle mais sensível à presença de causas especiais de variação. Para tanto, definem-se, no processo, *subgrupos racionais*, selecionados de forma a maximizar as diferenças entre subgrupos e minimizar as diferenças dentro dos subgrupos, sempre que causas especiais de variação estiverem presentes. Dependendo das características do processo, existem diferentes estratégias para a obtenção dos subgrupos racionais. A determinação de subgrupos racionais num contexto univariado pode ser encontrada em Montgomery (1996) e Duncan (1986), entre outros. Homes & Mergen (1993), entre outros, apresentam abordagens multivariadas para determinação de subgrupos racionais.

2.2.2- Carta de controle T^2 de Hotelling para monitorar as médias do processo

Na seção anterior, analisou-se um processo onde o vetor de médias μ_0 e a matriz de covariâncias Σ_0 das variáveis envolvidas eram conhecidos. Na prática, μ_0 e Σ_0 são freqüentemente estimados através de amostras preliminares recolhidas do processo, quando este está sob controle estatístico. Neste caso, o processo de construção da carta de controle está dividido em duas fases distintas (Alt, 1985). A Primeira fase consiste em verificar se o processo estava sob controle estatístico (apenas causas comuns presentes) quando amostras preliminares foram obtidas. O objetivo é determinar, através dessas amostras, a distribuição de referência para as variáveis envolvidas. Em outras palavras, deseja-se obter um grupo de amostras que ofereçam uma caracterização do processo sob controle estatístico. Se essas amostras representam o padrão, pode-se, a partir delas, estimar os parâmetros do processo, bem como encontrar limites de controle apropriados. A Segunda fase é análoga à situação descrita na seção anterior, onde os parâmetros do processo eram conhecidos *a priori*. As estimativas para estes parâmetros, geradas na Primeira fase, serão usadas para monitorar, através da carta de controle, amostras futuras. A expressão (3), apresentada anteriormente, será, então, reescrita, para considerar as estimativas dos parâmetros; ou seja:

$$T^2 = n(\overline{\boldsymbol{X}} - \overline{\overline{\boldsymbol{X}}})'\overline{\mathbf{S}}^{-1}(\overline{\boldsymbol{X}} - \overline{\overline{\boldsymbol{X}}}), \qquad (4)$$

onde $\overline{\overline{\boldsymbol{X}}}$ e $\overline{\mathbf{S}}$ representam, respectivamente, as estimativas para o vetor das médias e matriz de covariâncias do processo, sendo obtidos da seguinte forma:

$$\overline{\overline{\boldsymbol{X}}} = \frac{1}{g}\sum_{k=1}^{g}\overline{\boldsymbol{X}}_k \qquad (5)$$

e

$$\overline{\mathbf{S}} = \frac{1}{g}\sum_{k=1}^{g}\mathbf{S}_k \qquad (6)$$

com

$$\mathbf{S}_k = \frac{1}{n-1}\sum_{l=1}^{n}(\boldsymbol{X}_l - \overline{\boldsymbol{X}}_k)(\boldsymbol{X}_l - \overline{\boldsymbol{X}}_k)', \qquad (7)$$

onde g representa o número de amostras preliminares de tamanho n retiradas do processo e $\boldsymbol{X}_l$ é a $l^{ésima}$ observação p-variada referente à $k^{ésima}$ amostra.

A estatística escrita na expressão (4) é utilizada como base para a *Carta de Controle* T^2 *de Hotelling* (Lowry & Montgomery, 1995). Ryan (1989) derivou a distribuição dessa estatística, em ambas as fases. Na Primeira fase, os limites de controle para a estatística T^2 são os seguintes:

$$\text{LSC} = \frac{p(g-1)(n-1)}{gn-g-p+1}F_{\alpha,p,gn-g-p+1} \quad \text{e}$$

$$\text{LIC} = 0, \qquad (8)$$

onde $F_{\alpha,p,gn\text{-}g\text{-}p+1}$ representa o percentil da distribuição F de *Snedcor* com p e (gn-g-p+1) graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada.

Na Segunda fase, os limites de controle para a estatística T^2 são os seguintes:

$$\text{LSC} = \frac{p(m+1)(n-1)}{mn-m-p+1}F_{\alpha,p,mn-m-p+1} \quad \text{e}$$

$$\text{LIC} = 0, \qquad (9)$$

onde m é o número de amostras que representam o processo sob controle, filtradas a partir das g amostras preliminares. Nesta fase se $m < g$, ou seja, $g - m$ amostras foram descartadas por não caracterizarem o padrão das variáveis no processo, recalcula-se $\overline{\overline{\boldsymbol{X}}}$ e $\overline{\mathbf{S}}$ para monitoramento das amostras futuras substituindo g por m nas expressões (5) e (6). Alt (1985) faz uma extensa discussão sobre os limites de controle em processos multivariados, em ambas as fases, apresentando resultados teóricos e exemplos numéricos complementares.

Considere-se o exemplo numérico anterior, onde se deseja monitorar a resistência à tensão (X_1) e o diâmetro (X_2) de fibras têxteis através de amostras preliminares de tamanho 10. Suponha-se, agora, que o vetor de médias e a matriz de covariâncias do processo não são

conhecidos *a priori*. Neste caso, $\boldsymbol{\mu_0}$ e $\boldsymbol{\Sigma_0}$ serão estimados a partir das 20 amostras preliminares retiradas desse processo ($g = 20$).

A última linha da Tab. 2 apresenta as estimativas para as médias, variâncias e covariância do processo, obtidas a partir das 20 amostras preliminares. Supondo amostras obtidas a partir de um processo sob controle estatístico, essas estimativas serão efetivamente usadas em substituição a $\boldsymbol{\mu_0}$ e $\boldsymbol{\Sigma_0}$ na Segunda fase. O vetor $\bar{\bar{\mathbf{x}}}$ e a matriz **S** resultantes são:

$$\bar{\bar{\mathbf{x}}} = \begin{bmatrix} 115{,}59 \\ 1{,}06 \end{bmatrix} \quad \text{e} \quad \mathbf{S} = \begin{bmatrix} 1{,}23 & 0{,}79 \\ 0{,}79 & 0{,}83 \end{bmatrix}. \tag{10}$$

Tabela 2. Médias, desvios padrões, covariâncias e os respectivos valores de Hotelling. Exemplo proposto por Montgomery (1996).

AMOSTRA	MÉDIAS AMOSTRAIS		VARIÂNCIAS E COVARIÂNCIA AMOSTRAIS			VALOR CALCULADO	DETERMI-NANTE
(k)	TENSÃO (X_1)	DIÂMETRO (X_2)	$s^2_{1,k}$	$s^2_{2,k}$	$s_{12,k}$	$T^2_{p,k}$	$\lvert S_k \rvert$
1	115,25	1,04	1,25	0,87	0,8	2,16	0,45
2	115,91	1,06	1,26	0,85	0,81	2,14	0,41
3	115,05	1,09	1,30	0,90	0,82	6,77	0,50
4	116,21	1,05	1,02	0,85	0,81	8,29	0,21
5	115,90	1,07	1,16	0,73	0,8	1,89	0,21
6	115,55	1,06	1,01	0,8	0,76	0,03	0,23
7	114,98	1,05	1,25	0,78	0,75	7,54	0,41
8	115,25	1,10	1,40	0,83	0,80	3,01	0,52
9	116,15	1,09	1,19	0,87	0,83	5,92	0,35
10	115,92	1,05	1,17	0,86	0,95	2,41	0,10

Tabela 2. Médias, desvios padrões, covariâncias e os respectivos valores de Hotelling. Exemplo proposto por Montgomery (1996). (Continuação).

AMOSTRA	MÉDIAS AMOSTRAIS		VARIÂNCIAS E COVARIÂNCIA AMOSTRAIS			VALOR CALCULADO	DETERMI-NANTE
(k)	TENSÃO (X_1)	DIÂMETRO (X_2)	$s_{1,k}^2$	$s_{2,k}^2$	$s_{12,k}$	$T_{p,k}^2$	$\lvert S_k \rvert$
11	115,75	0,99	1,45	0,79	0,78	1,13	0,54
12	114,90	1,06	1,24	0,82	0,81	9,96	0,36
13	116,01	1,05	1,26	0,55	0,72	3,86	0,17
14	115,83	1,07	1,17	0,76	0,75	1,11	0,33
15	115,29	1,11	1,23	0,89	0,82	2,56	0,42
16	115,63	1,04	1,24	0,91	0,83	0,70	0,19
17	115,47	1,03	1,20	0,95	0,7	0,19	0,65
18	115,58	1,05	1,18	0,83	0,79	0,00	0,36
19	115,72	1,06	1,31	0,89	0,76	0,35	0,59
20	115,40	1,04	1,29	0,85	0,68	0,62	0,63
MÉDIAS	115,59	1,06	1,23	0,83	0,79		

A Fig. 3 apresenta a carta de controle T^2 de Hotelling para este exemplo. A linha horizontal na parte superior representa o limite superior de controle. Adotando $\alpha = 0{,}001$ e calculando esse limite superior através da expressão (8), obtém-se LSC = 13,72.

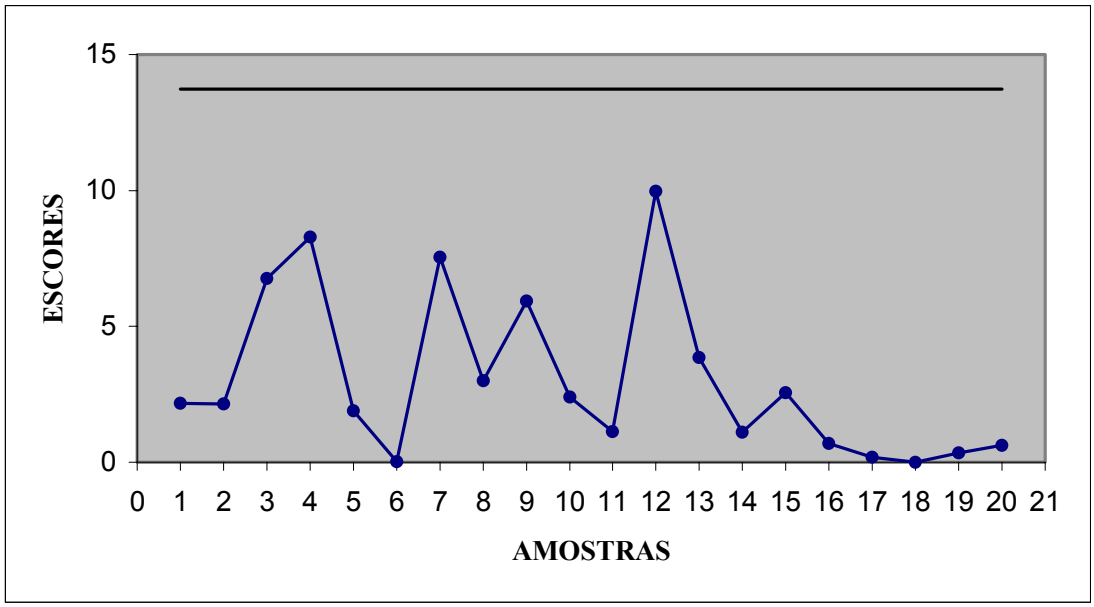

Figura 3. Carta de controle de Hotelling para monitorar a resistência à tensão e o diâmetro da fibra têxtil (Primeira fase do controle).

Nenhum dos 20 pontos plotados na carta de controle da Fig. 3 ultrapassaram LSC. Assim, admite-se que as amostras coletadas representam o padrão das variáveis quando o processo está sob controle estatístico. Como conseqüência, o vetor de médias e a matriz de covariâncias encontrados na expressão (10) serão efetivamente usados como estimativas das médias, variâncias e covariância do processo. Desta forma, as 20 amostras serão utilizadas para o cálculo dos limites de controle na Segunda fase (isto é, m = 20). Selecionando α = 0,001 e utilizando a expressão (9), obtém-se um LSC = 15,16. Se um vetor de médias $\bar{\mathbf{x}}'_f = [\bar{x}_{1,f}, \bar{x}_{2,f}]$ referente a uma futura amostra gerar, através da expressão (4), um valor superior a este limite (isto é, $T_f^2 > 15{,}16$), existirão indícios de que o vetor de médias do processo tenha se alterado, configurando um situação de processo fora de controle.

Quando o vetor de médias $\boldsymbol{\mu}_0$ e a matriz de covariâncias $\boldsymbol{\Sigma}_0$ são estimados a partir de um número grande de amostras preliminares, usa-se $\chi^2_{p,\alpha}$ (limite de controle obtido via aproximação pela distribuição do Qui-Quadrado) como limite superior de controle em ambas as fases (Montgomery, 1996). Se no exemplo acima tal aproximação houvesse sido usada no cálculo dos limites de controle, então LSC= 13,816 em ambas as fases. No exemplo, essa seria uma boa aproximação, já que o valor está próximo aos limites exatos obtidos nas duas fases. No caso

univariado ($p = 1$), para n fixo, sabe-se que com um número de amostras preliminares maior ou igual a 20 ($m \geq 20$), a distinção entre os limites de controle, na Primeira e na Segunda fases, pode ser desconsiderada, pois os limites exatos obtidos em ambas as fases resultarão muito próximos. No entanto, quando $p \geq 1$, existem restrições ao uso de tal aproximação. Lowry & Montgomery (1995) mostram que, na Segunda fase, à medida que o número de variáveis envolvidas cresce, necessita-se de um número maior de amostras preliminares para que o uso da distribuição do Qui-Quadrado se constitua em uma boa aproximação, conforme exemplificado a seguir.

Suponha-se 3 variáveis monitoradas através de amostras periódicas de 10 observações de cada variável. Após a Primeira fase, 20 amostras preliminares foram selecionadas para o controle na Segunda fase. Usando a expressão (9), com $\propto = 0{,}001$, obtém-se $LSC_1 = 15{,}1649$. Através da aproximação do Qui-Quadrado, tem-se $LSC_2 = 13{,}816$. O erro relativo é de 1,098 (LSC_1 / LSC_2). Suponha-se, agora, que 10 variáveis, com $n = 10$, estejam sendo monitoradas. Utilizando também 20 amostras preliminares obtidas na Primeira fase e $\propto = 0{,}001$, obtém-se, na Segunda fase, $LSC_1 = 34{,}833$. Através da aproximação do Qui-Quadrado, obtém-se $LSC_2 = 29{,}5884$. O erro relativo é de 1,18. Neste exemplo, é necessário um número de amostras preliminares superior a 30 para obter-se um erro relativo inferior a 1,098.

O tamanho das amostras preliminares (n) também influencia na aproximação dos limites de controle, na Segunda fase. Se, por exemplo, na Segunda fase do exemplo acima (p=10) estivessem disponíveis 20 amostras preliminares de tamanho 5, então $LSC_1 = 40{,}7491$; o limite, obtido pela aproximação do Qui-Quadrado, seria $LSC_2 = 29{,}5884$. O erro relativo, neste caso, seria 1,38 e seriam necessárias mais de 45 amostras preliminares para reduzir o erro a 1,18. De uma maneira geral, quanto menor o tamanho das amostras preliminares e/ou quanto maior o número de variáveis envolvidas, mais amostras são necessárias para manter um mesmo erro relativo.

2.2.2.1- Cálculo dos limites de controle quando n= 1

Em muitas situações práticas, o processo é monitorado utilizando amostras de tamanho unitário (n = 1). Isto pode acontecer, entre outros motivos, pelo fato de que, em alguns processos, o intervalo de tempo entre duas unidades consecutivas produzidas é muito grande. Da mesma forma, em alguns processos (geralmente processos químicos), mensurações repetidas diferem apenas devido a erros de análise; ver Tracy *et al.* (1992) e Montgomery (1996). Particularmente nos processos químicos em bateladas, foco de interesse nesta dissertação, existe a necessidade da utilização de amostras unitárias quando o objetivo é controlar as variáveis de processo. Isto se deve ao fato de que, durante uma batelada, existem diferentes padrões desejados para cada variável do processo, dependendo do instante em que a batelada está se desenvolvendo (o processo não é estacionário). Em outras palavras, se o processo estiver sob controle ou a batelada estiver conforme, as variáveis do processo seguirão suas trajetórias-padrão. Neste contexto, não faz sentido agrupar uma seqüência de observações consecutivas das variáveis, pois deseja-se verificar, em cada instante, se as variáveis estão seguindo essas trajetórias.

Suponha-se que, para um determinado processo, deseja-se construir uma carta de controle multivariada para monitorar p variáveis, a partir de m amostras preliminares, de tamanho n = 1, retiradas do processo sob controle estatístico. Neste caso, a estatística de controle T^2 de Hotelling, na expressão (4), é reescrita da seguinte forma (Lowry & Montgomery, 1995):

$$T^2=(\boldsymbol{X}-\overline{\boldsymbol{X}})'\mathbf{S}^{-1}(\boldsymbol{X}-\overline{\boldsymbol{X}}), \qquad (11)$$

onde $\overline{\boldsymbol{X}}$ e $\mathbf{S}$ representam, respectivamente, as estimativas para ao vetor das médias e a matriz de covariâncias do processo, e são calculados da seguinte forma:

$$\overline{\mathbf{X}}=\frac{1}{g}\sum_{k=1}^{g}\mathbf{X}_k$$

e

$$\mathbf{S}=\frac{1}{g-1}\sum_{k=1}^{g}(\boldsymbol{X}_k-\overline{\boldsymbol{X}})(\boldsymbol{X}_k-\overline{\boldsymbol{X}})'$$

Dada a matriz **X**, com dimensão $m \times p$, de dados *centrados*, contendo m observações p-variadas independentes coletadas do processo (ver seção 2.1.2), pode-se obter a matriz **S** através da seguinte expressão (Jackson, 1991):

$$\mathbf{S} = \frac{\mathbf{X}'\mathbf{X}}{m-1}. \qquad (12)$$

Na Primeira fase, quando selecionam-se g amostras preliminares que representem o processo sob controle, Tracy *et al.* (1992) mostraram que os limites de controle para a estatística T^2 são dados por:

$$\text{LSC} = \frac{(g-1)^2}{g} B(\alpha; p/2, (g-p-1)/2) \quad \text{e}$$

$$\text{LIC} = 0,$$

onde $B(\alpha;\ p/2, (g\text{-}p\text{-}1)/2)$ representa o percentil de uma distribuição Beta com $p/2$ e g-p-1 graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada. Valores na distribuição Beta podem ser aproximados pela distribuição F de Snedcor. Assim, pode-se obter os limites de controle na expressão (11) da seguinte forma (Lowry & Montgomery, 1995):

$$\text{LSC} = \frac{(g-1)^2}{g} \times \frac{(p/(g-p-1))F(\alpha; p, (g-p-1)}{1+(p/(g-p-1))F(\alpha; p, g-p-1)},$$

$$\text{LIC} = 0. \qquad (13)$$

Na Segunda fase, após obtidas as estimativas para os parâmetros do processo sob controle estatístico a partir de m amostras, filtradas das g amostras preliminares, os limites de controle para a estatística T^2 são:

$$\text{LSC} = \frac{p(m+1)(m-1)}{m(m-p)} F_{\alpha,p,m-p} \quad \text{e}$$

$$\text{LIC} = 0. \qquad (14)$$

Sullivan e Woodall (1996) apresentam critérios alternativos para estimar a matriz de covariâncias $\boldsymbol{\Sigma}_0$ do processo, ao utilizarem-se cartas de controle multivariadas com tamanho de

amostra unitário. Esses autores apresentam alguns procedimentos para o cálculo de **S** que tornam a carta de controle mais sensível a mudanças no vetor de médias do processo. Dois tipos de mudança no vetor de médias são abordados: (*i*) deslocamento abrupto no vetor de médias no processo (mudança súbita; ver Fig.4); e (*ii*) deslocamento gradativo (tendência) no vetor de médias do processo (mudança gradativa; ver Fig.5).

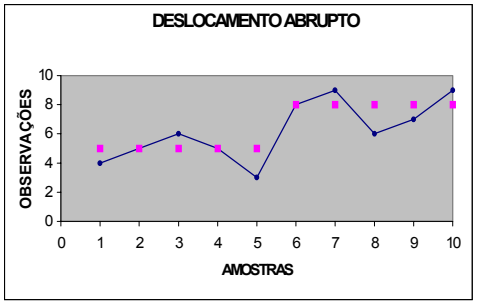

Figura 4. Mudança súbita na média.

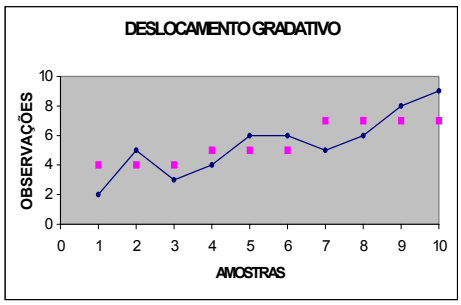

Figura 5. Mudança gradativa na média.

Na Fig.4, através de um exemplo simulado, tem-se um processo que, seguindo uma determinada média, sofre, num determinado momento, uma mudança súbita na média (as 5 primeiras amostras foram retiradas do processo com um determinado valor de média, que muda abruptamente a partir da 6ª amostra). A Fig.5, por outro lado, ilustra uma situação onde sucessivas amostras foram retiradas de um processo cuja média está sofrendo uma mudança gradativa.

Sullivan e Woodall (1996) mostram que o cálculo da matriz de covariâncias amostral, em função das diferenças entre sucessivas observações ao invés das próprias observações, produz um estimador mais robusto para Σ_0 em relação ao cálculo usual apresentado na expressão (12). Este critério foi primeiramente proposto por Holmes e Mergen (1993). A idéia básica é que observações adjacentes tendem a ser obtidas do mesmo vetor de médias do processo. Desta forma, se houver uma mudança abrupta no vetor de médias, somente um par de observações sucessivas será obtido a partir de diferentes vetores de médias. Se houver uma mudança gradativa, em contrapartida, sucessivos pares serão obtidos a partir de diferentes vetores de médias; estas diferenças, porém, serão pequenas se comparadas a duas observações aleatoriamente escolhidas. Genericamente, pode-se representar a diferença entre duas observações sucessivas *p*-variadas por:

$$\mathbf{v}'_k = \mathbf{x}'_{k+1} - \mathbf{x}'_k \ , \quad \text{para} \quad k = 1, ..., m-1,$$

Assim:

$$\mathbf{S}^* = \frac{\mathbf{V}'\mathbf{V}}{2(m-1)} \ ,$$

onde

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}'_1 \\ \mathbf{v}'_2 \\ .... \\ \mathbf{v}'_{m-1} \end{bmatrix} \quad \text{e}$$

$\mathbf{S}^*$ representa a matriz de covariâncias dessas diferenças. Usando $\mathbf{S}^*$ em lugar de **S** na estatística de controle T^2 de Hotelling na expressão (11), obtém-se uma carta de controle mais sensível a possíveis mudanças, abruptas ou gradativas, no vetor de médias do processo. Este procedimento é análogo ao que, no caso univariado, Nelson (1982) denominou por método da *amplitude móvel*.

2.2.3- Cartas de controle multivariadas para monitorar a variabilidade do processo

Quando o vetor de médias do processo está sendo monitorado, monitora-se também sua estrutura de covariância, já que, se houver uma mudança na matriz de covariâncias Σ_0, os escores obtidos através da estatística de Hotelling poderão ser afetados (Tracy *et al*., 1992). No entanto, para exercer um monitoramento mais efetivo sobre a variabilidade do processo, deve-se construir uma carta de controle específica para monitorar a sua matriz de covariâncias.

A variabilidade de um processo multivariado de ordem p é descrita através de uma matriz de covariâncias $p \times p$. A diagonal principal da matriz de covariâncias contém as variâncias das p

variáveis envolvidas; elementos fora da diagonal principal representam as covariâncias entre as variáveis na matriz.

Alt (1985) propõe dois procedimentos para o controle da variabilidade em processos multivariados. O primeiro procedimento é uma extensão da carta de controle univariada ($p = 1$) para variabilidade (*carta de controle* S^2; ver Montgomery, 1996). Este procedimento consiste em testar, a partir de sucessivas amostras, se a estrutura de covariâncias do processo está bem representada por uma determinada matriz de covariâncias $\boldsymbol{\Sigma_0}$. A estatística de controle, calculada para a $k^{ésima}$ amostra, é obtida da seguinte forma:

$$W_k = -pn + pn\ln(n) - n\ln\left(\frac{|\mathbf{A}_k|}{\boldsymbol{\Sigma_0}}\right) + tr(\boldsymbol{\Sigma_0}^{-1}\mathbf{A}_k)), \quad \text{para } k = 1,2,....,$$

onde $\mathbf{A}_k = (n\text{-}1)\,\mathbf{S}_k$, $\mathbf{S}_k$ é a matriz de covariâncias amostrais obtidas a partir da $k^{ésima}$ amostra, e *tr* representa o traço da matriz $\boldsymbol{\Sigma_0}^{-1}\mathbf{A}_k$. Os limites de controle desta estatística são os seguintes:

$$\text{LSC} = \chi^2_{\alpha,p(p+1)/2} \quad \text{e}$$

$$\text{LIC} = 0,$$

onde $\chi^2_{\alpha,p(p+1)/2}$ representa o percentil da distribuição do Qui-Quadrado com $p(p+1)/2$ graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada. Se o valor de W_k for superior à LSC, existirão fortes indícios de que a variabilidade do processo tenha se alterado, caracterizando assim uma situação de processo fora de controle ($\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma_0}$).

O segundo procedimento é baseado numa medida denominada *variância generalizada*. A *variância generalizada* de um processo, estimada através da $k^{ésima}$ amostra, é representada por $|\mathbf{S}_k|$, ou seja, pelo determinante da matriz de covariâncias amostrais. Montgomery e Wadsworth, *apud* Lowry & Montgomery (1995) sugerem uma aproximação assintótica, através da distribuição Normal, para obter uma carta de controle para $|\mathbf{S}|$.

Existe ainda um terceiro procedimento que consiste em obter uma carta de controle utilizando a média e da variância de $|\mathbf{S}|$, usando a idéia de que grande parte da distribuição de probabilidade desta estatística está contida no intervalo $E(|\mathbf{S}|) \pm 3\sqrt{V(|\mathbf{S}|)}$, onde $E(|\mathbf{S}|)$ e

$V(|\mathbf{S}|)$ representam, respectivamente, a média e a variância de |**S**|; Montgomery (1996). Escreve-se:

$$\mathrm{E}(|\mathbf{S}|) = b_1 |\boldsymbol{\Sigma}_0| \quad \text{e} \quad V(|\mathbf{S}|) = b_2 |\boldsymbol{\Sigma}_0|^2 ,$$

onde

$$b_1 = \frac{1}{(n-1)^p} \prod_{i=1}^{p} (n-i) \qquad \text{e}$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{i=1}^{p} (n-1) \Big[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j) \Big] .$$

Os limites de controle são dados por:

$$\mathrm{LSC} = |\boldsymbol{\Sigma}_0| (b_1 + 3 b_2^{1/2}) ,$$

$$\mathrm{LC} = b_1 |\boldsymbol{\Sigma}_0| ,$$

$$\mathrm{LIC} = |\boldsymbol{\Sigma}_0| (b_1 - 3 b_2^{1/2}) ,$$

onde o limite inferior de controle (LIC) deve ser substituído por zero se o valor calculado for negativo.

Na prática, necessita-se estimar a matriz de covariâncias do processo ($\boldsymbol{\Sigma}_0$). Através da análise de *m* amostras preliminares (Primeira fase), obtém-se a matriz de covariâncias amostrais **S**. Neste caso, deve-se substituir $|\boldsymbol{\Sigma}_0|$ por $|\mathbf{S}|/b_1$, pois este é o estimador não tendencioso de $|\boldsymbol{\Sigma}_0|$. Assim, os limites de controle apresentados acima são reescritos da seguinte forma:

$$\mathrm{LSC} = \frac{|\mathbf{S}|}{b_1} (b_1 + 3 b_2^{1/2}) ,$$

$$\mathrm{LC} = |\mathbf{S}| ,$$

$$\mathrm{LIC} = \frac{|\mathbf{S}|}{b_1} (b_1 - 3 b_2^{1/2}) .$$

Considere-se o exemplo numérico apresentado na seção 2.3.1. Na última coluna da Tab. 2, são mostrados os valores de $|\mathbf{S}_k|$ obtidos em cada uma das 20 (m=20) amostras preliminares. Obtém-se através da expressões (6) e (7) a matriz de covariâncias amostrais:

$$\mathbf{S} = \begin{bmatrix} 1{,}23 & 0{,}79 \\ 0{,}79 & 0{,}83 \end{bmatrix}, \qquad \text{com} \qquad |\mathbf{S}| = 0{,}3968 \ .$$

A Fig.6 apresenta a carta de controle $|\mathbf{S}|$, para monitorar a variabilidade do processo no exemplo numérico acima. As linhas horizontais vermelhas representam, respectivamente, o limite superior de controle (LSC = 1,26), a linha de centro (LC = 0,3968) e o limite inferior de controle (LIC = 0). Há evidências de que as 20 amostras preliminares representam a estrutura padrão de covariâncias do processo, ou seja, a variabilidade do processo sob controle estatístico. Deve-se monitorar continuamente a variabilidade desse processo através de futuras amostras, plotando na carta os respectivos valores $|\mathbf{S}_f|$.

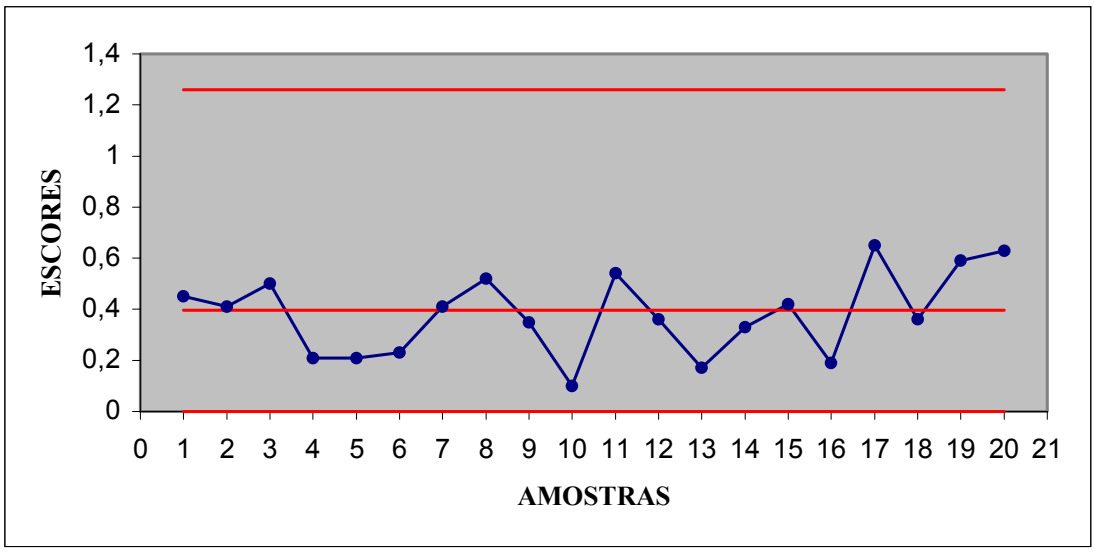

Figura 6. Carta de controle $|\mathbf{S}|$ para monitorar a variabilidade do processo.

É importante ressaltar que o uso da *variância generalizada* pode resultar em uma caracterização insuficiente da estrutura de covariância do processo. Montgomery (1996) cita, como exemplo, três diferentes matrizes de covariâncias:

$$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad \mathbf{S}_2 = \begin{bmatrix} 2{,}32 & 0{,}40 \\ 0{,}40 & 0{,}50 \end{bmatrix} \quad \text{e} \quad \mathbf{S}_3 = \begin{bmatrix} 1{,}68 & -0{,}40 \\ -0{,}40 & 0{,}69 \end{bmatrix}$$

A primeira matriz representa a estrutura de covariâncias estimada de um processo onde as duas variáveis são independentes (covariância zero). A segunda matriz representa a estrutura de covariância de um processo cujas variáveis apresentam correlação positiva, enquanto que na terceira as variáveis possuem correlação negativa. No entanto, apesar de possuírem estruturas de covariância diferente, $|\mathbf{S}_1| = |\mathbf{S}_2| = |\mathbf{S}_3| = 1$. Sugere-se, assim, a utilização de p cartas de controle univariadas para monitorar a variabilidade do processo, simultaneamente ao uso da carta de controle para $|\mathbf{S}|$.

2.2.4- Carta de controle multivariada para a soma acumulada (MCUSUM) para monitoramento das médias de processos

Os procedimentos de controle apresentados nas seções anteriores utilizam apenas a informação contida na amostra mais recente para inferir sobre o estado do processo. Informações contidas na sequência de amostras antecedentes à amostra mais recente são ignoradas. Tal esquema de monitoramento é bastante efetivo na detecção de mudanças estatisticamente significantes no vetor de médias do processo, mas relativamente insensível a pequenas mudanças no vetor de médias.

Para detectar pequenas mudanças no vetor de médias do processo, dois procedimentos de controle surgem como alternativa: a *carta de controle multivariada para a soma acumulada* (cartas de controle MCUSUM) e a *carta de controle multivariada para a média móvel ponderada exponencialmente* (carta de controle MEWMA).

Considere-se, inicialmente, uma situação em que $p = 1$. Amostras periódicas de tamanho unitário (n=1) são obtidas do processo. Suponha-se uma variável de interesse normalmente distribuída, com média $\mu = \mu_0$ (média do processo sob controle estatístico ou no valor alvo do

processo) e desvio-padrão $\sigma = \sigma_0$; $X \sim N(\mu_0, \sigma_0)$. Se σ_0 não for conhecido, deve-se estimá–lo através de amostras preliminares independentes obtidas do processo. Define-se, então, a seguinte estatística (Pignatiello & Runger, 1990):

$$C_k = \sum_{b=1}^{k} (x_b - \mu_0) \ , \quad \text{para } k = 1,2,..... \ , \tag{15}$$

onde C_k representa a soma das diferenças entre cada observação unitária (n=1) x_b e a média do processo, acumuladas até a $k^{ésima}$ amostra. Se o processo permanece sob controle estatístico, C_k se constitui em um passeio aleatório em torno do zero. Por outro lado, quando a média do processo muda para algum valor $\mu_1 > (<) \ \mu_0$, sucessivas diferenças positivas (negativas) serão acumuladas causando uma inflação positiva (negativa) no valor de C_k. Para exercer efetivamente o monitoramento da média do processo através dessas somas acumuladas, determinam-se os limites estatísticos de tolerância para C_k, apresentados mais a diante.

Montgomery (1996) define um procedimento de controle denominado *Algorithmic CUSUM* (*CUSUM* algorítmico). Escreve-se:

$$C_k^+ = \text{máx} \ [\ 0, x_k - (\mu_0 + a\sigma_0) + \ C_{k-1}^+ \ ] \quad \text{e}$$

$$C_k^- = \ \text{min} \ [\ 0, \ (\ \mu_0 - a\sigma_0 \ ) - x_k \ + \ C_{k-1}^- \ ], \tag{16}$$

onde x_k representa a observação mais recente, sendo $C_k^+ = \ C_k^- = 0$ os valores iniciais desta estatística. O parâmetro a usualmente representa a metade da distância entre a média do processo sob controle μ_0 e a média do processo fora de controle $\mu(\neq\mu_0)$, que se deseja detectar. Esta distância é expressa pelo número de desvios padrões em relação μ_0. Assim, com $\lambda(\mu) = (\mu - \mu_0) / \sigma_0$, pode-se expressar este deslocamento como $\mu = \mu_0 + \lambda(\mu) \ \sigma_0$, com a=$\lambda(\mu)/2$. Os escores C_k^+ e C_k^- acumulam, respectivamente, desvios positivos e negativos, em relação à média do processo, maiores do que a. Se C_k^+ ou C_k^- for superior a h, considera-se o processo fora de controle. Geralmente, o valor de h adotado é cinco vezes o desvio-padrão do processo (h=$5\sigma_0$). Os limites de controle dessa estatística são obtidos da seguinte forma:

$$\text{LSC} = \ h \quad \text{e}$$

$$LIC = -h \quad (17)$$

É importante ressaltar que a escolha de diferentes valores para os parâmetros *a* e *h* acarretam um impacto substancial no desempenho na carta CUSUM. O critério de escolha desses parâmetros será discutido a seguir; mas antes precisa-se definir a medida de desempenho para cartas de controle conhecida por Tamanho Médio de Corrida (*Average Run Length*, ARL).

Considere-se o procedimento de controle de Shewhart, apresentado na seção 2.3.1. A partir da medição da variável de interesse em sucessivas amostras unitárias, monitora-se a média μ_0 do processo. Os limites de controle utilizados na carta de controle são dados por $\mu_0 \pm 3\sigma_0$ ($Z_{\alpha/2} = \pm 3$). Assim, $\alpha = 0{,}0027$, ou seja, existe uma probabilidade de 0,27% da carta de controle acusar incorretamente uma mudança na média do processo (erro tipo I). O Tamanho Médio de Corrida para este processo, na ausência de causas especiais, é dado por $ARL_0 = 1/\alpha = 1/0{,}0027 \approx 370$ amostras. Isto significa que, estando o processo sob controle, espera-se retirar, em média, 370 amostras sucessivas antes que a carta gere um alarme falso.

Suponha-se, agora, que a média do processo tenha se desviado do valor nominal em um desvio-padrão ($\mu = \mu_0 + 1\sigma_0$). Adotando os mesmos limites de controle (ou seja, $Z_{\alpha/2} = \pm 3$, com $\alpha = 0{,}0027$), tem-se $ARL = 1/(1-\beta) = 1/0{,}0228 \approx 44$ amostras, onde 0,0228 é a probabilidade de que a carta detecte um desvio de $1\sigma_0$ na média do processo. Isto significa que serão necessárias em média 44 amostras sucessivas para que a carta sinalize uma mudança na média do processo, de μ_0 para $\mu(\neq\mu_0)$. O ARL deve ser entendido como sendo um medida de desempenho da carta para processos desviados da média ou valor alvo μ_0, obtida a partir da probabilidade α adotada.

ARL_0 e ARL são os parâmetros usados para medir o desempenho de uma carta CUSUM, a partir de uma determinada escolha de *a* e *h*. Assim, uma vez escolhido o valor de *a*, para um dado ARL_0, deve-se escolher *h* de forma a maximizar a chance de detectar um deslocamento $\lambda(\mu)$ na média μ_0 do processo. Em outras palavras, dada a probabilidade de alarme falso tolerada e o tamanho do deslocamento de interesse da média do processo, escolhe-se o valor de *h* que minimize o número médio de amostras sucessivas necessárias (ARL) para detectar o deslocamento na média do processo. Para desvios em torno de $1\sigma_0$ em relação a μ_0, a=1/2 e $h = 5\sigma_0$ acarretam um desempenho ARL bastante razoável para a carta CUSUM (Montgomery, 1996).

Algumas técnicas podem ser usadas para o cálculo de ARL_0 e ARL numa carta de controle CUSUM. Brook & Evans (1972) obtêm valores para os ARLs utilizando uma cadeia de Markov, para obter as probabilidades de transição de um estado sob controle para um estado fora de controle no processo. Woodall & Adams, *apud* Montgomery (1996) sugerem um procedimento para o cálculo dos ARLs que utiliza uma aproximação apresentada por Siegmund, *apud* Montgomery (1996). Dados *a* e *h*, pode-se obter o valor de ARL para uma carta CUSUM unilateral (C_k^+ ou C_k^-) através da seguinte expressão:

$$\text{ARL} = [\exp(-2\,\Delta\, b) + 2\,\Delta\, b - 1] / 2\,\Delta^2 \text{ , se } \Delta \neq 0 \quad \text{e}$$

$$= b^2 \text{ , se } \Delta = 0 \text{ ,} \qquad (18)$$

onde $\Delta = \lambda(\mu) - a$ para o caso da soma acumulada positiva C_k^+, $\Delta = -\lambda(\mu) - a$ para o caso da soma acumulada negativa C_k^-, e $b = h + 1{,}166$. Quando $\lambda(\mu) = 0$, a expressão (18) resulta em ARL_0. Quando $\lambda(\mu) \neq 0$, obtém-se o valor de ARL referente ao desvio de ordem $\lambda(\mu)$ em relação a μ_0 que se deseja detectar.

O valor de ARL para a carta CUSUM bilateral é dado por:

$$1 / \text{ARL} = 1 / \text{ARL}^+ + 1 / \text{ARL}^- , \qquad (19)$$

onde ARL^+ refere-se à soma acumulada positiva (C_k^+) e ARL^- refere-se à soma acumulada negativa (C_k^-), ambos obtidos através da expressão (18).

Adotando, por exemplo, $a = ½$, $h = 5$ e $\lambda(\mu) = 0$, obtém-se, através da expressão (18), $ARL_0^+ = ARL_0^- = 938{,}2$. Através de (19) obtém-se $ARL_0 \approx 469$. Desta forma, espera-se que a carta CUSUM gere um alarme falso após 469 amostras sucessivas, em média. Suponha-se, agora, que exista o interesse em avaliar o desempenho desta carta para um deslocamento de $1\sigma_0$ na média μ_0 do processo ($\lambda(\mu) = 1$). Adotando $a = ½$ e $h = 5$, tem-se $\Delta = 0{,}5$ em relação a C_k^+ e $\Delta = -1{,}5$ em relação a C_k^-. Usando as expressões (18) e (19), obtém-se $ARL \approx 10$. Este resultado mostra que a carta CUSUM é mais sensível para detectar mudanças na ordem de $1\sigma_0$ em relação à carta de Shewhart ($ARL \approx 44$ amostras).

No caso multivariado (p >1), as cartas MCUSUM são boas alternativas em relação à carta de Hotelling quando pequenos desvios em relação ao vetor das médias $\boldsymbol{\mu_0}$ do processo devem ser detectados. No controle univariado, o tamanho do desvio de interesse é representado em unidades de desvio-padrão [$\lambda(\mu) = (\mu_0 - \mu)/\sigma_0$]. A extensão multivariada dessa distância é dada por (Crosier, 1988):

$$\lambda\,(\boldsymbol{\mu}) = [(\,\boldsymbol{\mu} - \boldsymbol{\mu_0})'\ \boldsymbol{\Sigma_0^{-1}}\ (\boldsymbol{\mu} - \boldsymbol{\mu_0})]^{1/2}\,,$$

onde $\lambda\,(\boldsymbol{\mu})$ representa a distância, em p dimensões, entre o vetor de médias $\boldsymbol{\mu}(\neq\boldsymbol{\mu_0})$ do processo fora de controle e o vetor de médias $\boldsymbol{\mu_0}$ do processo sob controle [*distância de Mahalanobis* ou *"distância estatística*", conforme denominado por Johnson & Wichern (1992).

É importante ressaltar que existem diferentes vetores $\boldsymbol{\mu}(\neq\boldsymbol{\mu_0})$ de médias do processo equidistantes a $\boldsymbol{\mu_0}$, ou seja, $\lambda(\boldsymbol{\mu_1}) = \lambda(\boldsymbol{\mu_2})$ para $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2} \neq \boldsymbol{\mu_0}$. Suponha-se, por exemplo, que o vetor bi-variado de médias de um determinado processo sob controle seja $\boldsymbol{\mu_0} = (0, 0)'$ e que a matriz de covariâncias (2 × 2) do processo seja a matriz identidade. O desempenho ARL da carta de Hotelling será o mesmo para vetores de médias equidistantes em relação a $\boldsymbol{\mu_0}$. Assim, a carta terá o mesmo poder para detectar, por exemplo, deslocamentos no vetor de médias do processo na direção de $\boldsymbol{\mu_1}$ = (0,-1)', $\boldsymbol{\mu_2}$ = (1,0)' ou $\boldsymbol{\mu_3}$ = $(0{,}5\sqrt{2}\,)'$, pois $\lambda(\boldsymbol{\mu_1}) = \lambda(\boldsymbol{\mu_2}) = \lambda(\boldsymbol{\mu_3})$. É importante observar que cada uma das 3 situações representam diferentes deslocamentos na média de cada uma das variáveis. Desta forma, classifica-se a carta de Hotelling com sendo uma carta de controle com *direcionalidade invariante*.

De uma maneira geral, qualquer carta de controle multivariada que utiliza a matriz de covariâncias $\boldsymbol{\Sigma_0}$ (ou a estimativa amostral **S**) do processo em sua estatística de controle possui direcionalidade invariante (Lowry & Montgomery, 1995). O desempenho ARL dessas cartas é determinado apenas em função da distância entre o vetor de médias $\boldsymbol{\mu_0}$ do processo sob controle e o vetor de médias $\boldsymbol{\mu} \neq \boldsymbol{\mu_0}$ do processo fora de controle, e não em função de um deslocamento específico na média de uma determinada variável do processo.

A maioria das cartas de controle univariadas bilaterais, entre elas as cartas para médias de Shewhart e a carta CUSUM, possuem direcionalidade invariante (isto é, desempenho ARL obtido em função de $\lambda(\mu) = (\mu - \mu_0\,)\,/\sigma_0$). No entanto, num processo com p variáveis de interesse,

quando o controle multivariado está sendo feito a partir de *p* cartas univariadas, desconsiderando a correlação entre as variáveis, não se está utilizando um procedimento de controle com direcionalidade invariante, pois o desempenho ARL não depende de $\lambda(\boldsymbol{\mu})$, e sim da direção específica do deslocamento do vetor $\boldsymbol{\mu_0}$ (ver Pignatiello e Runger, 1990, para maiores esclarecimentos). Estes procedimentos de controle são chamados *de direcionalidade específica*.

Os procedimentos de controle multivariado que utilizam a filosofia CUSUM estão divididos em: (*i*) procedimentos de controle que utilizam múltiplas cartas CUSUM univariadas (abreviados por MCU), desconsiderando assim a correlação entre as variáveis (portanto, planos de direcionalidade específica); e (*ii*) procedimentos de controle que utilizam uma carta CUSUM multivariada (MCUSUM), isto é, utilizam a matriz $\boldsymbol{\Sigma_0}$ de covariâncias das variáveis (portanto, planos de direcionalidade invariante).

O primeiro procedimento MCU proposto por Woodall & Ncube (1985), consiste no monitoramento de *p* variáveis através de *p* cartas CUSUM univariadas. Neste procedimento, a matriz de covariâncias $\boldsymbol{\Sigma_0}$ do processo é utilizada para avaliar o pior desempenho ARL_0 (carta com maior chance de gerar um alarme falso) dentre as *p* cartas univariadas CUSUM. A média de cada variável é monitorada através da estatística de controle na expressão (16), com os limites de controle apresentados na expressão (17). Assim, a média da $i^{ésima}$ variável é monitorada obtendo-se os valores $C_{i,k}^{+}$ e $C_{i,k}^{-}$ a cada nova amostra dessa variável. A $i^{ésima}$ carta CUSUM sinaliza que a média da $i^{ésima}$ variável sofreu um deslocamento quando $C_{i,k}^{+} > h_i$ ou $C_{i,k}^{-} < -h_i$ para valores específicos de a_i e h_i. O MCU indicará que o processo está fora de controle quando pelo menos uma das *p* cartas univariadas detectar algum desvio em relação à média da respectiva variável.

Devido ao fato de possuir direcionalidade específica, o MCU deve ser usado num processo onde existe o interesse em detectar desvios numa direção específica preestabelecida em relação ao vetor $\boldsymbol{\mu_0}$ (Pignatiello e Runger, 1990). Estes desvios podem ser caracterizados através de deslocamentos das variáveis do processo sobre seus eixos (desvios específicos em relação às médias das variáveis) ou na direção dos eixos dos *Componentes Principais* (variáveis independentes que são combinações lineares das variáveis originais; ver seção 2.3.1).

Healy (1987) demonstrou que, quando existe o interesse em detectar um desvio da média em uma determinada direção preestabelecida, uma carta CUSUM univariada baseada na

combinação linear das p variáveis originais, estruturada naquela direção, fornecerá os bons resultados em termos de desempenho ARL. No entanto, se existir o interesse em detectar deslocamentos em várias direções, a estratégia proposta por Healy (1987) pode ser pouco eficiente (Pignatiello e Runger, 1990).

Hawkins (1991) aperfeiçoou a estratégia sugerida por Healy (1987), propondo um MCU, que mostrou ser mais eficiente em relação ao MCU apresentado por Woodall & Ncube (1985), para detectar desvios no vetor $\boldsymbol{\mu_0}$ em várias direções.

O primeiro procedimento MCUSUM foi proposto por Crosier (1988), consistindo em duas cartas de controle multivariadas CUSUM. A carta com melhor desempenho ARL, que será abreviada por CCUSUM, se constitui em uma extensão multivariada da carta univariada CUSUM. Considere novamente a estatística $C_k^+ = \text{máx}[0, C_{k-1}^+ + x_k - \mu_0 - a\sigma_0]$, apresentada na expressão (16). No caso multivariado, substituindo na expressão os escalares por vetores, tem-se:

$$\mathbf{C}_k = \text{máx}[\mathbf{0}, \mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0 - \mathbf{a}]$$

No caso univariado, sabe-se que a quantidade $C_{k-1}^+ + x_k - \mu_0 - a\sigma_0$ será exatamente zero (isto é, C_k=0) se, dado uma observação x_k, $C_{k-1}^+ + x_k - \mu_0 = a\sigma_0$; e será cada vez maior em relação a $a\sigma_0$ (isto é, C_k>0), à medida que consecutivas observações x_k sejam superiores à média μ_0 em mais do que a desvios-padrão σ_0. Já na expressão acima, ao avaliar o impacto da diferença $(\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu_0}) - \mathbf{a}$ no valor $\mathbf{C}_k$, tem-se dificuldade em encontrar e interpretar os valores de **a**, associados ao valores máximo e mínimo de $\mathbf{C}_k$.

Considera-se $a=[\mathbf{a}'\sum_0^{-1}\mathbf{a}]^{1/2}$ o tamanho do desvio de interesse em relação à $\boldsymbol{\mu_0}$ (em p dimensões), a ser detectado na carta CCUSUM. Neste caso, o vetor **a** deve ter comprimento a (comprimento calculado em função da matriz $\boldsymbol{\Sigma_0}$ de covariâncias). Se a diferença $(\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu_0}) - \mathbf{a}$ for para zero, então **a** deverá estar na mesma direção de $(\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu_0})$. Desta forma, tem-se

$$\mathbf{a}=\left(\frac{a}{F_k}\right)(\mathbf{C}_{k-1}+ \mathbf{x}_k - \boldsymbol{\mu_0}),$$

onde

$$F_k = [(\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0)]^{1/2}$$

representa o comprimento da parcela $\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0$. Rescreve-se então a diferença $\mathbf{C}_k = (\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0) - \mathbf{a}$ da seguinte forma:

$$\mathbf{C}_k = (\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0)\,(1 - a / F_k), \quad \text{se} \quad F_k > a,$$

isto é, se o comprimento de $\mathbf{C}_{k-1} + \mathbf{x}_k - \boldsymbol{\mu}_0$ for maior do que a; e

$$\mathbf{C}_k = \mathbf{0}, \text{ se } F_k \leq a.$$

Analogamente ao caso univariado (carta CUSUM), sugere-se como valor de a metade da distância entre o vetor de médias do processo sob controle $\boldsymbol{\mu}_0$ e o vetor de médias do processo fora de controle $\boldsymbol{\mu}(\neq\boldsymbol{\mu}_0)$, que se deseja detectar; assim $a = \lambda\,(\boldsymbol{\mu})/2$. A estatística de controle a ser plotada na carta é dada por:

$$Y_k = (\mathbf{C}_k' \, \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_k)^{1/2} \qquad k=1,2,....,$$

onde $\mathbf{C}_0 = \mathbf{0}$.

O processo será considerado fora de controle sempre que $Y_k > h_1$, para a e h_1 fixos. De maneira análoga ao procedimento CUSUM, dado a e especificado ARL_0, devemos escolher h de forma a maximizar a chance de detectar um deslocamento $\lambda\,(\boldsymbol{\mu})$ na média $\boldsymbol{\mu}_0$ do processo. Maiores detalhes para a obtenção de h_1 podem ser encontrados em Crosier (1988).

A carta CCUSUM possui uma característica desejada, quando monitoram-se p variáveis simultaneamente. Se o vetor de médias do processo deslocar-se para $\boldsymbol{\mu}(\neq\boldsymbol{\mu}_0)$, a carta fornece indicações em relação à possível direção deste deslocamento. Croiser (1988) demostra que, de uma maneira geral, esta carta possui um melhor desempenho ARL em relação à carta de Hotelling na detecção de deslocamentos no vetor de médias do processo.

Pignatiello & Runger (1990) também propuseram duas cartas CUSUM para controle multivariado de processos. O procedimento com melhor desempenho ARL, denominado MC1, é baseado no seguinte vetor de somas acumuladas:

$$\mathbf{C}_k = \sum_{b=1}^{k} (\mathbf{x}_b - \boldsymbol{\mu}_0) \ , \quad \text{para } k = 1,2,... \ .$$

Esta estatística corresponde à expressão (15) em sua versão multivariada. A estatística de controle da carta MC1 é dada por:

$$MC_k = \text{máx } \{ 0, \ (\mathbf{C}'_k \ \Sigma_0^{-1} \mathbf{C}_k)^{1/2} \ - an_k \},$$

onde $n_k = n_{k-1} + 1$ se $MC_{k-1} > 0$ e 1, caso contrário. Este parâmetro pode ser interpretado como sendo o número de amostras consecutivas desde a última renovação (isto é, desde que MC_k=0 pela última vez). O processo será considerado fora de controle se $MC_k > h_2$, para a e h_2 fixos. Novamente, sugere-se utilizar um valor de $a = \lambda(\boldsymbol{\mu}_1)/2$. Maiores detalhes para a obtenção de h_2 podem ser encontrados em Pignatiello & Runger (1990).

Pignatiello & Runger (1990) demonstraram que a carta MC1 possui melhor desempenho ARL em relação à carta de Hotelling na detecção de deslocamentos na média $\boldsymbol{\mu}_0$ do processo inferiores a 3 (λ ($\boldsymbol{\mu}$) < 3). Esses autores também mostraram que a carta MCU, proposta por Woodall e Ncube (1985), e a carta MC1 parecem ter o mesmo desempenho ARL, quando o deslocamento no vetor de médias do processo for provocado por apenas uma das variáveis em monitoramento. No entanto, a carta MC1 possui, em geral, um melhor desempenho ARL em relação a MCU quando o deslocamento no vetor de médias do processo for provocado por desvios simultâneos nas médias de mais de uma variável.

Outros procedimentos propostos por Crosier (1988) e Pignatiello e Runger (1990), baseiam-se na carta univariada CUSUM em função dos valores obtidos através da usual estatística de Hotelling .

Segundo Pignatiello e Runger (1990), de uma maneira geral, as cartas de controle com direcionalidade invariante (carta de Hotelling, carta CCUSUM, carta MC1), são mais sensíveis em relação às cartas de direcionalidade específica (carta de Shewhart e cartas MCU) quando

múltiplos deslocamentos nas médias das variáveis, que acarretam o mesmo desvio no vetor de médias do processo, devem ser monitorados. Isto acontece porque as cartas de direcionalidade invariantes são, simultaneamente, sensíveis à magnitude e à direção do deslocamento. Em contrapartida, quando estas cartas geram um sinal indicando um deslocamento no vetor de médias, geralmente não se tem idéia da direção deste deslocamento, ou seja, das variáveis que efetivamente sofreram alteração em relação as suas médias. O mais aconselhado é usar uma carta de direcionalidade invariante para monitorar o vetor de médias e cartas de direcionalidade específica para diagnosticar um sinal de processo fora de controle.

2.2.5- Carta de controle multivariada para a média móvel exponencialmente ponderada (MEWMA)

A carta de controle MEWMA também se constitui numa boa alternativa em relação ao procedimento de Hotelling quando existe o interesse em detectar pequenos desvios no vetor $\boldsymbol{\mu_0}$ de médias do processo (Lowry & Montgomery, 1995). A carta MEWMA é uma extensão multivariada da carta univariada EWMA. Para o caso univariado (p= 1), define-se, para n = 1, a seguinte estatística de controle:

$$M_k = r\,x_k + (1-r)M_{k\text{-}1}\,, \quad \text{para} \quad k = 1,2,\ldots\,, \tag{20}$$

onde x_k representa a observação referente à $k^{\text{ésima}}$ amostra unitária, r é a constante de ponderação $\{\, r \in (0,1] \,\}$ e $M_{k\text{-}1}$ é o valor obtido a partir da amostra k-1. Suponha-se que $M_0 = \mu_0$ (média do processo sob controle). Considerando a suposição de normalidade das observações x_k ($X_k \sim N(\mu_0\,,\, \sigma_0)$) e independência entre amostras, o desvio-padrão de M_k é dado por (Lowry & Woodall, 1992)

$$\sigma_{0M_k} = \sqrt{\sigma_0^2 \left(\frac{r}{2-r}\right)\left[1-(1-r)^{2k}\right]}\,.$$

Os limites de controle para a estatística na expressão (20) são dados por:

$$\text{LSC} = \mu_0 + L\,\sigma_{0M_k} \ ,$$

$$\text{LC} = \mu_0 \ ,$$

$$\text{LIC} = \mu_0 - L\,\sigma_{0M_k} \ ,$$

onde L representa o número de desvios-padrão tolerados em relação a M_k.

O critério para a escolha dos valores de r e L é semelhante ao critério de escolha de a e h para a carta CUSUM (*algoritmo CUSUM*). Especifica-se, inicialmente, o ARL_0 (número médio de amostras consecutivas até que a carta gere um alarme falso) a ser tolerado e a magnitude do desvio a ser detectado pela carta. A partir daí, determinam-se os valores de r e L que minimizam ARL (número médio de amostras consecutivas necessárias para que a carta detecte o desvio especificado). Lucas e Saccucci (1990) comparam o desempenho ARL da carta EWMA para diferentes valores de r e L. Segundo Montgomery (1996), $L = 3$ e pequenos valores de r devem ser utilizados na detecção de pequenos desvios na média do processo.

Para o caso multivariado ($p > 1$), a estatística na expressão (20) pode ser adaptada da seguinte forma (Lowry & Woodall, 1992):

$$\mathbf{M}_k = \mathbf{R}\,\mathbf{x}_k + (\mathbf{I}_p - \mathbf{R})\,\mathbf{M}_{k-1} \ , \quad \text{para} \quad k = 1,2,... \ ,$$

onde $\mathbf{x}_k$ é uma observação representada por um vetor p-dimensional referente a $k^{ésima}$ amostra unitária ($\boldsymbol{X}_k \sim N(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$, a matriz $\mathbf{R} = \text{diag}(r_1, r_2,...,r_p)$ contém p constantes de ponderação $\{r_i \in (0,1]$, para $i = 1, 2,..., p\}$ e $\mathbf{M}_{k-1}$ é o vetor p-dimensional dos valores referentes à amostra k-1. Suponha-se $\mathbf{M}_0 = \mathbf{0}$. Se não existir razão, *a priori*, para atribuir diferentes pesos para as variáveis monitoradas na carta, então $r_1 = r_2 = ... = r_P = r$. Monitora-se um processo multivariado utilizando o procedimento MEWMA através da seguinte estatística de controle (Lowry & Woodall, 1992):

$$TM_k^2 = \mathbf{M}_k' \ \boldsymbol{\Sigma}_{\mathbf{0}M_k}^{-1} \mathbf{M}_k \ ,$$

onde $\boldsymbol{\Sigma}_{M_k} = \{r\,[1 - (1 - r)^{2k}] / (2 - r)\}\boldsymbol{\Sigma}_0$. O processo será considerado fora de controle se $TM_k^2 > h_3$. O valor de h_3 é escolhido a partir do desempenho ARL desejado para a carta (para maiores esclarecimentos; ver Lowry *et al.*, 1992). Deve-se observar que quando as p características de qualidade recebem o mesmo peso r, o desempenho ARL da carta MEWMA depende apenas do parâmetro $\lambda(\boldsymbol{\mu})$. Desta forma, este procedimento de controle também apresenta a característica de *direcionalidade invariante*. Por esta razão, pode-se comparar o desempenho dessa carta em relação as cartas de Hotelling, CCUSUM e MC1.

Lorwy *et al.* (1992) mostram que a carta MEWMA possui um desempenho ARL superior em relação as demais cartas (Hotelling, CCUSUM e MC1), quando essa é implantada num processo inicialmente fora de controle ($\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$). Quando o processo está inicialmente sobre controle, a carta MEWMA é pelo menos tão boa quanto as demais na detecção de deslocamentos no vetor de médias $\boldsymbol{\mu}_0$.

Yashchin (1987) chama a atenção para o elevado grau de "inércia" da carta univariada EWMA em relação a outros procedimentos, tais como a carta CUSUM. Se, por exemplo, o valor de M_k, na expressão (20), estiver no lado negativo em relação à linha central (M_k < LC) na carta e existir um desvio positivo na média do processo, este aumento será detectado com algum retardo. Lucas e Saccucci (1990) propõem, para minimizar o grau de inércia na carta EWMA, o uso simultâneo da carta EWMA e da carta de Shewhart (deve-se lembrar que esta última não considera as observações anteriores coletadas do processo).

Um elevado grau de inércia também pode ocorrer nas cartas MCUSUM e MEWMA. De maneira análoga ao caso univariado, Lowry & Woodall (1992) recomendam o uso da carta de Hotelling em conjunto com estas cartas. Dessa maneira, estabelece-se um compromisso entre baixa inércia (carta de Hotelling) e rápida detecção de pequenos deslocamentos no vetor de médias (MCUSUM e MEWMA).

2.2.6- Procedimentos de controle para dados de processo autocorrelacionados

As cartas de controle apresentadas foram construídas a partir de *m* amostras preliminares, $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_m$, consideradas estatisticamente independentes, com amostras subseqüentes também independentes. Em indústrias químicas e de processos, por exemplo, esta hipótese, com freqüência, não se verifica, pois o intervalo de tempo entre duas amostras costuma ser pequeno. Montgomery (1992) e Montgomery & Mastrangelo (1991), entre outros, mostram que, em tais situações, as variáveis individuais mensuradas em cada vetor $\boldsymbol{X}_k$ apresentarão elevada autocorrelação, o que pode prejudicar o desempenho ARL_0 das cartas de controle.

Montgomery *et al.*, *apud* Lowry & Montgomery (1995) mostram que a estrutura de autocorrelação das variáveis de processo pode ser identificada através da Análise de Componentes Principais (ACP), usando dados obtidos do processo sob controle. Devido à forte autocorrelação entre as variáveis de processo, dois ou três Componentes Principais costumam oferecer uma caracterização satisfatória dessas variáveis. Quando dispõem-se de dados do processo (**X**) e dados de qualidade (**Y**), Kresta *et al.* (1991) recomendam o uso de um procedimento denominado Projeção de Estruturas Latentes (PEL), para capturar a estrutura de autocorrelação dos dados, em lugar da ACP. Para identificar um sinal fora de controle obtido via ACP ou PEL, pode-se utilizar cartas de controle univariadas (uma para cada variável de processo envolvida).

Montgomery (1996) sugere alguns procedimentos de controle univariado para dados autocorrelacionados baseados na utilização de modelos de séries temporais. Nesses procedimentos, ao invés de monitorarem-se os valores observados de uma variável de interesse, monitoram-se os resíduos por eles gerados nos modelos de séries temporais, através de cartas de Shewhart. Mason *et al.* (1997) comentam que o ajuste de modelos de séries temporais para dados multivariados também é possível, apesar de não muito favorecido em aplicações práticas.

2.2.7- Procedimentos *a posteriori* para identificar variáveis *fora de controle* em cartas multivariadas

De acordo com Jackson (1991), um procedimento de controle multivariado deve: (*i*) permitir a rápida identificação do estado do processo (sob controle ou fora de controle), (*ii*) permitir o estabelecimento da probabilidade associada ao erro tipo I, (*iii*) levar em conta a estrutura de correlação das variáveis envolvidas, e (*iv*) oferecer uma rápida identificação da origem de causas especiais incidentes sobre o processo.

O item (*iv*) acima é o de mais difícil obtenção. Sempre que uma carta multivariada detecta causas especiais de variação no processo, não se dispõe de informações claras acerca das variáveis que saíram do padrão preestabelecido de variação. Em outras palavras, não se sabe, ao certo, as razões que levaram o processo ao estado fora de controle. Este problema se torna mais grave à medida que o número de variáveis aumenta. A dificuldade de interpretação de um sinal fora de controle numa carta multivariada tem sido exaustivamente discutida por inúmeros autores. Os vários métodos propostos podem ser basicamente divididos em dois grupos: (*i*) procedimentos de diagnóstico via cartas univariadas, e (*ii*) procedimentos de diagnóstico via decomposição da estatística T^2 de Hotelling.

2.2.7.1- Métodos de diagnóstico via cartas univariadas

Alt (1985) sugere o uso da carta multivariada de Hotelling simultaneamente ao uso de p cartas univariadas de Shewhart no controle do processo. A idéia é utilizar a carta multivariada para detectar mudanças no processo. Cartas multivariadas são mais robustas se comparadas às cartas univariadas, por considerarem a estrutura de correlação das variáveis monitoradas. Uma vez tendo sido gerado um sinal fora de controle, verifica-se, através das cartas univariadas, qual grupo de variáveis pode ter sido responsável pelo sinal. Para minimizar o erro tipo I, aconselha-se

o uso dos limites de controle de *Bonferroni* nas cartas univariadas. Estes limites são obtidos substituindo $\alpha/2$ por $\alpha/2p$ em cada uma das p cartas univariadas. Desta forma, abrem-se os limites de controle nestas cartas buscando obter uma probabilidade total de alarme falso de no máximo α (erro tipo I arbitrado na carta de Hotelling).

O procedimentos MCU, sugeridos por Woodall e Ncube (1985) , Healy (1987) e Hawkins (1991) também usam cartas univariadas (cartas CUSUM, no caso) para detectar mudanças no vetor $\boldsymbol{\mu_0}$ de médias do processo, sendo que os procedimentos MCU apresentados por Healy (1987) e Hawkins (1991) utilizam cartas CUSUM univariadas projetadas para detectar um desvio numa direção específica preestabelecida, isto é, para sinalizar a possível presença de uma causa especial de variação conhecida.

Pignatiello e Runger (1990) também sugerem a utilização de cartas univariadas para auxiliar na interpretação dos sinais fora de controle gerados pela carta MC1.

Mason *et al.* (1997), entre outros, apresentam restrições ao uso de cartas univariadas para diagnosticar um sinal fora de controle num processo multivariado. Em primeiro lugar, é difícil determinar a probabilidade total α de alarme falso associada ao uso simultâneo de p cartas univariadas. Alt (1985) propõe o uso dos limites de Bonferroni para minimizar a probabilidade α. No entanto, se existem muitas variáveis, esses limites serão demasiadamente alargados nas p cartas univariadas, dificultando assim a detecção de um efetivo deslocamento na média de alguma variável do processo. Em segundo lugar, quando a carta multivariada detecta um desvio no processo, não significa necessariamente que apenas a média de uma ou mais variáveis tenha mudado, mas sim que a atual estrutura de correlação pode não estar de acordo com a estrutura de correlação das variáveis no processo sob controle. Por não levar em consideração a correlação entre as variáveis (pois não usam a matriz de covariâncias $\boldsymbol{\Sigma_0}$ do processo), as cartas univariadas tornam-se pouco sensíveis para detectar mudanças desta natureza.

2.2.7.2- Métodos de diagnóstico via decomposição da estatística T^2 de Hotelling

Diversos autores têm sugerido técnicas para diagnosticar sinais fora de controle em cartas multivariadas baseadas na decomposição da estatística T^2 de Hotelling. A idéia é dividir T^2 em duas ou mais partes, onde cada parte representa um subgrupo de q ($< p$) variáveis.

Considere o caso onde exista o interesse em monitorar o vetor p dimensional $\boldsymbol{\mu_0}$ de médias do processo, a partir de m amostras preliminares unitárias (n=1) recolhidas do processo sob controle estatístico (Segunda fase). Quando uma nova amostra é coletada, uma observação $\mathbf{x}'_k$ (com $k = m+1,...$) é obtida e o devido valor T_k^2 de Hotelling é calculado através da expressão (11). Para um dado valor de α, se $T_k^2 > [p(m+1)(m-1)/m(m-p)]F_{\alpha,p,m-p}$, o limite superior de controle na expressão (14), um sinal de processo fora de controle será gerado pela carta de Hotelling. No entanto, para verificar as variáveis responsáveis por esse sinal, pode-se decompor T^2 de várias maneiras.

Mason *et al.* (1995) sugerem que a estatística T^2 seja dividida em componentes independentes, cada um refletindo a contribuição de uma variável individual. Utilizando um resultado originalmente obtido por Rencher (1993), estes autores apresentam, inicialmente, a seguinte decomposição:

$$T^2 = T^2_{p-1} + T^2_{p.1,2,...,p-1} \qquad (21)$$

onde

$$T^2_{p-1} = (\boldsymbol{X}_k^{(p-1)} - \overline{\boldsymbol{X}}^{(p-1)})' \mathbf{S}^{-1(p-1)} (\boldsymbol{X}_k^{(p-1)} - \overline{\boldsymbol{X}}^{(p-1)})$$

é a estatística T^2 de Hotelling em relação às p-1 primeiras variáveis, representando a distância padronizada, em p-1 dimensões, entre o $k^{ésimo}$ vetor de observações $\boldsymbol{X}_k^{(p-1)}$ (k=1,2,...) e o vetor $\overline{\boldsymbol{X}}^{(p-1)}$ das médias estimadas do processo, com $\mathbf{S}^{(p-1)}$ representando a matriz de covariâncias estimada do processo, ambos ($\overline{\boldsymbol{X}}^{(p-1)}$ e $\mathbf{S}^{(p-1)}$) obtidos a partir de m amostras preliminares;

$$T^2_{p.1,2,...,p-1} = \frac{X_{p,k} - \overline{X}_{p.1,...,p-1}}{S_{p.1,...,p-1}}$$

é o $p^{\text{ésimo}}$ componente do vetor $\boldsymbol{X}_k$, ajustado pelas estimativas da média $\overline{X}_{p.1,...,p-1}$ e do desvio-padrão $S_{p.1,...p\text{-}1}$ da distribuição condicional de X_p dado $X_1, X_2,...,X_{p\text{-}1}$. Essas estimativas são obtidas através de um modelo de regressão linear múltipla, usando m amostras preliminares, de X_p em função de $X_1, X_2,...,X_{p\text{-}1}$. Desta forma, o escore $T^2_{p.1,2,...,p-1}$ representa a distância padronizada unidimensional entre a observação $X_{p,k}$ e a média esperada dessa variável, dados os valores $\boldsymbol{X}_k^{(p-1)}$ obtidos nas p-1 demais variáveis. Dada uma nova observação $\mathbf{x}'_k$ (com $k = m+1,..$), pode-se utilizar a estatística $T^2_{p.1,2,...,p-1}$ para verificar se a relação entre a variável p e as demais p-1 variáveis não se alterou, considerando a estrutura padrão de correlação entre elas.

De forma interativa, pode-se decompor a expressão (21) em p partes independentes, obtendo o seguinte resultado:

$$T^2 = T^2_1 + T^2_{2.1} + T^2_{3.1,2} + T^2_{4.1,2,3} + ... + T^2_{i+1.1,...,i} . \qquad (22)$$

Mason *et al.* (1995) mostraram que a estatística $T^2_{i+1.1,...,i}$ multiplicada por uma constante segue uma distribuição F, com parâmetros 1 e (m-1). A significância α de qualquer um dos termos apresentados em (22) pode ser verificada comparando os valores obtidos com os seguintes limites:

$$\text{LSC} = \frac{m+1}{m} F_{\alpha,(1,m-1)} \quad \text{e}$$

$$\text{LIC} = 0. \qquad (23)$$

Pode-se decompor a estatística T^2 de diferentes maneiras. Dado que o valor T^2 apresentou um resultado significativo (T^2>LSC), pode-se, por exemplo, verificar a relação entre cada variável e as p-1 variáveis restantes. Isto pode ser feito obtendo os seguinte escores:

$$T^2_{1.2,...,p}, T^2_{2.1,3,...,p}, T^2_{3.1,2,4,...,p}, ..., T^2_{p.1,...,p-1}, \qquad (24)$$

onde a significância de cada um desses termos pode ser verificada comparando estes valores com os limites na expressão (23). De uma maneira geral, deve-se adotar uma estratégia de decomposição considerando as características do processo em questão, tal que a determinação das razões que levaram esse processo a sair do controle sejam facilmente identificadas. Mason *et al.* (1995) mostram que os principais procedimentos de diagnóstico propostos na literatura são casos especiais do método de partição apresentado acima.

Hawkins (1993) apresenta três diferentes estratégias para a decomposição de T^2. Duas dessas estratégias destacam-se e relacionam-se com as decomposições propostas por Mason *et al.* (1995) nas expressões (22) e (24). O primeiro procedimento baseia-se na seguinte decomposição:

$$T^2 = \sum_{i=1}^{p} (X_i - \overline{X}_i) Z_i \ ,$$

onde Z_i é o resíduo padronizado obtido através da regressão linear múltipla da $i^{\text{ésima}}$ variável sobre as p-1 demais variáveis no vetor $\boldsymbol{X}$ (X_i em função de $X_1, X_2, ...X_{i-1}, X_{i+1}, ..., X_p$), obtido a partir de m amostras preliminares. T^2 é expressa como uma soma de produtos entre Z_i e o correspondente desvio de X_i em relação a sua média $\overline{X}_i$. Dado uma nova observação $\mathbf{x}'_k$ (com $k = m+1,...$), Hawkins (1993) sugere uma estratégia de diagnóstico utilizando p cartas de controle, uma para cada $Z_{i,k}$. Assim, se $T_k^2 >$LSC e, se para alguma variável i, o valor $Z_{i,k}$ ultrapassar certos limites, pode-se concluir que a relação desta variável com as demais p-1 variáveis não está condizente com a estrutura padrão de correlação existente entre elas. Mason *et al.* (1995) mostram que a estatística Z_i pode ser escrita da seguinte forma:

$$Z_i = \frac{1}{S_{i.1,...,i-1,i+1,...,p}} T^2_{i.1,...,i-1,i+1,...,p} \, .$$

Na expressão acima, Z_i pode ser obtido através do produto entre a estatística T^2, utilizando a decomposição na expressão (24), e o inverso do desvio-padrão condicional da $i^{ésima}$ variável, dados $X_1,...,X_{i-1},X_{i+1},...,X_p$.

A segunda estratégia proposta por Hawkins (1993) baseia-se na seguinte decomposição:

$$T^2 = \sum_{i=1}^{p} Y_i^2 \ ,$$

onde Y_i representa o resíduo padronizado obtido através da regressão linear múltipla da $i^{ésima}$ variável em função das i-1 variáveis anteriores no vetor $\boldsymbol{X}$ (X_i em função de X_1, X_2,...$X_{i\text{-}1}$), a partir de m amostras preliminares. Hawkins (1993) sugere uma estratégia de diagnóstico utilizando p cartas de controle, uma para cada Y_i. Assim, dada uma nova observação $\mathbf{x}'_k$ (com $k = m$+1,...), se T_k^2 >LSC e, se para alguma variável i, o valor $Y_{i,k}$ ultrapassar certos limites, pode-se concluir que a relação desta variável com as primeiras i-1 variáveis não está condizente com a estrutura padrão de correlação entre elas. Hawkins (1993) aconselha a utilização de sua segunda estratégia em processo denominados "em cascata", onde a matéria-prima passa por uma série de etapas até a obtenção do produto final. Neste caso, a qualidade dos subprodutos ao final de cada etapa de produção é função da conformidade com o padrão na etapa em questão e nas etapas anteriores. Assim, um eventual descontrole em alguma etapa do processo será refletido na qualidade do subproduto gerado nesta etapa e nas etapas posteriores. Em processos deste tipo, cada variável X_i representa a medição da qualidade do subproduto obtido ao final da $i^{\text{ésima}}$ etapa. Em cada amostra tem-se, então, o vetor p- dimensional $\boldsymbol{X}$ representando as medidas de qualidade dos subprodutos produzidos nas p etapas. Mason *et al.* (1995) mostram que Y_i pode ser escrita da seguinte forma:

$$Y_i = \frac{1}{S_{i.1,\ldots,i-1}} T^2_{i.1,\ldots,i-1}$$

Deste modo, Y_i pode ser obtido através do produto entre a estatística T^2, utilizando a decomposição apresentada na expressão (22), e o inverso do desvio-padrão condicional da $i^{ésima}$ variável, dados X_1,...,$X_{i\text{-}1}$.

Zhang, *apud* Wade & Woodall (1993) e Constable *et al.*, *apud* Lowry & Montgomery (1995) apresentaram um procedimento de diagnóstico denominado *cause-selecting chart* (carta de controle para detecção de causas). Estes autores defendem a idéia da construção de uma carta de controle para monitorar uma variável X_i, do vetor $\boldsymbol{X}$, mediante o ajuste de suas observações, via regressão. Esta estratégia é semelhante às estratégias propostas por Hawkins (1993). A

diferença dá-se pelo fato de que o modelo de regressão é proposto para monitorar a média de X_i considerando o efeito de alguma covariável preditora que não faz parte do vetor $\boldsymbol{X}$. Wade & Woodall (1993) fazem uma revisão deste procedimento, propondo alguns aperfeiçoamentos.

Wierda, *apud* Mason *et al*. (1997) apresenta uma estratégia para a decomposição de T^2 denominada *step-down*. Este procedimento baseia-se numa ordenação *a priori* das médias das p variáveis. De acordo com a ordenação, o procedimento particiona o vetor de médias em q ($< p$) subvetores e define q subhipóteses; cada subhipótese testa a possibilidade de um determinado subvetor de médias não ter sofrido alterações. Uma carta de controle é associada a cada uma das q subhipóteses. A estatística de controle utilizada na $i^{\text{ésima}}$ carta é:

$$F_i = \frac{T_i^2 - T_{i-1}^2}{1 + T_{i-1}^2/(m-1)} \quad ,$$

onde T_i^2 e T_{i-1}^2 representam, respectivamente, a estatística T^2 em relação as i e as i-1 primeiras variáveis do subvetor escolhido. Utilizando o resultado na expressão (22), podemos escrever:

$$F_i = \frac{1}{1 + T_{i-1}^2/(m-1)} T_{i.1,\ldots,p}^2$$

O resultado acima deixa claro que este procedimento é semelhante a segunda estratégia proposta por Hawkins (1993).

Murphy (1987) particiona as p variáveis em dois subgrupos, onde um deles deve representar as variáveis suspeitas de terem causado o sinal de processo fora de controle. A estatística de controle é dada por:

$$D_i = T^2 - T_i^2 ,$$

onde T_i^2 representa a estatística T^2 considerando as i primeiras variáveis do subgrupo suspeito. É importante observar que neste procedimento, ao contrário do método proposto por Wierda, *apud* Mason *et al*. (1997) não existe uma ordenação *a priori* das médias das p variáveis. Utilizando os resultados nas expressões (21) e (22), Mason *et al.* (1995) mostram que a estatística D_i pode ser escrita como:

$$D_i = \sum_{j=i+1}^{p} T^2_{j.1,2,...,j-1} \ .$$

Doganaksoy *et al.* (1991) propõem uma estratégia de decomposição denominada ranqueamento univariado. Este procedimento consiste em particionar a estatística T^2 em p termos incondicionais. Em outras palavras, obtém-se o valor da estatística de Hotelling referente a cada variável individualmente. Temos assim p cartas univariadas de Hotelling (na verdade, p cartas de Shewhart). Esta abordagem é semelhante ao procedimento de Alt (1985), e também representa um caso particular em relação aos critérios de decomposição apresentados por Mason *et al.* (1995).

Jackson (1980), a exemplo de Alt (1985), sugeriu o uso da carta de Hotelling para detectar mudanças no processo, em conjunto com as cartas univariadas, para diagnosticar o problema. O procedimento em Jackson (1980) se diferencia do procedimento de Alt (1985) pelo fato de utilizar as cartas univariadas para monitorar p Componentes Principais (ao invés das p variáveis originais). Os Componentes Principais (CPs) representam a projeção das variáveis originais de controle nas direções de máxima variabilidade dos dados. Jackson mostra que a estatística T^2 pode ser escrita em função dos CPs. A expressão (11) pode então ser reescrita da seguinte forma:

$$T^2 = \sum_{i=1}^{p} \frac{C_i^2}{\lambda_i} \ , \qquad (25)$$

onde C_i , para i = 1, 2, ..., p, representa o $i^{\text{ésimo}}$ Componente Principal e λ_i, sua respectiva variância.

Por serem independentes, alguns CPs poderão refletir um conjunto específico de eventos anômalos no processo (causas especiais de variação). Desta forma, se uma causa especial está presente, um dos CPs deve apresentar um desvio em relação ao seu padrão. Como cada CP é uma combinação linear das p variáveis originais, pode-se determinar as variáveis que mais contribuem para sua formação, encontrando, assim, o grupo de variáveis responsáveis pelo sinal fora de controle, gerado pela carta de Hotelling.

A idéia de diagnosticar o processo através dos CPs foi inicialmente criticada pelo fato destes não possuírem, na maioria dos casos, interpretação física. Miller *et al.* (1993) apresentaram um método para determinar o efeito exercido por cada variável em um Componente Principal. Este procedimento gráfico denominado Gráficos de Contribuição será descrito em detalhes no Capítulo 3.

Jackson (1991), aperfeiçoando sua idéia original, sugere, para fins de diagnóstico, o uso de cartas de controle univariadas para os Componentes Principais, utilizando limites de controle de Bonferroni. A escolha desses limites resolve o problema do erro tipo I, apresentado pelo procedimento de Alt (1985), devido à independência entre Componentes Principais.

2.3- Cartas de controle multivariadas baseadas em métodos de projeção

Através da utilização de sistemas de controle informatizados, empresas industriais buscam uma caracterização mais completa de seus processos, gerando, numa base freqüente, um grande volume de dados acerca de inúmeras variáveis de desempenho dos processos (variáveis de processo). Tais variáveis tendem a apresentar dependência entre si, visto que, em aplicações práticas, poucos eventos caracterizam o processo em qualquer instante do tempo. Conseqüentemente, existem redundâncias na representação do sistema, já que diferentes combinações de variáveis são representações alternativas de um mesmo grupo de eventos subjacentes ao processo. Em contrapartida, cada variável traz alguma informação distinta sobre o processo, o que torna necessária a sua análise conjunta.

McGregor (1995) cita algumas razões que justificam a necessidade do uso de procedimentos de controle multivariados em processos industriais informatizados:

- Um maior número de variáveis do processo (***X***) são mensuradas, se comparadas ao número de variáveis de qualidade do produto final (***Y***). Em processos químicos, em

particular, dados do processo são obtidos na razão de 100/1 em relação aos dados de qualidade final.

- Variáveis de processo são mensuradas a cada minuto ou segundo. Em contrapartida, a medição de variáveis de qualidade costuma tomar horas.

- Devido à estrutura informatizada dos sistemas de controle, variáveis do processo são mensuradas em tempo real, enquanto que mensurações da qualidade são obtidas mediante análise do produto final em laboratório. Desta forma, mensurações do processo estão instantaneamente disponíveis, enquanto que mensurações de qualidade estão disponíveis com algumas horas de atraso.

- Mensurações das variáveis do processo são mais precisas em relação às variáveis de qualidade, pois as primeiras são feitas por sensores conectados ao processo, enquanto mensurações de qualidade envolvem análises químicas complexas.

- Se um evento especial presente no processo afetar a qualidade do produto, provavelmente causará alterações nas variáveis do processo. Por outro lado, causas especiais podem não provocar alterações nas variáveis de qualidade, pois, em determinados processos, estas variáveis não oferecem uma caracterização completa da qualidade do produto final. O monitoramento de variáveis de qualidade em conjunto com variáveis do processo oferece uma caracterização mais completa da qualidade do produto final.

Os métodos tradicionais de controle multivariado (carta de Hotelling, carta MCUSUM e carta MEWMA, por exemplo) não são suficientemente robustos para tratar com um grande número de variáveis correlacionadas, pois foram concebidos para monitorar variáveis de qualidade do produto final (isto é, um pequeno número de variáveis correlacionadas). A estrutura colinear das variáveis do processo faz com que as estatísticas de controle utilizadas nestas cartas forneçam sinalizações distorcidas acerca do estado do processo. Em algumas situações, tais estatísticas não poderão ser calculadas, devido à dificuldade de inversão da matriz de covariâncias das variáveis de processo. Surge então a necessidade de utilizar cartas de controle baseadas em *métodos de projeção de dados*.

O controle estatístico multivariado do processo (CEMP) está alicerçado nos métodos de projeção de dados via *Componentes Principais* (ACP) e via *Projeção de Estruturas Latentes* (PEL). O objetivo dos métodos de projeção é transformar variáveis de processo (e/ou variáveis de qualidade) em um grupo reduzido de variáveis (variáveis latentes), preservando informações relevantes contidas nas variáveis originais e eliminando redundâncias no sistema inicial, representadas por variáveis colineares. As variáveis latentes são combinações lineares das variáveis originais. Desta forma, quando os dados originais são projetados num espaço de dimensões menores, trazem consigo a estrutura de covariância das variáveis originais. Normalmente, duas ou três variáveis latentes são suficientes para representar variáveis originais em maior número.

2.3.1- Carta de controle baseada na projeção de dados via ACP

Suponha p variáveis de processo sendo simultaneamente monitoradas. Seja $\boldsymbol{X} = [X_1,X_2,...,X_p]$ um vetor contendo p variáveis aleatórias, pode-se obter um vetor com p novas variáveis latentes $\boldsymbol{C} = [C_1,C_2,...,C_p]$, onde cada C_i representa uma combinação linear das p variáveis iniciais. Seja $\mathbf{X}$ uma matriz, $m \times p$, de dados preliminares do processo *centrados* (referentes a m amostras independentes), contendo uma observação p-variada em cada linha (ver seção 2.1.2). Em situações práticas, como não se conhece a matriz de covariâncias $\boldsymbol{\Sigma_0}$ do processo, usa-se a matriz $\mathbf{S} = (m-1)^{-1}\mathbf{X}'\mathbf{X}$ de covariâncias amostrais. Utilizando-se no lugar de $\mathbf{X}$ a matriz $\mathbf{X_d}$ de dados *padronizados*, conforme apresentado na seção 2.1.2, obtém-se, ao invés de $\mathbf{S}$, a matriz $\mathbf{S^{(cor)}}$ de correlacões amostrais. Como a matriz $\mathbf{S}$ é simétrica e não singular, existe uma matriz ortonormal $\mathbf{U}=[\mathbf{u}_1,\mathbf{u}_2,...,\mathbf{u}_p]$ que transforma a matriz $\mathbf{S}$ numa matriz diagonal

$$\mathbf{S}_C = \mathbf{U}'\mathbf{S}\mathbf{U} .$$

$\mathbf{S_C} = \text{diag}[\lambda_1,\lambda_2,...,\lambda_p]$, onde λ_i são os autovalores de S (Rencher, 1995). A matriz $\mathbf{S_C}$ apresenta as covariâncias amostrais das novas variáveis independentes $\boldsymbol{C}$ (covariância entre C_i e C_j é zero,

para $i \neq j$), denominadas *Componentes Principais*. Os autovalores de $\mathbf{S}_{\boldsymbol{C}}$ estão dispostos em ordem decrescente, com λ_i representando a variância do componente C_i.

Cada autovetor $\mathbf{u}_i$, $i = 1,...,p$, de $\mathbf{U}$, obtido a partir do autovalor λ_i, apresenta os coeficientes da $i^{ésima}$ combinação linear das variáveis iniciais $\boldsymbol{X}$. A matriz $\mathbf{U}$ é chamada de matriz das cargas, com cada vetor $\mathbf{u}_{\boldsymbol{i}}$ contendo as cargas devidas ao componente C_i. Desta forma, o vetor $\boldsymbol{C}$ de variáveis latentes pode ser escrito por

$$\boldsymbol{C} = \mathbf{U}' \boldsymbol{X},$$

O primeiro componente representa geometricamente a projeção das variáveis originais na direção de máxima variabilidade dos dados (representada por λ_1). Este componente é uma combinação linear das variáveis originais cujas cargas foram obtidas através do primeiro autovetor $\mathbf{u}_1$ de $\mathbf{U}$, condicionado a $\mathbf{u}_1'\mathbf{u}_1 = 1$ (vetor normalizado). O segundo componente representa a projeção das variáveis originais na segunda maior direção de variabilidade (ortogonal em relação à primeira direção) dos dados (representada por λ_2). O segundo autovetor $\mathbf{u}_2$ de $\mathbf{U}$ (com $\mathbf{u}_2'\mathbf{u}_2 = 1$ e $\mathbf{u}_1'\mathbf{u}_2 = 0$), contém as cargas utilizadas na segunda projeção. O $p^{ésimo}$ componente representa a projeção das variáveis originais na $p^{ésima}$ maior direção de variabilidade (ortogonal em relação às p-1 anteriores) dos dados (representada por λ_p). O $p^{ésimo}$ autovetor $\mathbf{u}_p$ de $\mathbf{U}$ (com $\mathbf{u}_p'\mathbf{u}_p = 1$ e $\mathbf{u}_i'\mathbf{u}_j = 0$ para $i \neq j$) contém as cargas utilizadas na $p^{ésima}$ projeção.

Dada a matriz de dados $\mathbf{X}$, o vetor $\mathbf{x}_k'$ referente a uma observação pode ser transformado no vetor de escores $\mathbf{c}_k'$ fazendo

$$\mathbf{c}_k = \mathbf{U}'\mathbf{x}_k, \quad k = 1,...,m .$$

O escore devido à $i^{ésima}$ variável latente referente à observação $\mathbf{x}_k'$ é obtido fazendo

$$c_{i,k} = \mathbf{u}_i'\mathbf{x}_k, \quad k = 1,...,m .$$

Obtêm-se os escores relativos ao $i^{ésimo}$ Componente Principal através da seguinte expressão:

$$\mathbf{c}_i = \mathbf{X}\mathbf{u}_i, \quad i = 1,2,...,p.$$

A ACP decompõe a matriz **X** de observações da seguinte maneira:

$$\mathbf{X} = \sum_{i=1}^{p} \mathbf{c}_i \mathbf{u}_i' \ .$$

Na prática, raramente existe a necessidade de considerar todos os p autovalores λ_i, visto que, devido à forte correlação entre as variáveis originais, poucos componentes costumam capturar a maior parte da variabilidade dos dados. Jackson (1991) apresenta algumas estratégias para escolha do número de componentes necessários na obtenção de uma descrição eficiente dos dados. Dentre as estratégias apresentadas, o procedimento de *validação cruzada* (*cross validation*; Wold, 1978) parece ser o mais seguro; este procedimento será descrito no Capítulo 3 desta dissertação. Para q ($< p$) componentes retidos (segundo algum critério pré-determinado), a matriz **X** é aproximada pelo modelo

$$\hat{\mathbf{X}} = \sum_{i=1}^{q} \mathbf{c}_i \mathbf{u}_i' \ ,$$

onde

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}$$

representa a matriz $m \times p$ dos resíduos referentes a cada uma das observações reescritas em função dos q componentes retidos no modelo.

Monitora-se o processo através de amostras futuras, projetando cada vetor $\mathbf{x}_f'$ futuro, p dimensional, no plano definido pelos q autovetores retidos, obtendo os escores

$$c_{i,f} = \mathbf{u}_i' \mathbf{x}_f \ , \quad \text{para} \quad i = 1,...,q \ ,$$

e os resíduos

$$\mathbf{e}_f' = \mathbf{x}_f' - \hat{\mathbf{x}}_f' \ ,$$

onde $\hat{\mathbf{x}}_f = \mathbf{U}_q \mathbf{c}_{q,f}$, e $\mathbf{c}_{q,f}$ é o vetor, $q \times 1$, de escores gerados do modelo e $\mathbf{U}_q$ é a matriz, $p \times q$, das cargas referentes aos q autovetores retidos. Utiliza-se a carta multivariada baseada na

estatística de Hotelling apresentada na expressão (25), em função dos q CPs retidos. Escreve-se então:

$$T_{q,f}^2 = \sum_{i=1}^{q} \frac{c_{i,f}^2}{\lambda_i}, \qquad (26)$$

onde λ_i representa a variância de C_i . A estatística T_q^2 baseada nos q primeiros componentes retidos é utilizada para monitorar desvios nas variáveis de processo que exercem maior contribuição na variação apresentada em **X** (MacGregor & Kourti, 1995). Se $T_{q,f}^2 >$ LSC [(limite de controle obtido através da expressão (14)], então algumas dessas variáveis de processo apresentaram desvios significativos em relação ao seus padrões (provável presença de uma causa especial). Analisando os respectivos escores $c_{i,f}$, referentes aos q Componentes Principais C_i retidos, pode-se, através da utilização de Gráficos de Contribuição (descritos no Capítulo 3 deste trabalho), determinar quais variáveis são efetivamente responsáveis por sinais apresentados na carta de controle.

A possível ocorrência de um evento que não está presente na distribuição de referência (obtida a partir da matriz **X**) provocará alterações na estrutura de covariância das variáveis de processo. Tais eventos podem ser detectados calculando o erro quadrático de predição referente aos resíduos gerados através do vetor $\mathbf{x}'_f$ (MacGregor & Kourti, 1995). Escreve-se:

$$EQP = \sum_{i=1}^{p} (x_{i,f} - \hat{x}_{i,f})^2 \qquad (27)$$

A estatística na expressão (27), também designada Q por Jackson (1991), representa o quadrado da distância perpendicular entre o vetor $\mathbf{x}'_f$ e o plano definido pelos q componentes retidos. Se o processo está sob controle, então o valor de Q não deve ultrapassar certos limites. Quando apenas causas comuns estão presentes (isto é, apenas eventos representados pela distribuição de referência), Q representa flutuações aleatórias ocorridas no processo. Em contrapartida, quando um evento não usual ocorre e altera a estrutura de covariância apresentada em **X** (provável causa especial de variação presente), um valor alto de Q_f é obtido. Isso significa que a observação $\mathbf{x}'_f$ não foi bem representada pelo modelo quando projetada no plano q dimensional. Os limites

de controle usados na estatística Q são calculados, a partir da matriz $\mathbf{X}$ dos dados de referência, utilizando resultados aproximados da distribuição das formas quadráticas [Jackson & Mudholkar (1979)]; e serão apresentados no Capítulo 3 deste trabalho.

O monitoramento efetivo das variáveis de processo é feito, assim, utilizando-se uma carta de Hotelling a partir do q Componentes Principais retidos, a carta Q para monitorar os resíduos do modelo [expressões (26) e (27), respectivamente] e os Gráficos de Contribuição, para diagnosticar um sinal de processo fora de controle gerado a partir das estatísticas $T^2_{q,f}$ e/ou Q_f.

2.3.2- Carta de controle baseada na projeção de dados via PEL

As cartas de controle via ACP, descritas na seção 2.3.1, podem ser utilizadas para monitorar processos, reduzindo a dimensão das variáveis $\boldsymbol{X}$ de processo ou das variáveis $\boldsymbol{Y}$ de qualidade final do produto. O procedimento PEL consiste na redução simultânea das dimensões de $\boldsymbol{X}$ e $\boldsymbol{Y}$, buscando o monitoramento de $q(< p)$ variáveis latentes que explicam grande parte da covariância entre $\boldsymbol{X}$ e $\boldsymbol{Y}$. Dadas uma matriz $\mathbf{X}$ ($m \times p$), de dados do processo, e uma matriz $\mathbf{Y}$ ($m \times b$), de dados de qualidade do produto, deseja-se obter um grupo reduzido de variáveis latentes que expliquem não apenas a variação em $\mathbf{X}$, ou em $\mathbf{Y}$, mais a variação em $\mathbf{X}$ que melhor representa os dados de qualidade $\mathbf{Y}$, (MacGregor & Kourti, 1995).

Através da versão mais usual do método PEL, devida a Hoskuldsson (1988), obtém-se, a partir da matriz $\mathbf{X'YY'X}$ de covariâncias amostrais (esta matriz apresenta simultaneamente as covariâncias das variáveis $\boldsymbol{X}$ e $\boldsymbol{Y}$ e entre as variáveis $\boldsymbol{X}$ e $\boldsymbol{Y}$), a primeira variável latente através da expressão:

$$C_1 = \mathbf{w}_1' \boldsymbol{X},$$

onde C_1 é a combinação linear das variáveis $\boldsymbol{X}$ que maximiza a covariância entre a matriz $\mathbf{X}$ e a matriz $\mathbf{Y}$ de observações. O vetor $\mathbf{w}_1$ representa o primeiro autovetor da matriz $\mathbf{X'YY'X}$ de

covariâncias amostrais. Os escores referentes à primeira variável latente são obtidos, para cada um das *m* observações preliminares $\mathbf{x}'_k$ de **X**, da seguinte forma:

$$\mathbf{c}_1 = \mathbf{X}\,\mathbf{w}_{1.}$$

Uma vez gerado o vetor $\mathbf{c}_1$ (vetor, de dimensão $m \times 1$, contendo os escores devidos à primeira variável latente, referentes às *m* observações), regridem-se as colunas da matriz **X** (*m* observações de cada variável X_i, i=1,...,p) em função de $\mathbf{c}_1$, obtendo-se o vetor dos coeficientes da regressão

$$\mathbf{u}_1 = \frac{\mathbf{X}'\mathbf{c}_1}{\mathbf{c}'_1\mathbf{c}_1}.$$

Deflaciona-se então a matriz **X** para obter os resíduos

$$\mathbf{X}_2 = \mathbf{X} - \mathbf{c}_1\mathbf{u}'_1.$$

A matriz **Y** também é deflacionada resultando em

$$\mathbf{Y}_2 = \mathbf{Y} - \mathbf{c}_1\mathbf{v}'_1,$$

onde $\mathbf{v}_1$ representa o vetor dos coeficientes da regressão das colunas da matriz **Y** (*m* observações de cada variável Y_i, i=1,...,b) em função de $\mathbf{c}_1$, dado por:

$$\mathbf{v}_1 = \frac{\mathbf{c}'_1\mathbf{Y}}{\mathbf{c}'_1\mathbf{c}_1}.$$

A segunda variável latente é obtida através de

$$C_2 = \mathbf{w}'_2\boldsymbol{X},$$

onde $\mathbf{w}_2$ representa o primeiro autovetor da matriz $\mathbf{X}'_2\mathbf{Y}_2\mathbf{Y}'_2\mathbf{X}_2$. Uma vez gerado o vetor $\mathbf{c}_2$, regridem-se as colunas da matriz $\mathbf{X}_2$ e da matriz $\mathbf{Y}_2$ em função de $\mathbf{c}_2$, obtendo-se os respectivos vetores $\mathbf{u}_2$ e $\mathbf{v}_2$ dos coeficientes das duas regressões. Deflacionando as matrizes $\mathbf{X}_2$ e $\mathbf{Y}_2$, obtêm-se os respectivos resíduos

$$\mathbf{X}_3 = \mathbf{X} - \mathbf{c}_1\mathbf{u}'_1 - \mathbf{c}_2\mathbf{u}'_2 = \mathbf{X}_2 - \mathbf{c}_2\mathbf{u}'_2 \quad \text{e} \quad \mathbf{Y}_2 = \mathbf{Y} - \mathbf{c}_1\mathbf{v}'_1 - \mathbf{c}_2\mathbf{v}'_2 = \mathbf{Y}_2 - \mathbf{c}_2\mathbf{v}'_2.$$

O processo pode ser continuado até a obtenção do $p^{ésimo}$ autovetor. A exemplo do modelo ACP, as variáveis latentes (C_1, C_2,...) são independentes e os autovetores ($\mathbf{w}_1$, $\mathbf{w}_2$,...) são ortogonais entre si.

A exemplo da ACP, raramente existe a necessidade de considerar todos os p autovalores λ_i na PEL, visto que, devido à forte correlação entre as variáveis originais $\boldsymbol{X}$ e $\boldsymbol{Y}$, poucas variáveis latentes costumam capturar a maior parte da variabilidade dos dados. Através do procedimento de *validação cruzada* (*cross validation*, Wold, 1978), pode-se determinar um número $q(< p)$ suficiente de variáveis latentes que serão utilizadas no modelo.

As cartas de controle baseadas na projeção de dados via PEL adotam o esquema de monitoramento semelhante ao esquema utilizado nas cartas via ACP. O processo é monitorado através de amostras futuras, projetando cada observação $\mathbf{x}'_f$, de dimensão p, no plano definido pelos q autovetores retidos, obtendo recursivamente os respectivos escores $c_{i,f}$'s (usando $\mathbf{w}_i$'s, $\mathbf{u}_i$'s e $\mathbf{v}_i$'s, para i=1,...,q), calculados através do $i^{ésimo}$ deflacionamento de $\mathbf{x}'_f$. Utilizando a estatística de Hotelling decomposta na expressão (26), em função de q variáveis latentes retidas, monitora-se a variação das variáveis do processo que causam maior impacto nas variáveis de qualidade (Kourti & MacGregor, 1996). Através da estatística Q_f [(expressão (27)], monitoram-se os resíduos referentes à estimativa $\hat{\mathbf{x}}_f = \mathbf{U}_q \mathbf{c}_{q,f}$ (observação $\mathbf{x}'_f$ reescrita através do modelo PEL de referência, construído utilizando q variáveis latentes), com o objetivo de verificar possíveis alterações na estrutura de covariância do processo. Finalmente, utilizam-se os Gráficos de Contribuição para diagnosticar um possível sinal de processo fora de controle gerado através das estatísticas $T^2_{q,f}$ e/ou Q_f. A base teórica do método, bem como aplicações da PEL no monitoramento de processos contínuos, pode ser encontradas em Lorber *et al.* (1987), Geladi & Kowalski (1986a), Kresta *et al.* (1991), MacGregor & Kourti (1995), Kourti & MacGregor (1996), entre outros.

2.3.2.1- Projeção de dados via PEL usando múltiplos blocos

Os métodos de projeção de dados constituem-se em boa alternativa em relação aos métodos usuais de controle estatístico de processos, sempre que existir uma grande massa de dados coletados dos processos em estudo. No entanto as cartas de controle e de diagnóstico via métodos de projeção, discutidas acima, ainda podem ser de difícil interpretação quando um número grande de variáveis de processo estão envolvidas. Em alguns processos, pode-se decompor as variáveis de processo $\boldsymbol{X}$ em blocos $\boldsymbol{X_1}$,$\boldsymbol{X_2}$,..., onde em cada bloco, que representa uma parte (uma seção de produção) do processo, existe um grupo peculiar de eventos subjacentes ocorrendo quando o processo está sob controle.

A PEL-MB não equivale a realizar a PEL, separadamente, em cada bloco. Na PEL-MB, os blocos são analisados conjuntamente e ponderados de forma a maximizar sua covariância com o conjunto de variáveis $\boldsymbol{Y}$. Desta forma, dado um vetor $\mathbf{x}'_f$, obtêm-se para cada bloco os respectivos valores $T^2_{q,f}$ (calculados em função das q variáveis latentes C_i retiradas) e Q_f. Se algum evento especial estiver presente em algum bloco, os valores $T^2_{q,f}$ e Q_f, obtidos a partir das variáveis daquele bloco, devem ser afetados. Através de Gráficos de Contribuição, pode-se diagnosticar possíveis causas especiais associadas a eventos anômalos. Geladi & Kowalski (1986b), Kourti *et al.* (1996), Kourti & MacGregor (1996), MacGregor & Kourti (1996), Macgregor *et al.* (1994), entre outros, apresentam a base teórica do método, bem como aplicações da PEL-MB no monitoramento de processos contínuos e por bateladas.

2.3.3- Métodos não lineares de projeção de dados

Na maioria dos casos, os modelos lineares apresentados descrevem, satisfatoriamente, as flutuações em torno dos eventos comuns presentes no processo (causas comuns de variação). Nos

processos em bateladas (foco principal desta dissertação), embora o comportamento das variáveis seja não linear, através das abordagens ACPM (Análise de Componentes Principais Multidirecionais) ou PELM (Projeção de Estruturas Latentes Multidirecionais), a trajetória não linear das variáveis de processo é removida e os desvios de cada variável em relação a suas trajetórias médias são modelados via modelos lineares (ACP ou PEL). Entretanto existem processos industriais complexos com fortes características não lineares que não podem ser modificados e monitorados via métodos lineares. Para tratar deste problema, técnicas não lineares vêm sendo desenvolvidas, incluindo PCA não linear, PLS não linear, análise de correlação não linear e procedimentos envolvendo redes neurais. Martin *et al.* (1996), Martin & Morris (1996), entre outros, fazem uma discussão acerca dessas abordagens.

2.4- Considerações finais

Nas seções 2.3.1 e 2.3.2, foram apresentados de maneira resumida a base teórica do monitoramento de processos multivariados contínuos, através dos métodos de projeção de dados via ACP e PEL, respectivamente. Salienta-se que, num processo contínuo, a trajetória das variáveis é constante (média ou valor alvo de cada variável constante) se o processo está sob controle estatístico, ou seja, apenas pequenas flutuações em torno da média de cada variável ocorrem. Neste caso, obtém-se a distribuição de referência para monitorar o processo a partir de *m* amostras preliminares independentes, que representem estas variações aleatórias (devidas a causas comuns). No capítulo 3, será apresentada detalhadamente a base teórica da Análise de Componentes Principais, assim como da carta de controle via ACP para monitoramento de processos multivariados contínuos. Utilizando um exemplo simulado, será feito o monitoramento de um processo multivariado através da carta de controle proposta.

O foco de interesse deste trabalho são as cartas de controle para processos em bateladas, via utilização da abordagem ACP. Num processo em batelada, as médias das variáveis não são constantes, isto é, se o processo está sob controle as variáveis devem apresentar pequenas

flutuações em torno de suas trajetórias médias, dentro de cada batelada. Desta forma, a distribuição de referência deve ser obtida considerando não apenas a estrutura de correlação das variáveis do processo (com acontece nos processos contínuos), mas também a autocorrelação das variáveis durante a batelada. No capítulo 4, será detalhada a filosofia de controle dos processos multivariados em bateladas através da *Análise de Componentes Principais Multidirecionais*; ACPM (*Multiway Principal Component Analysis*; *MPCA*). No capítulo 5, a partir de um exemplo simulado, será construída uma carta de controle multivariada, utilizando a abordagem ACPM. Também será feito o monitoramento em tempo real do processo multivariado proposto através da carta obtida.

CAPÍTULO 3

3- COMPONENTES PRINCIPAIS

A *Análise de Componentes Principais* (ACP) é uma técnica de análise multivariada que visa transformar um grupo de variáveis correlacionadas em variáveis independentes denominadas Componentes Principais (CPs). Os CPs são combinações lineares das variáveis originais e representam suas projeções nas direções de máxima variabilidade dos dados.

3.1- Base teórica

Considere **X** uma matriz de dados *centrados*, $m \times p$, onde cada linha contém uma observação independente p-variada, representada por um vetor $\mathbf{x}_k'$, $1 \times p$. A estrutura de covariância destes dados é dada pela matriz de covariâncias amostral **S**, $p \times p$, obtida a partir de m amostras preliminares, é dada por

$$\mathbf{S} = \begin{bmatrix} s_1^2 & s_{12} & \ldots & s_{1p} \\ s_{21} & s_2^2 & \ldots & s_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ s_{p1} & s_{p2} & \ldots & s_p^2 \end{bmatrix}.$$

Na matriz **S**, s_i^2 representa a variância da $i^{ésima}$ variável (i=1,...,p) e s_{ij} representa a covariância entre a $i^{ésima}$ e a $j^{ésima}$ variável (i,j=1,...,p e $i \neq j$). Como as variáveis estão correlacionadas, $s_{ij} \neq 0$.

A ACP foi proposta por Hotelling e baseia-se num resultado da teoria de álgebra linear. Como a matriz de covariâncias é simétrica e não singular ($|\mathbf{S}| \neq 0$), existe uma matriz **U** ortonormal que diagonaliza **S**. Então pode-se escrever:

$$\mathbf{U}'\mathbf{SU} = \mathbf{S}_c, \tag{28}$$

onde

$$\mathbf{S}_c = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_p \end{bmatrix}.$$

$\mathbf{S}_c$ é uma matriz diagonal cujos elementos λ_1,..., λ_p representam os autovalores de **S**. As colunas da matriz **U**, $\mathbf{u}_1$,..., $\mathbf{u}_p$ representam os autovetores de **S**. Os autovalores são obtidos através da solução da seguinte equação (Rencher, 1995):

$$|\mathbf{S} - \lambda \mathbf{I}_p| = 0, \tag{29}$$

onde $\mathbf{I}_p$ é uma matriz identidade de ordem $p \times p$. A equação (29) é denominada *equação característica* da matriz **S**. Os autovetores de **S** são obtidos através da solução da seguinte equação:

$$[\mathbf{S} - \lambda \mathbf{I}_p]\,\mathbf{f}_i = 0, \tag{30}$$

tal que

$$\mathbf{u}_i = \frac{\mathbf{f}_i}{\sqrt{\mathbf{f}_i'\mathbf{f}_i}} \quad , \quad \text{para} \quad i = 1,.....,p \; , \tag{31}$$

representa os autovetores normalizados.

Dessa forma, a partir da expressão (28) pode-se transformar p variáveis correlacionadas $X_1,.....,X_p$ em p variáveis independentes $C_1,....,C_p$, denominadas Componentes Principais. Os autovetores $\mathbf{u}_i$ descrevem os eixos de coordenadas das novas variáveis. Estes vetores são combinações lineares das variáveis originais, projetando-as nas direções de máxima variabilidade em relação à nuvem de pontos p-dimensional, obtida a partir dos dados (Johnson & Wichern, 1992).

O vetor $\mathbf{x}_k'$, referente a $k^{ésima}$ observação das p variáveis, pode ser transformado em um vetor de escores $\mathbf{c}_k'$, obtidos através dos CPs, fazendo:

$$\mathbf{c}_k = \mathbf{U}'\mathbf{x}_k \, , \quad k = 1,...,m \; .$$

O escore devido a $i^{ésima}$ variável latente (C_i), referente a observação $\mathbf{x}_k'$, é obtido fazendo

$$c_{i,k} = \mathbf{u}_i'\mathbf{x}_k \, , \quad k = 1,...,m \; , \tag{32}$$

onde C_i possui média zero e variância λ_i, que representa o $i^{ésimo}$ autovalor de **S** (Gnanadesikan, 1997).

Obtêm-se os m escores relativos ao $i^{ésimo}$ Componente Principal através da seguinte expressão:

$$\mathbf{c}_i = \mathbf{X}\mathbf{u}_i \; , \quad i = 1,2,...,p.$$

Pode-se escrever a matriz **X** de observações em função dos componentes C_i, fazendo:

$$\mathbf{X} = \sum_{i=1}^{p} \mathbf{c}_i \mathbf{u}_i' \; . \tag{33}$$

Considere-se o seguinte exemplo introdutório bivariado (p=2), apresentado por Jackson (1980). Suponha-se que a concentração de um componente químico numa solução esteja sendo

monitorada através de dois diferentes métodos. Admite-se que o primeiro método seja o procedimento padrão e que o segundo esteja sendo utilizado para ratificar os resultados do primeiro, gerando assim uma maior confiabilidade nestes resultados. Obtêm-se 15 amostras da produção; cada amostra representa uma medição da concentração pelos dois métodos. A Tab. 3 apresenta os dados centrados (ou seja, com média zero) referentes aos resultados das 15 amostras.

Tabela 3. Resultado das 15 amostras preliminares $\mathbf{x}'_k$, escores $c_{i,k}$ (i=1,2) obtidos através da expressão (32) e referentes escores padronizados (ver seção 3.3). As amostras *16*, *17* e *18* representam observações futuras retiradas deste processo. Exemplo adaptado proposto por Jackson (1980).

AMOSTRA (k)	MÉTODO 1 (X_1)	MÉTODO 2 (X_2)	ESCORE (c_1)	ESCORE (c_2)	ESCORE (y_1)	ESCORE (y_2)
1	0,0	0,7	0,48	0,51	0,40	1,72
2	0,4	-0,2	0,15	-0,42	0,13	-1,43
3	-0,3	0,0	-0,22	0,21	-0,18	0,70
4	-0,3	0,1	-0,15	0,28	-0,12	0,95
5	1,7	1,5	2,27	-0,09	1,88	-0,30
6	1,0	0,8	1,28	-0,11	1,06	-0,38
7	-1,3	-1,2	-1,77	0,03	-1,47	0,10
8	-0,5	-0,7	-0,84	-0,16	-0,70	-0,55
9	0,1	-0,6	-0,34	-0,50	-0,28	-1,71
10	-0,4	-0,4	-0,57	-0,01	-0,47	-0,05
11	0,5	0,4	0,64	-0,06	0,53	-0,19

Tabela 3. Resultado das 15 amostras preliminares $\mathbf{x}'_k$, escores $c_{i,k}$ (i=1,2) obtidos através da expressão (32) e referentes escores padronizados (ver seção 3.3). As amostras *16*, *17* e *18* representam observações futuras retiradas deste processo. Exemplo adaptado proposto por Jackson (1980). (Continuação).

AMOSTRA (k)	MÉTODO 1 (X_1)	MÉTODO 2 (X_2)	ESCORE (c_1)	ESCORE (c_2)	ESCORE (y_1)	ESCORE (y_2)
12	-0,8	-1,0	-1,27	-0,17	-1,06	-0,58
13	1,3	1,6	2,04	0,26	1,70	0,89
14	0,1	-0,2	-0,07	-0,21	-0,05	-0,73
15	-1,5	-0,8	-1,64	0,46	-1,36	1,55
16	2,3	2,5	3,39	0,22	2,82	0,75
17	1,0	-1,0	0,03	-1,41	0,03	-4,81
18	-2,7	-0,9	-2,57	1,21	-2,14	4,12

A Fig. 7 apresenta um diagrama de dispersão dos dados, referentes às 15 mensurações geradas pelos dois métodos. Observa-se que estes resultados são correlacionados (o coeficiente de correlação amostral obtido foi r=0,887). Isto acontece porque X_1 e X_2 representam mensurações de concentração, realizadas através de dois métodos diferentes, de uma mesma unidade amostral (mesma amostra de solução).

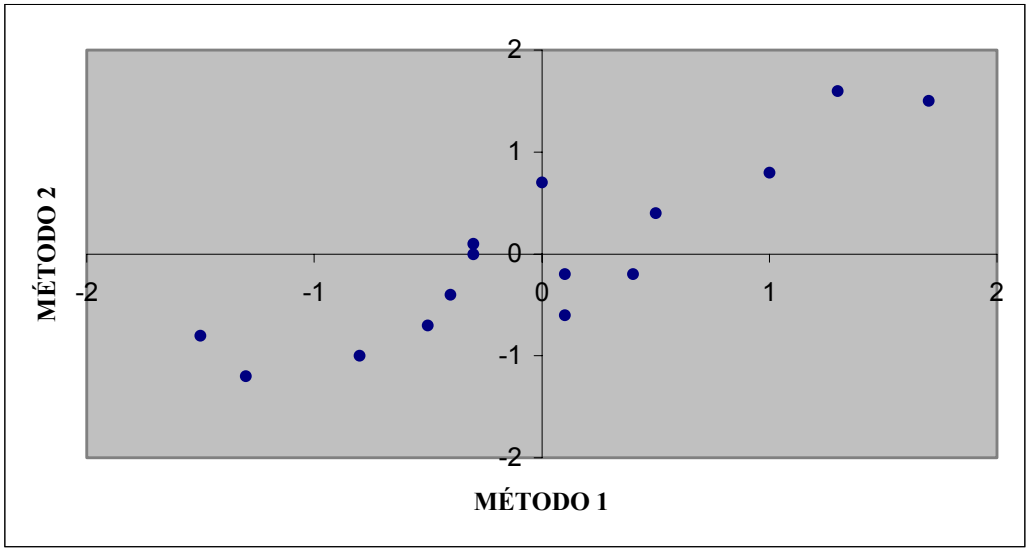

Figura 7. Diagrama de dispersão dos dados das análises químicas.

Antes de aplicar a ACP, necessitam-se das variâncias e a covariância entre as variáveis X_1 e X_2. O vetor de médias amostrais e a matriz de covariâncias dos dados são, respectivamente:

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{S} = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix} = \begin{bmatrix} 0{,}7986 & 0{,}6793 \\ 0{,}6793 & 0{,}7343 \end{bmatrix}$$

Através das expressões (29), (30) e (31), pode-se obter os autovalores e os respectivos autovetores associados à matriz **S**. Para o exemplo, tem-se:

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1{,}4465 \\ 0{,}0864 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = [\mathbf{u}_1 \mid \mathbf{u}_2] = \begin{bmatrix} 0{,}7236 & -0{,}6902 \\ 0{,}6902 & 0{,}7236 \end{bmatrix}$$

Através da expressão (28), diagonaliza-se a matriz de covariâncias **S**, fazendo:

$$\mathbf{U}'\mathbf{S}\mathbf{U} = \begin{bmatrix} 0{,}7236 & 0{,}6902 \\ -0{,}6902 & 0{,}7236 \end{bmatrix} \begin{bmatrix} 0{,}7986 & 0{,}6793 \\ 0{,}6793 & 0{,}7343 \end{bmatrix} \begin{bmatrix} 0{,}7236 & -0{,}6902 \\ 0{,}6902 & 0{,}7236 \end{bmatrix}$$

$$= \begin{bmatrix} 1{,}4465 & 0 \\ 0 & 0{,}0864 \end{bmatrix} = \mathbf{S}_c \ ,$$

onde $\mathbf{S}_c$ representa a matriz de covariâncias referente às duas novas variáveis independentes C_i (Componentes Principais) obtidas a partir de **S**.

A Tab. 3 também apresenta os escores referentes aos dois CPs acima. Os escores devidos ao primeiro CP podem ser obtidos através da expressão (32). No exemplo:

$$c_{1,k} = [0{,}7236 \quad 0{,}6902]\mathbf{x}_k \ ,$$

onde $\mathbf{x}'_k$ é o vetor que representa uma amostra das medições de concentração geradas pelos dois métodos.

A primeira amostra destes dados é representada pelo seguinte vetor:

$$\mathbf{x}'_1 = [0 \quad 0{,}7]$$

O escore devido ao primeiro CP, referente a esta observação é obtido fazendo:

$$c_{1,1} = [0{,}7236 \quad 0{,}6902]\begin{bmatrix} 0 \\ 0{,}7 \end{bmatrix} = 0{,}48$$

Os escores referentes ao segundo CP também são obtidos através da expressão (32), sendo dados por:

$$c_{2,k} = [-0{,}6902 \quad 0{,}7236]\mathbf{x}_k$$

Para a primeira amostra, o escore obtido a partir do segundo CP é:

$$c_{2,1} = [-0{,}6902 \quad 0{,}7236]\begin{bmatrix} 0 \\ 0{,}7 \end{bmatrix} = 0{,}51$$

3.2- Interpretação Geométrica dos Componentes Principais

Através da Fig. 8, pode-se observar os novos eixos coordenados devidos aos dois CPs. Estes eixos representam a rotação do eixos originais (X_1,X_2), em relação às suas médias, na

direção de máxima variabilidade dos dados (isto é, onde as observações estão mais separadas). A direção dessas projeções é fornecida pelos respectivos autovetores $\mathbf{u}_1$ e $\mathbf{u}_2$.

O primeiro CP (linha vermelha) representa a projeção das variáveis originais na direção de maior variabilidade dos dados, obtida através de $\mathbf{u}_1$. O maior autovalor encontrado representa a variância estimada de C_1 (estimada a partir das 15 amostras preliminares na Tab. 3). Desta forma, representa a variância das novas observações $c_{1,k}$. Assim, $s^2_{C_1} = \lambda_1 = 1{,}4465$.

O segundo CP (linha amarela) representa a projeção das variáveis originais na segunda (e última, neste caso) maior direção de variabilidade dos dados, ortogonal em relação à primeira, obtida através de $\mathbf{u}_2$. O segundo autovalor encontrado representa a variância estimada de C_2. Dessa forma, representa a variância das novas observações $c_{2,k}$. Assim, $s^2_{C_2} = \lambda_2 = 0{,}0864$.

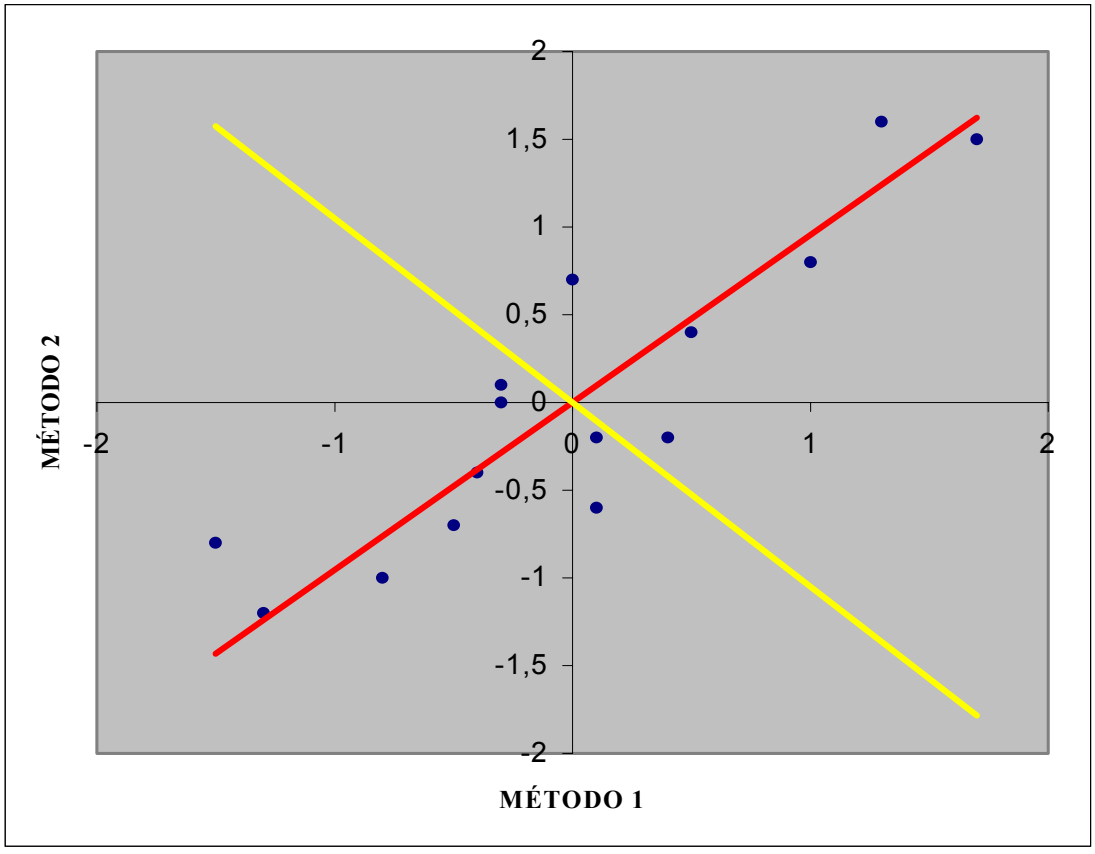

Figura 8. Representação geométrica dos Componentes Principais.

No caso bivariado, o primeiro CP sempre explica a variabilidade comum a X_1 e X_2, enquanto que o segundo CP explica a variabilidade devida às diferenças entre X_1 e X_2 (Jackson, 1991). Os coeficientes do primeiro autovetor $\mathbf{u}_1' = [0{,}7236 \quad 0{,}6902]$ são próximos e possuem mesmo sinal. Isso indica que o primeiro CP representa uma média ponderada das duas variáveis, refletindo, assim, as variações devidas ao processo. Os coeficientes do segundo autovetor $\mathbf{u}_2' = [-0{,}6902 \quad 0{,}7236]$ são próximos, mas possuem sinais trocados. O segundo CP representa, desta forma, as variações devidas às diferenças de mensuração entre os dois métodos. Assim, as duas fontes de variação existentes estão devidamente separadas e alocadas nos respectivos CPs.

Sabe-se que na análise multivariada (p>1), a estrutura de covariância de um processo é representada pela matriz de covariâncias $\boldsymbol{\Sigma_0}$ (quando não se conhece $\boldsymbol{\Sigma_0}$, usa-se **S** como estimativa). No entanto, em algumas situações, deseja-se representar a variabilidade total de um processo através de um valor. Uma medida de variabilidade usada, na prática, é dada pelo traço da matriz de covariâncias [tr(**S**)], ou seja, pela soma das variâncias das p variáveis (Rencher, 1995) . Para o exemplo, a variância total das variáveis originais pode ser determinada fazendo:

$$tr(\mathbf{S}) = s_1^2 + s_2^2 = 0{,}7986 + 0{,}7343 = 1{,}5329 .$$

Os CPs preservam a variância total das variáveis originais (Johnson & Wichern, 1992), o que pode ser constatado pela identidade abaixo:

$$tr(\mathbf{S}_c) = s_{C_1}^2 + s_{C_2}^2 = \lambda_1 + \lambda_2 = 1{,}4465 + 0{,}0864 = 1{,}5329 .$$

Como os CPs são independentes (isto é, $s_{C_1C_2} = 0$), a estrutura de covariância do processo (representada por estas novas variáveis) resume-se às variâncias λ_1 e λ_2 (variâncias dos CPs). Então o valor 1,4465 representa uma medida da variância comum aos dois métodos de mensuração (isto é, variação do processo), enquanto que o valor 0,0864 representa uma medida da variância devida às diferenças entre os dois métodos. Deve-se entender que o primeiro CP explica 94% (1,4465/1,5329) da variabilidade total, isto é, 94% da variabilidade do dados é devida às diferenças nos níveis de concentração obtidos nas amostras, independente do método de medição. O segundo CP explica 6% (0,0864/1,5329) da variabilidade total, isto é, 6% da variabilidade dos dados é devida às diferenças nas medições obtidas pelos dois métodos, em cada amostra.

3.3- Controle multivariado via Carta de Hotelling utilizando CPs

Quando se realiza o controle de um processo através de uma carta multivariada (carta de Hotelling tradicional, no caso), existe a dificuldade de diagnóstico das causas especiais, quando estas aparecem no processo. Sabe-se que quanto mais complexa a estrutura de covariâncias do processo (quanto mais variáveis envolvidas), maior a dificuldade de diagnóstico. Algumas estratégias de diagnóstico foram descritas no Capítulo 2 (seção 2.2.7). No entanto, mesmo quando o número de variáveis é pequeno (como é o caso do exemplo proposto na seção 3.2; p=2), existe alguma dificuldade de diagnóstico. Por exemplo, considere-se o monitoramento de um processo bivariado. Admita-se que a carta de Hotelling tenha gerado, a partir de uma nova observação coletada do processo, um sinal de processo fora de controle. Através de um procedimento de diagnóstico comum, procura-se encontrar a possível causa deste sinal observando o reflexo desta observação nas duas cartas individuais (cartas de Shewhart em relação às duas variáveis originais). No entanto, o fato de estas cartas não considerarem a correlação entre as duas variáveis envolvidas, pode dificultar uma rápida identificação da causa especial presente. Para resolver este problema, Jackson (1980) sugere a utilização da carta de Hotelling em conjunto com as cartas de Shewhart em relação aos Componentes Principais para fins de diagnóstico, pois estes consideram a estrutura de correlação das variáveis originais.

Considere-se novamente a situação apresentada no exemplo da seção 3.2, onde o nível de concentração de um componente químico numa solução está sendo monitorado através de dois métodos de medição. Considere-se também que as 15 amostras obtidas representem o padrão deste processo, apresentando apenas variabilidade devida a causas comuns (processo sob controle estatístico). A estatística de Hotelling apresentada na expressão (11) pode ser reecrita da seguinte forma:

$$T_k^2 = \frac{c_{1,k}^2}{\lambda_1} + \frac{c_{2,k}^2}{\lambda_2} = y_{1,k}^2 + y_{2,k}^2, \qquad (34)$$

onde $c_{i,k}$ é o escore devido à variável C_i (Componente Principal CP_i) , para i=1,2, referente à $k^{ésima}$ amostra.

Se o desvio-padrão $\sigma_{0,i}$ de C_i fosse conhecido, então esta variável teria distribuição Normal com média zero (pois as variáveis originais X_i estão centradas) e desvio-padrão $\sigma_{0,i}$; escreve-se $C_i \sim \boldsymbol{N}(0, \sigma_{0,i})$. Dessa forma, seria possível monitorar o $i^{ésimo}$ CP através da variável $Z_i (= C_i / \sigma_{0,i}) \sim \boldsymbol{N}(0,1)$, que representa a distância padronizada entre C_i e 0 (sua respectiva média, dado que o processo está sob controle estatístico); ver seção 2.2.1.

Na prática, o desvio-padrão $\sigma_{0,i}$ de C_i é estimado a partir das 15 amostras preliminares (m=15) através do desvio-padrão amostral $\sqrt{\lambda_i}$. Neste caso, C_i tem distribuição de Student com parâmetro m-1; simboliza-se $C_i \sim \boldsymbol{T}_{m\text{-}1}$. Pode-se então construir uma carta de Shewhart para monitorar a média de cada componente C_i com os seguintes limites:

$$\text{LSC} = + \boldsymbol{T}_{m\text{-}1,\alpha/2} \sqrt{\lambda_i} \ ,$$

$$\text{LC} = 0 \ ,$$

$$\text{LIC} = - \boldsymbol{T}_{m\text{-}1,\alpha/2} \sqrt{\lambda_i} \ ,$$

onde $\boldsymbol{T}_{m\text{-}1,\alpha/2}$ representa o percentil da distribuição de Student com m-1 graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada.

A média de C_i será efetivamente monitorada através da variável $Y_i = C_i / \sqrt{\lambda_i}$ (possui média zero e desvio-padrão unitário), que representa a distância padronizada entre C_i e 0. Para a variável padronizada Y_i, os limites de controle são:

$$\text{LSC} = \boldsymbol{T}_{m\text{-}1,\alpha/2} \ ,$$

$$\text{LC} = 0 \ ,$$

$$\text{LIC} = -\boldsymbol{T}_{m\text{-}1,\alpha/2} \ . \qquad (35)$$

A cada nova amostra f obtida do processo, os respectivos escores $y_{i,f}$ calculados serão comparados a estes limites de controle.

A média das variáveis originais X_i também serão monitoradas através da variável padronizada $Y_i = X_i / s_i$, onde s_i é o desvio-padrão estimado de X_i, obtido através das 15 amostras preliminares. Assim, dada uma nova amostra *f*, o respectivo escore $y_{i,f}$ será comparado com os mesmos limites de controle obtidos acima (a Tab. 3 apresenta os escores padronizados $y_{i,k}$ referentes às 15 observações preliminares).

No monitoramento de um processo multivariado através de ACP, deve-se seguir uma seqüência determinada de procedimentos (Jackson, 1991). Dada uma nova observação $\mathbf{x}'_f$ obtida do processo, deve-se (*i*) obter o valor T_f^2 [expressão (34)] calculado a partir dos escores padronizados $y_{i,f} (= c_{i,f} \Big/ \sqrt{\lambda_i})$; (*ii*) se T_f^2 ultrapassar o LSC [expressão (14)] estabelecido para a carta de controle, deve-se buscar um diagnóstico para o ponto fora de controle, verificando os escores $y_{i,f}$ através das respectivas cartas individuais, pois, como são independentes, cada CP representa um grupo específico de eventos do processo; e, finalmente, (*iii*) se o escore devido a algum CP ultrapassar o LSC, deve-se examinar as cartas individuais construídas a partir das variáveis originais. Se o valor de T_f^2 não ultrapassar o LSC, admite-se que o processo está sob controle estatístico. Mesmo assim, através dos escores $y_{i,f}$, pode-se detectar algum tipo de tendência nos dados, antevendo assim uma situação de processo fora de controle.

A Fig. 9 apresenta a carta de Hotelling com os resultados obtidos através das 15 amostras preliminares. As amostras *16*, *17, 18* apresentam escores superiores a LSC [$T_{16}^2 =$ 11,38, $T_{17}^2 =$ 23,14, $T_{18}^2 =$ 21,55; com α=5%, LSC=8,7 e representam futuras medições da concentração, obtidas a partir de dois métodos. Estas três situações simuladas servem para mostrar o poder de diagnóstico de causas especiais em processos multivariados monitorados via carta de Hotelling, com o auxílio de cartas individuais em relação aos Componentes Principais.

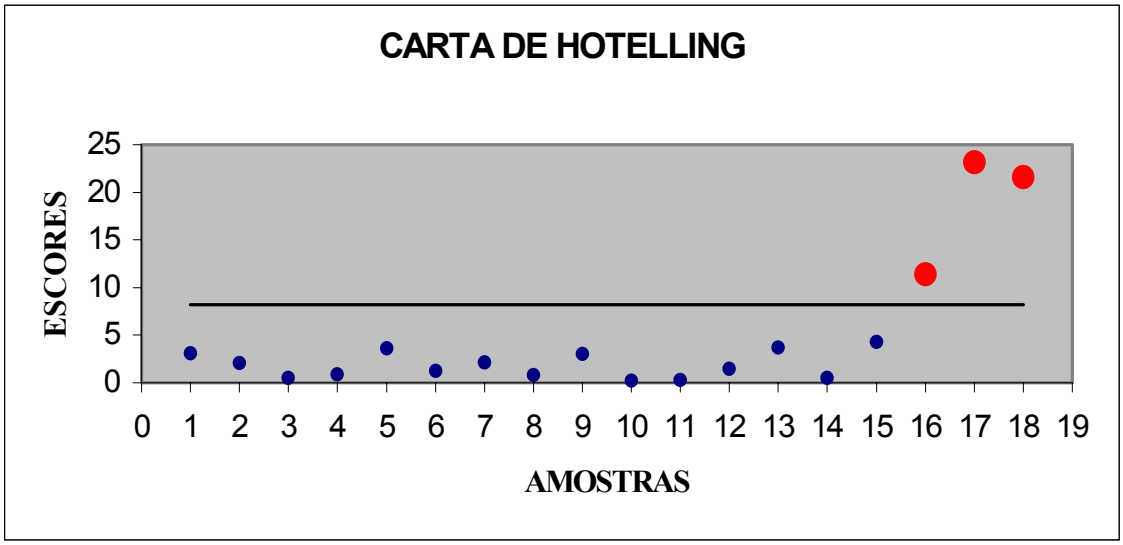

Figura 9. Carta de Hotelling em relação às amostras apresentadas na Tabela 3; significância de 5% (α=5%).

É importante ressaltar que o uso de α=5% (ao invés de α=0,1%, adotado nas cartas de controle mostradas no Capítulo 2), tanto para a carta de Hotelling quanto para as cartas individuais, se justifica pelo fato de se tratarem de amostras unitárias retiradas do processo. No exemplo, cada amostra representa uma medição de concentração (x_1 e x_2) fornecida por cada método e não uma média de n medições consecutivas ($\bar{x}_1$ e $\bar{x}_2$), obtidas de cada método. Desta forma, como se está trabalhando com a variabilidade máxima das variáveis envolvidas, diminuem-se os limites de controle para reduzir a probabilidade do erro tipo II (a probabilidade da carta não detectar uma efetiva mudança nas médias ou na estrutura de covariância das variáveis no processo).

Ressalta-se também que os limites LSC e LIC, obtidos para as cartas individuais (em relação aos CPs e às variáveis originais X_1 e X_2), são calculados usando $\alpha/2p$ ao invés de α/p. Esses limites de Bonferroni (ver seção 2.2.7.1) são usados para diminuir a chance de alarme falso (erro tipo I) no monitoramento de um processo multivariado, via p cartas univariadas. No exemplo proposto (p=2), dada a significância α=5% adotada (probabilidade de erro tipo I tolerada), trabalhando com os limites usuais, tem-se uma chance de aproximadamente 10% de que alguma das duas cartas individuais, em relação aos CPs, gere um alarme falso $[1-(0{,}95)^2]$, dado que as duas variáveis são independentes. Neste caso, para alcançar uma probabilidade de

5% de alarme falso, adota-se (5% / 2) = 2,5% [pois, $1-(0{,}975)^2 \cong 0{,}05$]. Apesar das variáveis originais não serem independentes, o uso dos limites de Bonferroni para as respectivas cartas individuais, de alguma forma, também mantém a chance de alarme falso em torno de 5%; ver Alt (1985).

As Fig. 10, 11, 12 e 13 representam, respectivamente, as cartas de controle individuais para monitorar as médias de C_1, C_2, X_1 e X_2. Os limites de controle inferior e superior (de Bonferroni) foram obtidos através da expressão (35), utilizando as 15 amostras preliminares, retiradas do processo sob controle estatístico. Então, para α=5%, LSC = $\boldsymbol{T}_{m\text{-}1,\alpha/2}$ = $\boldsymbol{T}_{14,0{,}025\,/\,2}$ = 2,51, e LIC = -2,51.

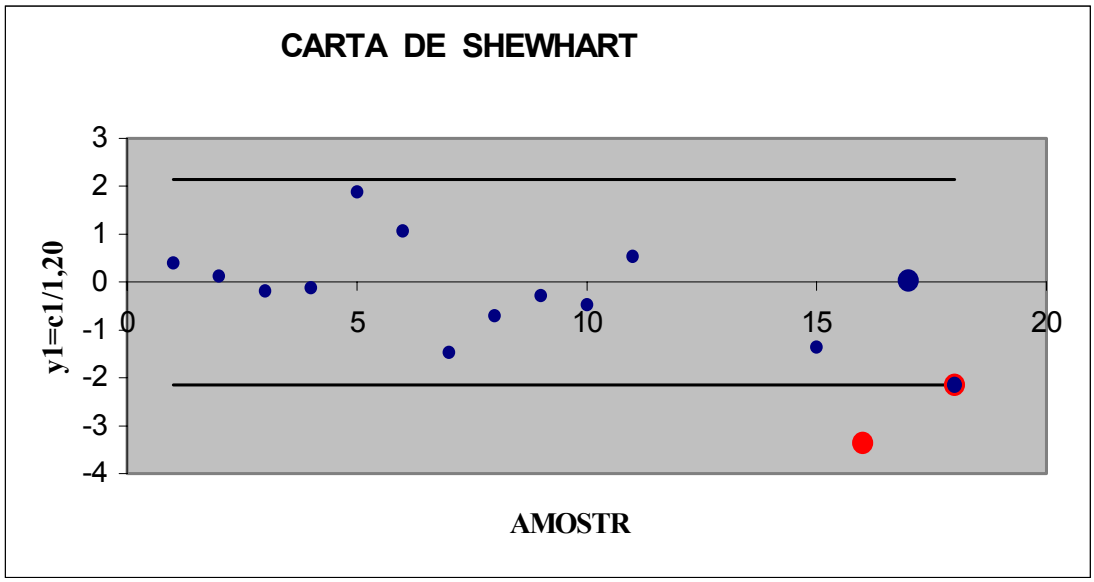

Figura 10. Carta de Shewhart para monitorar os escores padronizados referentes ao primeiro Componente Principal; significância de 5% (α=5%). Exemplo proposto por Jackson (1980).

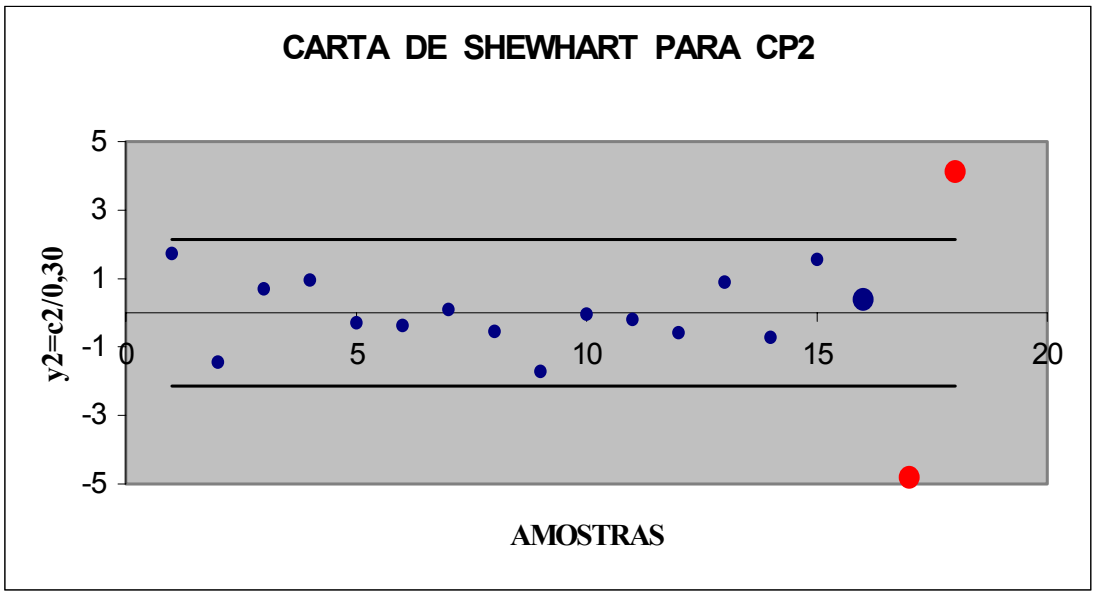

Figura 11. Carta de Shewhart para monitorar os escores padronizados referentes ao segundo Componente Principal; significância de 5% (α=5%). Exemplo proposto por Jackson (1980).

Observando a carta de Hotelling (Fig. 9), nota-se que a amostra 16 (primeiro ponto vermelho) apresenta um escore além dos limites de tolerância, configurando assim uma situação de processo fora de controle. Nesse processo, sabe-se que as variáveis C_1 (CP_1) e C_2 (CP_2) representam, respectivamente, a variabilidade devida ao processo e a variabilidade devido às diferenças de medição entre os dois métodos. Observando as cartas individuais em relação aos CPs (Fig. 10 e 11), nota-se que o devido escore obtido em relação ao CP_2 está dentro do padrão. Isso indica que não houve uma grande diferença entre as medições obtidas por cada método, nesta amostra. No entanto o escore obtido em relação ao CP_1 está fora dos limites de controle. Dessa forma, há evidências de que a concentração medida nesta amostra está efetivamente fora do padrão desejado, visto que este componente explica as variações nas medições obtidas das amostras, independente do método de medição. Essa conclusão pode ser ratificada observando os escores obtidos em relação às variáveis originais. Através das cartas individuais para X_1 e X_2 (Fig. 12 e 13) observa-se que, nesta amostra, a concentração medida pelos dois métodos está abaixo do limite mínimo tolerado.

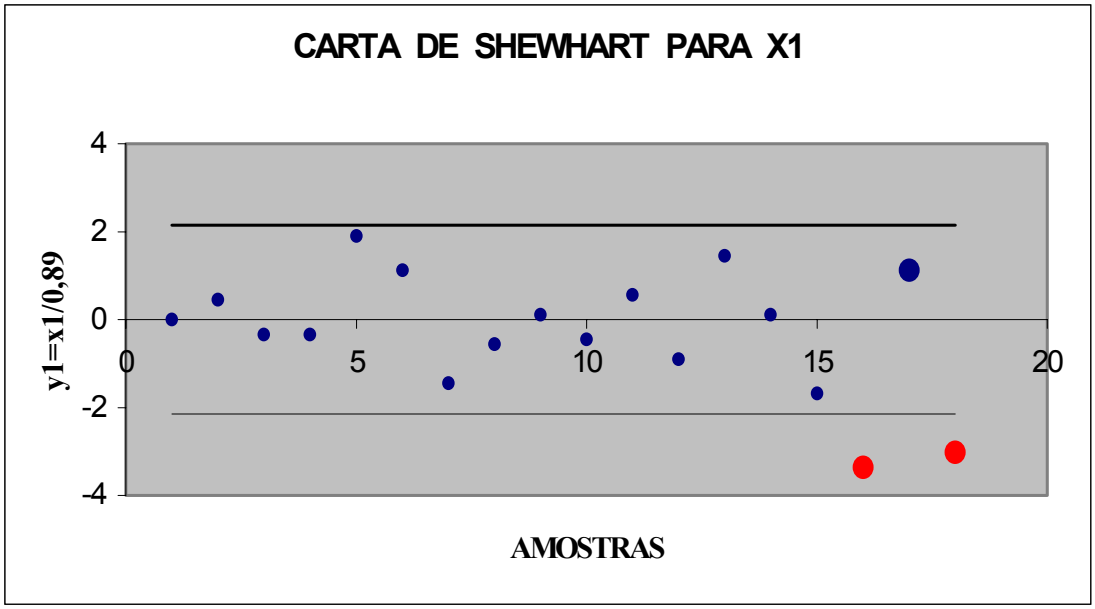

Figura 12. Carta de Shewhart para monitorar os escores padronizados referentes aos resultados obtidos através do primeiro método; significância de 5% (α=5%). Exemplo proposto por Jackson (1980).

O segundo ponto vermelho mostrado na carta de Hotelling representa o resultado de uma segunda amostra futura de medição da concentração retirada deste processo (está marcada como amostra *17*). Novamente tem-se um sinal de processo fora de controle. Observa-se que o escore referente ao CP_1 está dentro dos limites de tolerância, enquanto que o escore referente ao CP_2 está fora dos limites. Então, não há evidências de que os níveis de concentração obtidos nesta amostra estejam fora do padrão desejado, mas há evidências de que as respectivas medições geradas pelos dois métodos apresentam uma diferença significativa, sugerindo um provável erro de mensuração em algum dos métodos. Os escores referentes às variáveis originais X_1 e X_2 estão dentro dos limites de tolerância, ratificando assim a conclusão de que a concentração obtida nesta amostra está dentro do padrão.

A amostra *18* representa uma terceira amostra futura retirada deste processo. Mais uma vez, através da carta de Hotelling, tem-se um sinal de processo fora de controle. Observa-se que o escore referente ao CP_1, embora esteja muito próximo de LIC, está dentro dos limites de tolerância, enquanto que o escore referente ao CP_2 está fora dos limites. Pode-se concluir, a exemplo do que ocorreu na amostra *17*, que os dois métodos apresentaram resultados de concentração significativamente diferentes. Além disso, o resultado apresentado na carta para

CP_1 sugere, num primeiro momento que, embora ainda dentro dos limites de tolerância, os níveis de concentração podem estar começando a ficar abaixo do padrão desejado. Observando as cartas para X_1 e X_2 , nota-se que o primeiro método apresentou uma medição de concentração abaixo de LIC, enquanto que o segundo apresentou um resultado dentro dos limites tolerados. Há evidências de que nesta amostra existiu uma grande diferença entre os resultados obtidos pelos dois métodos (ver carta para CP_2), e que esta diferença deve ter sido causada por um erro de medição cometido pelo primeiro método (ver cartas para X_1 e X_2). Assim, pode-se entender que o escore referente ao CP_1 está muito próximo de LIC basicamente devido à influência do baixo nível de concentração apontado pelo primeiro método, e não porque o processo esteja efetivamente com um nível de concentração abaixo do desejado.

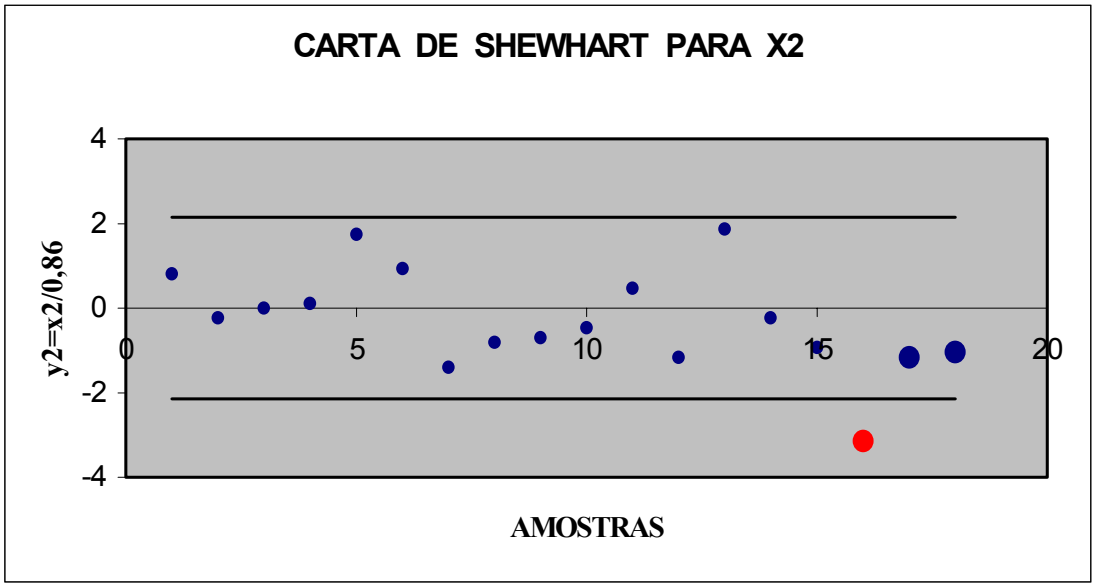

Figura 13. Carta de Shewhart para monitorar os escores padronizados referentes aos resultados obtidos através do segundo método; significância de 5% (α=5%). Exemplo proposto por Jackson (1980).

3.4 – Cartas de controle utilizando projeção de dados via CPs

Na seção anterior, mostrou-se a utilidade da ACP como ferramenta de auxílio para diagnóstico de causas especiais de variação. Mesmo com poucas variáveis envolvidas (no exemplo mostrado na seção anterior, p=2), através das cartas de controle via CPs, pode-se separar os diferentes tipos de variação presentes no processo, pois cada CP é reflexo de um grupo específico de causas de variação.

Considere-se agora um processo multivariado com um grande número (p) de variáveis correlacionadas sendo mensuradas periodicamente. Dado que o processo está sob controle, existe um determinado número de eventos comuns ocorrendo, sendo este número certamente pequeno em relação à quantidade de variáveis envolvidas. Neste caso, pode-se monitorar o processo usando Componentes Principais para reduzir a dimensão original dos dados, visto que, na verdade, a cada nova amostra de medições obtidas das p variáveis, tem-se uma diferente visualização dos mesmos eventos. Geralmente dois ou três CPs capturam grande parte da estrutura de correlação das variáveis originais, representando assim, de maneira satisfatória, os eventos comuns presentes no processo (MacGregor, 1995, MacGregor & Kourti, 1995, entre outros) .

Dadas m amostras independentes coletadas do processo sob controle estatístico, obtém-se novamente a matriz $\mathbf{X}$, $m \times p$, de dados centrados, onde cada linha contém uma observação p-variada, representada por um vetor $\mathbf{x}'_k$, $1 \times p$, para k=1,...,m. Através das expressões (29), (30) e (31) obtêm-se os p autovalores λ_i e os respectivos p autovetores $\mathbf{u}_i$. A expressão (33) apresenta a matriz dos $\mathbf{X}$ escrita em função dos p CPs. Esta expressão pode ser rescrita de seguinte forma (MacGregor & Kourti, 1995):

$$\mathbf{X} = \sum_{i=1}^{q} \mathbf{c}_i \mathbf{u}'_i + \sum_{i=q+1}^{p} \mathbf{c}_i \mathbf{u}'_i \ ,$$

onde $\mathbf{c}_i$ representa o vetor, $m \times 1$, dos escores referentes ao $i^{ésimo}$ CP, $\mathbf{u}_i$ representa o $i^{ésimo}$ autovetor, $p \times 1$, e q representa o número de CPs que serão utilizados no monitoramento do processo. Dessa forma, se os primeiros q CPs retidos oferecem uma caracterização satisfatória da

estrutura de correlação das p variáveis iniciais, a matriz $\mathbf{X}$ de dados é bem representada pelo modelo

$$\hat{\mathbf{X}} = \sum_{i=1}^{q} \mathbf{c}_i \mathbf{u}_i' + \mathbf{E}, \tag{36}$$

com

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}.$$

A matriz $\hat{\mathbf{X}}$ contém os dados originais estimados em função dos q CPs retidos, e a matriz $\mathbf{E}$ contém os resíduos obtidos do modelo.

No exemplo bivariado apresentado, todos os CPs foram utilizados no monitoramento do processo. No caso multivariado, a distribuição de referência das p variáveis originais é obtida a partir de k observações preliminares (contidas na matriz $\mathbf{X}$), considerando apenas q CPs. Dessa forma, monitora-se o processo através de futuras amostras, projetando a respectiva observação $\mathbf{x}_f'$ no plano definido pelos q CPs retidos, obtendo os escores

$$c_{i,f} = \mathbf{u}_i' \mathbf{x}_f, \quad \text{para} \quad i = 1,...,q. \tag{37}$$

Analogamente ao exemplo bivariado, utiliza-se a carta de Hotelling [(expressão (34)] juntamente com cartas individuais para os CPs para fins de diagnóstico. No caso multivariado, a estatística de Hotelling é decomposta em função dos q CPs retidos [expressão (26)]. Tem-se então:

$$T_{q,f}^2 = \sum_{i=1}^{q} \frac{c_{i,f}^2}{\lambda_i} = y_{1,f}^2 + ... + y_{q,f}^2, \tag{38}$$

onde λ_i representa a variância do CP_i. Na Primeira fase do controle, quando seleciona-se a distribuição de referência das variáveis no processo sob controle estatístico (seção 2.2.2.1), os limites de controle desta estatística são obtidos através da expressão (13), substituindo p por q. Assim,

$$\text{LSC} = \frac{(g-1)^2}{g} \times \frac{(q/(g-p-1))F(\alpha; q, (g-q-1)}{1+(q/(g-q-1))F(\alpha; q, g-q-1)} \quad \text{e}$$

$$\text{LIC} = 0. \tag{39}$$

Na Segunda fase do controle, quando a distribuição de referência foi obtida a partir das m ($\leq g$) amostras preliminares, os limites de controle são obtidos através da expressão (14), substituindo p por q. Tem-se então:

$$\text{LSC} = \frac{q(m+1)(m-1)}{m(m-q)} F_{\alpha,q,m-q} \quad \text{e}$$

$$\text{LIC} = 0. \tag{40}$$

É importante salientar que estes limites de controle são obtidos sob a suposição de que as variáveis (neste caso, os q CPs retidos) tenham distribuição Normal q-variada. Esta suposição é bastante razoável, pois o valor $T^2_{q,f}$ é obtido somando os devidos escores $c_{i,f}$ referentes aos q CPs utilizados. Como cada CP representa uma combinação linear das p variáveis originais, sabe-se, pelo *teorema do limite central* (Mood *et al*. 1974), que cada CP possui distribuição aproximadamente Normal. Mesmo que a combinação linear não envolva um número muito grande de variáveis de processo (situação desejada para obter-se uma rápida convergência em relação a distribuição Normal), ainda assim, pode-se admitir distribuição Normal para os CPs, pois dado que o processo está sob controle, as m amostras contém apenas flutuações aleatórias destas variáveis em torno de suas médias. Como cada um dos q CPs retidos representa uma soma de p variáveis iniciais, eles também apresentam flutuações aleatórias em torno das suas médias. Assim, os devidos escores $c_{i,f}$ gerados são bem descritos por uma distribuição Normal. Desta forma, os q CPs têm distribuição aproximadamente Normal q-variada com vetor de médias $\mathbf{0}$ (dados originais foram centrados) e matriz de covariâncias amostrais $\mathbf{S}$ (que é diagonal, pois os CPs são independentes), de dimensão $q \times q$.

Cada CP é monitorado através das cartas individuais, comparando o respectivo escore

$$y_{i,f} = \frac{c_{i,f}}{\sqrt{\lambda_i}}\ , \tag{41}$$

com $c_{i,f}$ obtido através da expressão (37), aos limites de controle apresentados na expressão (35).

Quando se utiliza um número de CPs inferior ao número de variáveis originais, busca-se a eliminação de redundâncias no processo. Estas redundâncias ocorrem pelo fato de existirem variáveis colineares no processo. Cada uma dessas variáveis é praticamente uma combinação linear das *p*-1 demais variáveis, trazendo, assim, pouca informação exclusiva sobre o processo. Na verdade, essas variáveis refletem muito pouco as principais fontes de variabilidade do processo (representam ruído), visto que duplicam informação já contida em outras variáveis. Dessa forma, através dos *q* primeiros CPs, monitoram-se os eventos que representam as maiores fontes de variabilidade. A partir das cartas individuais em relação aos CPs, podem-se detectar possíveis deslocamentos nas médias, bem como alterações nas variâncias das variáveis originais que exercem maior influência na variação total contida na matriz **X**. Utilizando a estatística de Hotelling em relação aos *q* primeiros CPs [expressão (38)], pode-se detectar se a variação das *p* variáveis de processo, na dimensão definida pelos *q* CPs retidos, é maior do que a tolerada, assumindo que apenas causas comuns de variação estão presentes.

A utilização dos *p* CPs no monitoramento pode, inclusive, gerar sinalizações erradas sobre o estado do processo (alarme falso); Kourti & MacGregor (1996). Sabe-se que os últimos CPs descrevem uma porção muito pequena da variação total de **X**, ou seja, possuem autovalores λ_i (variâncias) muito próximos de zero. Incluindo estes CPs na estatística de Hotelling, nota-se que pequenos desvios nos escores $y_{i,f}$ podem gerar um alto valor para $T^2_{q,f}$, pois esses escores são divididos por seus respectivos autovalores, muito próximos de zero.

Monitorar o processo via carta de Hotelling e cartas de Shewhart baseados nos primeiros *q* CPs, entretanto não é uma estratégia suficiente. Se um evento totalmente novo (ou seja, que não está na distribuição de referência do processo, construída a partir da matriz **X** de amostras preliminares) aparecer no processo, os escores $y_{i,f}$ referentes aos *q* CPs retidos poderão não ser afetados, visto que esse evento não refletirá necessariamente numa alteração na média ou na variabilidade de alguma variável de processo. No entanto a ocorrência deste novo evento causará, certamente, uma alteração na estrutura de correlação original dessas variáveis, capturada na distribuição de referência dos dados. Dessa forma, uma futura observação $\mathbf{x}'_f$ não será mais satisfatoriamente representada pelo modelo construído através dos primeiros *q* CPs, ou seja, um

resíduo relevante será gerado ao obter-se a estimativa $\hat{\mathbf{x}}'_f$ através do modelo apresentado na expressão (36); Martin & Morris (1996).

A possível presença de um novo evento no processo pode ser detectada através do monitoramento dos erros quadráticos de predição (*EQP*), gerados a partir de uma observação futura $\mathbf{x}'_f = [x_{1,f}, ..., x_{p,f}]$. Rescrevendo a expressão (27), tem-se:

$$EQP = \sum_{i=1}^{p} (x_{i,f} - \hat{x}_{i,f})^2 . \qquad (42)$$

A observação $\mathbf{x}'_f$ é estimada através do modelo $\hat{\mathbf{x}}_f = \mathbf{U}_q \mathbf{c}_{q,f}$, onde $\mathbf{c}_{q,f}$ é o vetor, $q \times 1$, de escores obtidos do modelo (cada um referente ao devido CP retido) e $\mathbf{U}_q = [\mathbf{u}_1 | \mathbf{u}_2 | ... | \mathbf{u}_q]$ é a matriz, $p \times q$, dos autovetores ou das cargas referentes aos q CPs retidos.

Na prática, como o monitoramento é realizado através dos escores padronizados $y_{i,f}$ $(= c_{i,f} / \sqrt{\lambda_i})$, então o vetor $\hat{\mathbf{x}}'_f$ é obtido através de

$$\hat{\mathbf{x}}_f = \mathbf{V}_q \mathbf{y}_{q,f}, \qquad (43)$$

onde $\mathbf{V}_q = [\mathbf{u}_1 \sqrt{\lambda_1} | \mathbf{u}_2 \sqrt{\lambda_2} | ... | \mathbf{u}_q \sqrt{\lambda_q}]$ e o vetor $\mathbf{y}_{q,f}$ representa os escores referentes aos q CPs utilizados no modelo.

A estatística apresentada na expressão (42), conhecida como estatística *Q* (Jackson, 1991), representa a distância quadrada perpendicular entre uma observação *p*-variada e o plano definido pelos primeiros *q* CPs retidos no modelo. Se um evento incomum aparecer no processo e alterar a estrutura de correlação contida na matriz **X**, um alto valor de Q_f será obtido. Os limites de controle para a estatística Q_f são obtidos, a partir da matriz **X** de dados preliminares do processo, utilizando resultados aproximados da distribuição das *formas quadráticas* (Jackson & Mudholkar, 1979).

Fazendo:

$$\theta_1 = \sum_{i=q+1}^{p} \lambda_i \,, \quad \theta_2 = \sum_{i=q+1}^{p} \lambda_i^2 \,, \quad \theta_3 = \sum_{i=q+1}^{p} \lambda_i^3 \quad \text{e} \quad h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2};$$

e admitindo que as p variáveis seguem um distribuição Normal p-variada, Jackson & Mudholkar (1979) demonstram que a razão

$$Z = \frac{\theta_1[(Q/\theta_1)^{h_0} - 1 - \theta_2 h_0(h_0 - 1)/\theta_1^2]}{\sqrt{2\theta_2 h_0^2}}$$

segue uma distribuição Normal com média zero e desvio-padrão unitário [$Z \sim \boldsymbol{N}(0,1)$]. A suposição de Multinormalidade das variáveis é bastante razoável, dado que as k amostras preliminares representam o processo sob controle estatístico, isto é, trazem pequenas flutuações aleatórias das variáveis em torno de suas médias. Os limites de controle da estatística Q são os seguintes:

$$\text{LSC} = \theta_1[\frac{Z_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2}]^{1/h_0} \quad \text{e}$$

$$\text{LIC} = 0, \qquad (44)$$

onde $|Z_\alpha|$ representa o desvio positivo máximo tolerado na distribuição Normal padronizada, dado à probabilidade de erro tipo I (α) admitida, se o valor de h_0 é positivo; e representa o desvio negativo máximo tolerado se o valor h_0 for negativo.

3.4.1- Padronização dos dados preliminares do processo

No exemplo químico apresentado na seção 3.3, os autovetores $\mathbf{u}_i$ foram calculados através das expressões (29), (30) e (31), utilizando a matriz **S** de covariâncias obtida a partir da matriz **X** (matriz contendo m amostras preliminares do processo) de dados centrados, isto é, com média zero). Em algumas situações práticas, principalmente quando existe um grande número de

variáveis de processo, necessita-se *padronizar* os dados da matriz **X**, ou seja, os valores referentes a cada variável passam a se distribuir com média zero e variância unitária (ver seção 2.1.2). Neste caso, a matriz **S** obtida a partir de **X** passa a ser a matriz de correlações, gerando assim autovetores $\mathbf{u}_i$ distintos daqueles obtidos a partir da matriz de dados centrados.

No exemplo químico, utilizou-se a matriz **X** de dados centrados (Tab. 3), encontrando-se a matriz de covariância

$$\mathbf{S} = \begin{bmatrix} 0{,}7986 & 0{,}6793 \\ 0{,}6793 & 0{,}7343 \end{bmatrix},$$

e obtendo-se a matriz de autovetores

$$\mathbf{U} = \left[\mathbf{u}_1 \mid \mathbf{u}_2\right] = \begin{bmatrix} 0{,}7236 & -0{,}6902 \\ 0{,}6902 & 0{,}7236 \end{bmatrix}.$$

Se na matriz **X** os dados estão padronizados, então a partir de matriz de correlação

$$\mathbf{S} = \begin{bmatrix} 1 & 0{,}887 \\ 0{,}887 & 1 \end{bmatrix},$$

têm-se, respectivamente, através das expressões (29), (30) e (31), o vetor $\boldsymbol{\lambda}$ de autovalores λ_i e a matriz **U** de autovetores

$$\boldsymbol{\lambda} = \begin{bmatrix} 1{,}887 \\ 0{,}113 \end{bmatrix} \qquad \text{e} \qquad \mathbf{U} = \begin{bmatrix} 0{,}7071 & -0{,}7071 \\ 0{,}7071 & 0{,}7071 \end{bmatrix}.$$

Observa-se que, quando se padronizam os dados preliminares, os CPs resultantes também preservam a variância total das variáveis originais, pois $tr(\mathbf{S}) = 1+1 = 2 = 1{,}887 + 0{,}113$.

A padronização dos dados é necessária já que variáveis de processo costumam representar diferentes grandezas (temperatura, pressão, comprimento, área, altura, peso,...). Se não se padronizam os dados preliminares, pode-se ter uma idéia distorcida da influência de cada variável na variabilidade total dos dados. Admita-se, por exemplo, que se a média amostral dos pesos seja de 60 kg com desvio-padrão de 7 kg (isto é, variância de 49), enquanto que a média amostral das alturas é 180 cm com desvio-padrão de 20 cm (isto é, variância de 400). Então, apenas

centralizando os dados na matriz **X**, a altura será responsável por 89% da variabilidade total dos dados [400/(400+49)].

No exemplo acima, os CPs obtidos a partir de dados apenas centrados serão fortemente dominados pela variável altura. Dessa forma, um desvio considerável no peso pode não ser refletido no primeiro CP (que, por sua vez, explica a maior porção da variabilidade total dos dados), pois a carga do autovetor $\mathbf{u}_1$, devido à variável peso, é muito baixa. Isto, sem dúvida, não é verdade, pois quando se considera a variação proporcional de cada variável [para a altura 11% (20/180) e para o peso 12% (7/60)], conclui-se que ambas possuem uma variabilidade semelhante, devendo assim ter pesos semelhantes nos CPs. A padronização dos dados corrige esta distorção.

Mesmo que as variáveis envolvidas tenham a mesma unidade de medida, a padronização pode ser necessária, pois se as respectivas variâncias forem muito heterogêneas, os CPs serão dominados pelas variáveis com maior variância.

Não existe uma relação entre os CPs obtidos através da matriz de covariâncias e os CPs obtidos a partir da matriz de correlações. No entanto, quanto mais heterogêneos forem as variâncias das variáveis originais, mais distantes estarão os dois grupos de CPs (Jackson, 1991).

Considere-se, por exemplo, a amostra futura *18*, $\mathbf{x}'_{18} = [-2{,}7, -0{,}9]$, com dados apenas centrados, apresentada na Tab. 3. Padronizando estes dados, obtém-se $\hat{\mathbf{x}}'_{18}$ = [-2,7/0,8936, -0,9/0,8569] = [-3,02, -1,05] . Através da expressão (41), calculam-se os escores $y_{1,18}$ e $y_{2,18}$, utilizando os respectivos autovetores obtidos a partir das 15 amostras preliminares em relação aos dados padronizados. Tem-se então:

$$y_{1,18} = \begin{bmatrix} \dfrac{0{,}7071}{\sqrt{1{,}887}} & \dfrac{0{,}7071}{\sqrt{1{,}887}} \end{bmatrix} \begin{bmatrix} -3{,}02 \\ -1{,}05 \end{bmatrix} = -2{,}09 \quad \text{e}$$

$$y_{2,18} = \begin{bmatrix} \dfrac{-0{,}7071}{\sqrt{0{,}113}} & \dfrac{0{,}7071}{\sqrt{0{,}113}} \end{bmatrix} \begin{bmatrix} -3{,}02 \\ -1{,}05 \end{bmatrix} = 4{,}14$$

Para obter-se a estimativa $\hat{\mathbf{x}}'_{18}$, em função dos dados padronizados, a partir do modelo e utilizando apenas o primeiro CP, faz-se, através da expressão (44),

$$\hat{\mathbf{x}}_{18} = \begin{bmatrix} 0{,}7071\sqrt{1{,}887} \\ 0{,}7071\sqrt{1{,}887} \end{bmatrix}(-2{,}09) = \begin{bmatrix} -2{,}03 \\ -2{,}03 \end{bmatrix}.$$

É importante ressaltar que, retendo apenas o primeiro CP, o modelo explica 94% [1,887/(1,887+0,113)] da variabilidade total dos dados originais (**X**).

Retendo os dois CPs (no exemplo apresentado, p=2), então a observação $\mathbf{x}'_{18}$ padronizada pode ser rescrita em função do modelo através de

$$\mathbf{x}_{18} = \begin{bmatrix} 0{,}7071\sqrt{1{,}887} & -0{,}7071\sqrt{0{,}113} \\ 0{,}7071\sqrt{1{,}887} & 0{,}7071\sqrt{0{,}113} \end{bmatrix}\begin{bmatrix} -2{,}09 \\ 4{,}14 \end{bmatrix} = \begin{bmatrix} -3{,}02 \\ -1{,}05 \end{bmatrix}.$$

3.4.2- Métodos de escolha do número *q* de CPs necessários no modelo

Sabe-se que quando p variáveis são fortemente correlacionadas, pode-se representar sua estrutura de covariância através de poucos CPs (geralmente dois ou três). No entanto é necessário algum critério estatístico de escolha para determinar o número q de CPs a ser utilizado na análise. Existem, na literatura, inúmeros critérios de seleção de CPs, sendo que cada um é mais adequado a situações específicas de aplicação da ACP.

No CEP multivariado, buscam-se m amostras preliminares que ofereçam uma caracterização suficiente do comportamento das p variáveis envolvidas, quando apenas causas comuns de variação estão presentes no processo. A partir daí, o processo será monitorado através de amostras futuras, comparando-as com a distribuição de referência obtida em função das amostras preliminares. Dentro desse contexto, o interesse é determinar um número suficiente (q)

de CPs que representem satisfatoriamente esta distribuição de referência, isto é, que sejam reflexo das principais fontes de variabilidade do processo.

Para atingir este objetivo, dois critérios surgem como boas alternativas (Nomikos & MacGregor, 1995): (*i*) *A regra da linha de corte* (*broken Stick rule;* Jolliffe, *apud* Jackson, 1991) e (*ii*) *A regra da validação cruzada* (*cross-validation rule*; Wold, 1978).

3.4.2.1- A regra da linha de corte

Considere-se um segmento de reta de comprimento unitário. Se esse segmento for dividido em *p* segmentos, admite-se que o comprimento esperado do $q^{ésimo}$ maior segmento seja

$$G_q = \frac{1}{p}\sum_{i=q}^{p}\frac{1}{i} \ .$$

Se a proporção da variancia total das *p* variáveis originais explicada por cada autovalor λ_q obtido (variância do $q^{ésimo}$ maior CP, obtido a partir da matriz **X** das *m* amostras preliminares) for superior ao valor G_q calculado, retém-se então o CP na análise (Nomikos & MacGregor, 1995).

Este critério, embora seja um tanto grosseiro, permite uma obtenção rápida e eficaz do número *q* de CPs suficientes na representação da matriz **X**.

3.4.2.2- A regra da validação cruzada

Admita-se que m amostras preliminares das p variáveis de processo tenham sido obtidas. Tem-se então uma matriz $\mathbf{X}$, $m \times p$, de dados padronizados, onde cada linha contém um vetor $\mathbf{x}'_k$ (para $k = 1,...,m$), que representa uma observação p-variada. Seguem-se então os seguintes passos (Jackson, 1991):

1. Desconsidera-se inicialmente o vetor $\mathbf{x}'_1$ e aplica-se a ACP nas $m - 1$ observações restantes, encontrando-se os p autovetores através das expressões (29), (30) e (31).

2. Através da expressão (41), obtêm-se os p escores $y_{i,1}$ referentes à observação $\mathbf{x}'_1$, utilizando os p autovalores obtidos em função das $m - 1$ amostras preliminares.

3. Através da expressão (43), retendo o primeiro CP, os dois primeiros CPs, os três primeiros, e assim por diante, até que todos os p CPs tenham sido considerados, encontram-se as respectivas estimativas $\hat{\mathbf{x}}'_1$ para a observação $\mathbf{x}'_1$ deletada.

4. Através da expressão (42), obtêm-se os erros quadráticos de predição Q_1's, referentes a cada uma das p estimativas geradas na etapa 3.

5. Retorna-se à primeira etapa e desconsidera-se agora apenas a segunda observação $\mathbf{x}'_2$, obtendo novamente p autovalores a partir das $m - 1$ observações restantes. No passo 4, obtêm-se os valores Q_2's referentes às p estimativas obtidas na etapa 3. O processo é repetido até que a ultima observação $\mathbf{x}'_m$ seja deletada e os devidos p valores Q_m's sejam obtidos.

6. Para cada um dos p modelos construídos (ou seja, retendo o primeiro CP, os dois primeiros CPs, e assim por diante) obtém-se a média dos m valores Q_k's e divide-se o valor resultante por p. O resultado obtido é conhecido como estatística *SEQP* (*soma dos erros quadráticos de predição*). Dessa forma, para o primeiro modelo (com o primeiro CP retido), obtém-se a estatística *SEQP*(1), para o segundo modelo *SEQP*(2),

para o terceiro modelo *SEQP*(3), e assim por diante. Necessita-se também obter o valor *SEQP*(0), que representa a soma dos quadrados dos *mp* dados originais.

7. Para determinar se a inclusão do $q^{ésimo}$ CP no modelo agrega informação substancial não trazida pelos q-1 CPs já incluídos sobre a variabilidade presente nos dados originais, calcula-se o valor da estatística

$$W = \frac{[SEQP(q-1) - SEQP(q)]/Dm}{SEQP(q)/Dr} \quad ,$$

onde

$$Dm = m + p - 2q \qquad \text{e} \qquad Dr = p(m-1) - \sum_{i=1}^{q}(m + p - 2i) \, .$$

Se $W > 1$, então o $q^{ésimo}$ CP deve ser retido, pois traz um acréscimo substancial no poder de predição do modelo.

Este critério é mais consistente em relação ao método apresentado na seção anterior. No entanto a obtenção dos q CPs não é tão rápida e simples quanto no critério da linha de corte.

3.4.3- Exemplo

A aplicação do CEP multivariado em processos contínuos, utilizando métodos de projeção (particularmente a ACP), tem sido ilustrada por alguns autores; entre eles MacGregor & Kourti (1995) e Kourti & MacGregor (1996). Esses autores apresentam exemplos aplicados no controle de processos químicos, mais especificamente em um processo químico para produção de um polímero denominado polietileno de baixa densidade.

Em um processo de produção de polímeros, alguns componentes químicos são periodicamente depositados em um reator de polimerização; a partir daí, estes componentes

atravessam as seções deste reator, sofrendo em cada uma delas transformações químicas até que o polímero seja produzido. Medições freqüentes das variáveis de processo, tais como temperaturas e pressões em cada seção do reator, percentuais de adição de componentes no reator e de mistura destes componentes são realizadas por sensores.

Considere-se hipoteticamente um processo químico com as características acima, onde mensurações acerca de 8 variáveis de processo (X_1, X_2,..., X_8) são realizadas periodicamente. Seguindo a filosofia do CEP tradicional, amostras periódicas preliminares dessas variáveis são recolhidas, afim de obter a sua distribuição de referência, quando o processo está sob controle estatístico, isto é, apenas sob a presença de causas comuns de variação. Através do *software SPSS* (Akey *et al.*, 2000), 30 amostras preliminares foram obtidas via simulação.

A partir da matriz **X**, de dimensão 30 × 8, contendo os dados padronizados, determina-se o número de CPs que serão utilizados no modelo para construir a distribuição de referência. Através da regra da *linha de corte* (seção 3.4.2.1), retém-se apenas o primeiro CP no modelo. Aplicando a regra da *validação cruzada* (seção 3.4.2.2), conclui-se que os três primeiros CPs oferecem uma caracterização suficiente da estrutura de correlação encontrada nos dados preliminares. Como este segundo critério é mais robusto, serão efetivamente retidos no modelo os três primeiros CPs.

Utilizando as expressões (29), (30) e (31) calculam-se os três maiores autovalores e os autovetores associados. Em seguida, através da expressões (37) e (41), obtêm-se os escores $y_{i,k}$ ($y_{1,k}$, $y_{2,k}$ e $y_{3,k}$) para cada uma das 30 observações preliminares.

Calculado o valor da estatística de Hotelling [expressão (38)] em relação às 30 observações, e o limite de controle LSC através da expressão (39), obtém-se então a carta de Hotelling, em relação aos três CPs retidos, ilustrada na Fig. 14. Observa-se que as amostras 6,7 e 24 apresentam escores acima do LSC (pontos vermelhos). Isto indica que, a partir da projeção destas observações no plano definido pelos três CPs, existem evidências que de que a variação ocorrida nas variáveis de processo, neste período, esteve acima do tolerado, dado que apenas causas comuns estão presentes no processo. A Fig. 15 mostra a projeção destas amostras no plano tridimensional definido pelos CPs retidos. Nota-se claramente que estas amostras estão fora do padrão de variação encontrado nas amostras preliminares.

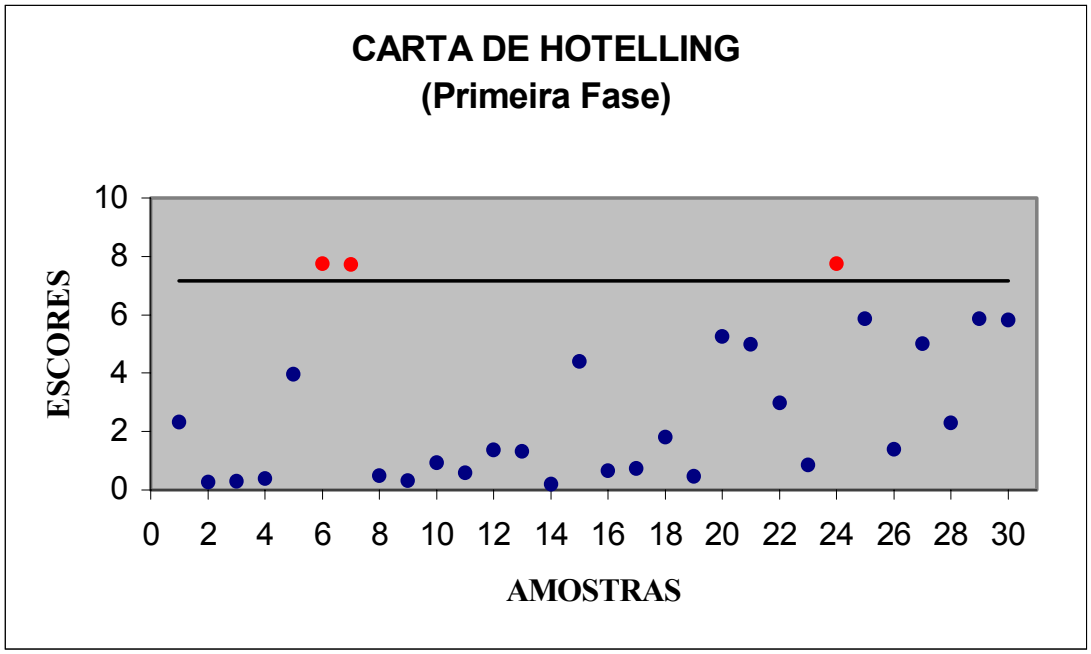

Figura 14. Carta de Hotelling em relação às 30 amostras preliminares; LSC obtido com significância de 5% (α=5%).

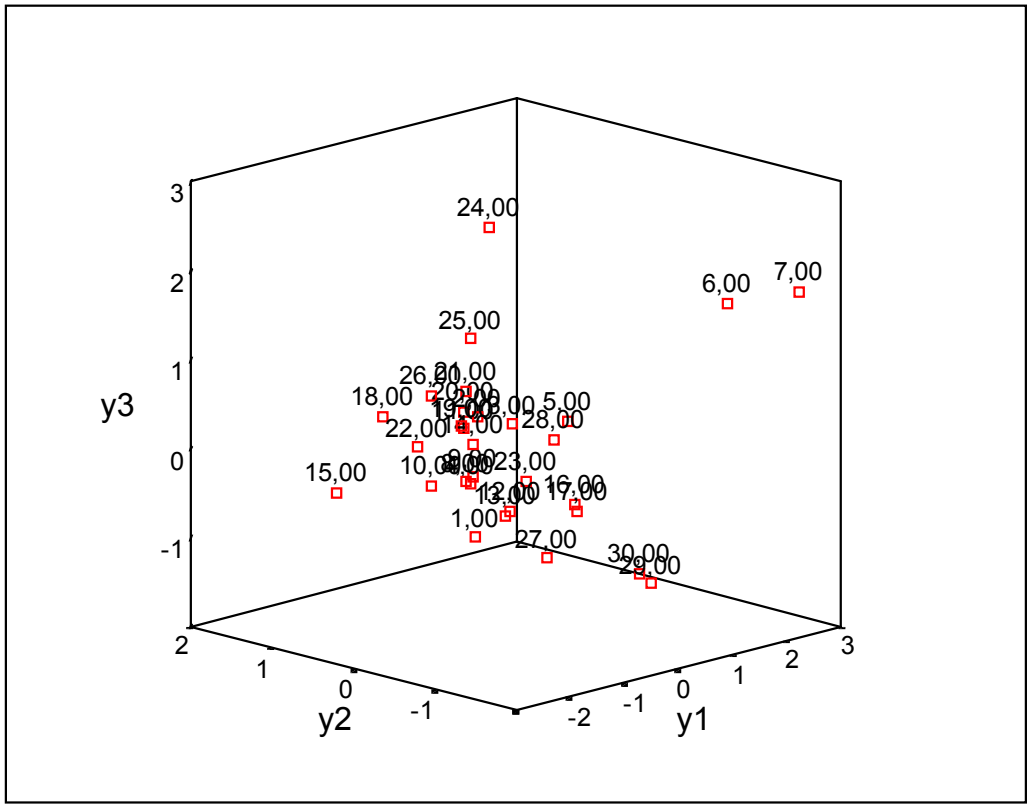

Figura 15. Projeção das 30 observações preliminares no plano definido pelos três CPs retidos.

Através da carta *Q* ilustrada na Fig. 16 [expressões (43) e (42)], observa-se que as amostras 25 e 27, embora estejam dentro do padrão na carta de Hotelling, apresentam escores acima do LSC [(expressão (44)]. Isso indica que o modelo construído em função dos três primeiros CPs não oferece uma caracterização suficiente destas observações, gerando assim um resíduo significativo. Dessa forma, existem evidências de que algum evento incomum ocorreu no processo quando estas amostras foram recolhidas. Observa-se também que as amostras 6, 7 e 24 apresentam valores *Q* acima do LSC, indicando também a possível ocorrência de um evento incomum no momento em que estas amostras foram recolhidas.

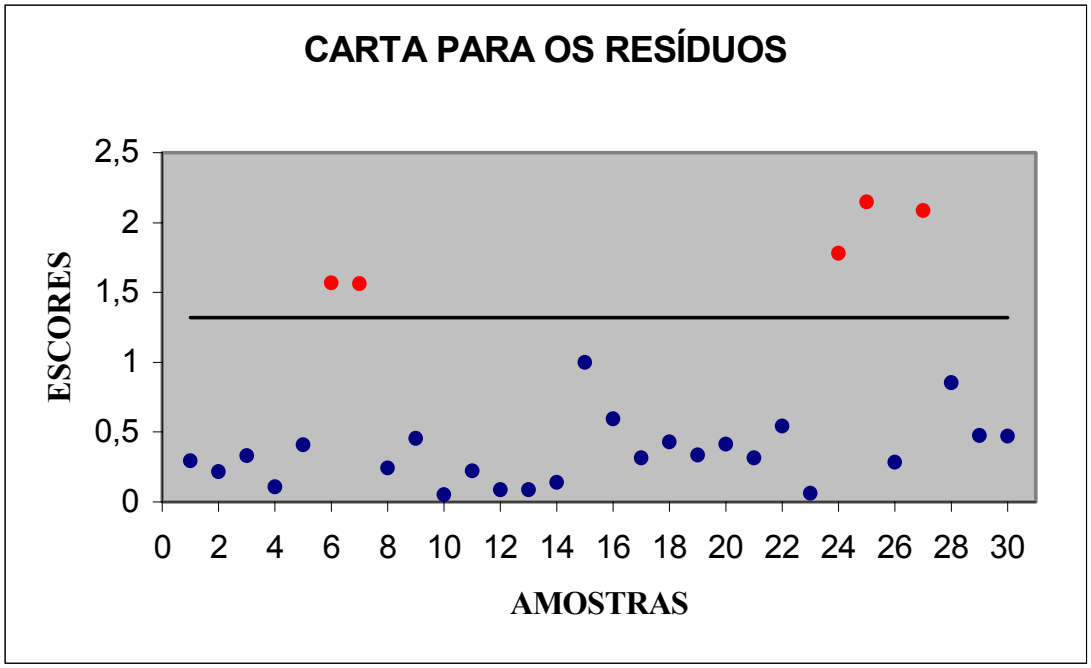

Figura 16. Carta *Q* em relação às 30 amostras preliminares; LSC obtido com significância de 5% (α=5%).

Após essa análise, considera-se que, no momento em que as amostras 6, 7, 24, 25 e 27 foram obtidas, o processo estava fora de controle estatístico. Dessa forma, essas amostras serão excluídas da distribuição de referência, por não caracterizarem suficientemente a estrutura de correlação das 8 variáveis envolvidas. Utilizando as 25 amostras preliminares restantes (com os dados novamente padronizados), obtêm-se novamente os três maiores autovalores e autovetores associados.

Utilizando novamente a estatística de Hotelling [expressão (38)] com os limites de controle da expressão (39), e a estatística *Q* [expressões (43) e (42)] com os limites de controle da expressão (44), conclui-se que as 25 amostras oferecem uma caracterização satisfatória da distribuição de referência das 8 variáveis de processo, pois nenhuma ultrapassa os respectivos LSC (ou seja, processo considerado sob controle quando estas 25 amostras foram obtidas). Dessa forma, os autovalores e autovetores, calculados a partir destas amostras, serão usados no monitoramento de futuras amostras retiradas deste processo. Obtém-se, então, o modelo ACP de referência [expressões (36) e (43)]. Nesta fase, os limites de controle LSC para as carta de Hotelling e para a carta *Q* (Fig. 16 e 17, respectivamente) são calculados através das expressões (40) e (44), respectivamente.

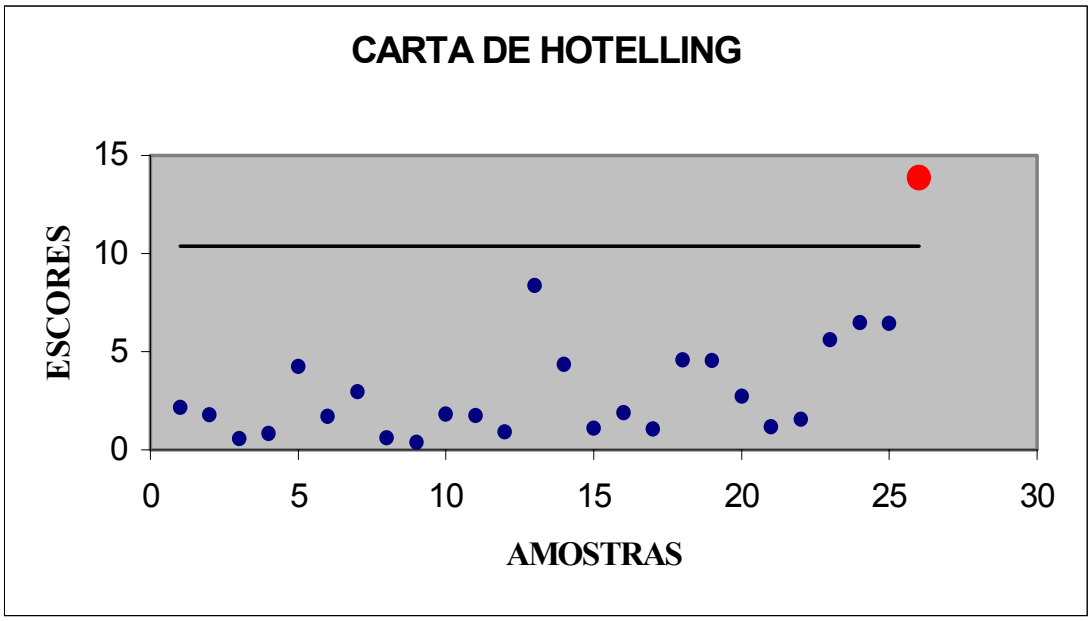

Figura 17. Carta de Hotelling em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); e nova amostra retirada do processo (amostra *26*).

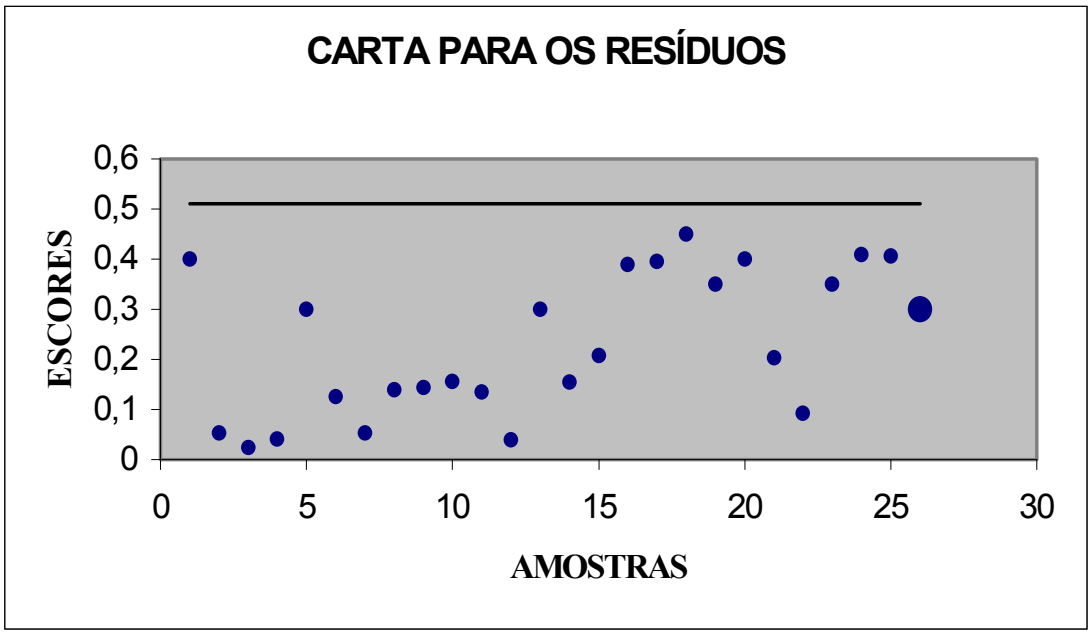

Figura 18. Carta *Q* em relação as 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); e nova amostra retirada do processo (amostra *26*).

Suponha-se que uma nova amostra tenha sido obtida do processo, representada pela amostra *26*; $\mathbf{x}'_{26}$. Através da carta *Q* (Fig. 18), observa-se que o resíduo obtido em função desta observação não é significativo (Q_{26} < LSC). Assim, não existe evidência da ocorrência de algum evento atípico no processo, pois esta observação é bem representada pelo modelo obtido em função dos três primeiros CPs. Em contrapartida, observando a carta de Hotelling , conclui-se que esta observação acusa uma variação nas variáveis de processo acima do tolerado ($T^2_{3,26}$ > LSC), supondo que apenas causas comuns de variabilidade estejam presentes no processo. Dessa forma, existe evidência de que alguma fonte de variabilidade no processo apresentou variação além do normal, refletindo numa alteração na variabilidade padrão de uma ou mais variáveis de processo.

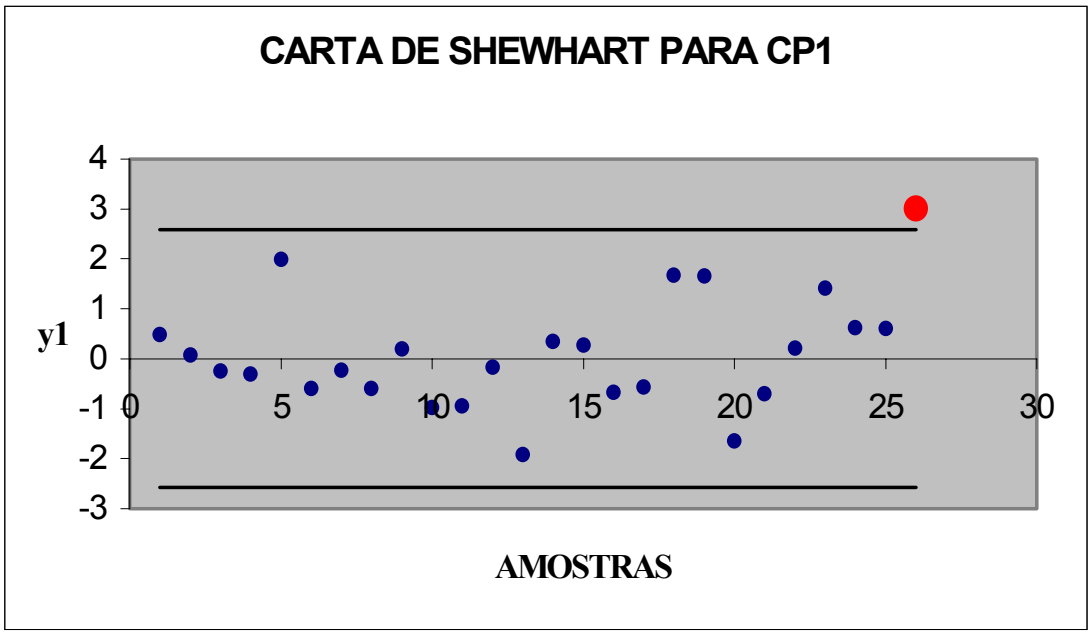

Figura 19. Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*).

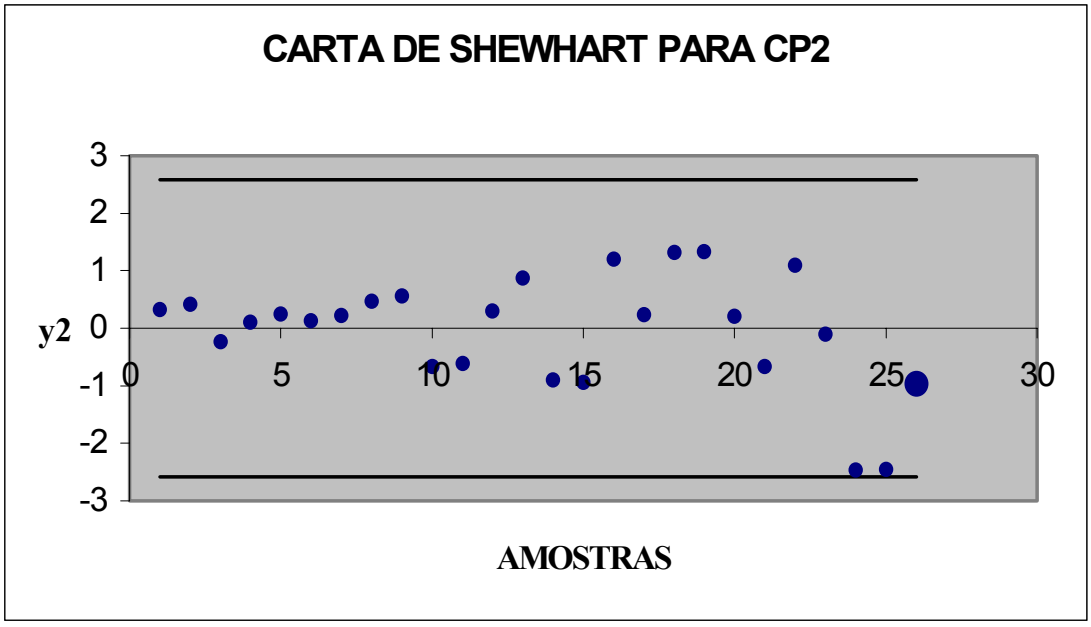

Figura 20. Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*).

Observando as cartas individuais para os CPs (Fig. 19, 20 e 21), pode-se notar que apenas o escore referente ao primeiro CP [$y_{1,26}$ obtido através da expressões (37) e (41)] está acima do

LSC [expressão (35)]. Assim, deve-se concentrar esforços neste CP na busca de um diagnóstico para o sinal de processo fora de controle, gerado a partir da carta de Hotelling.

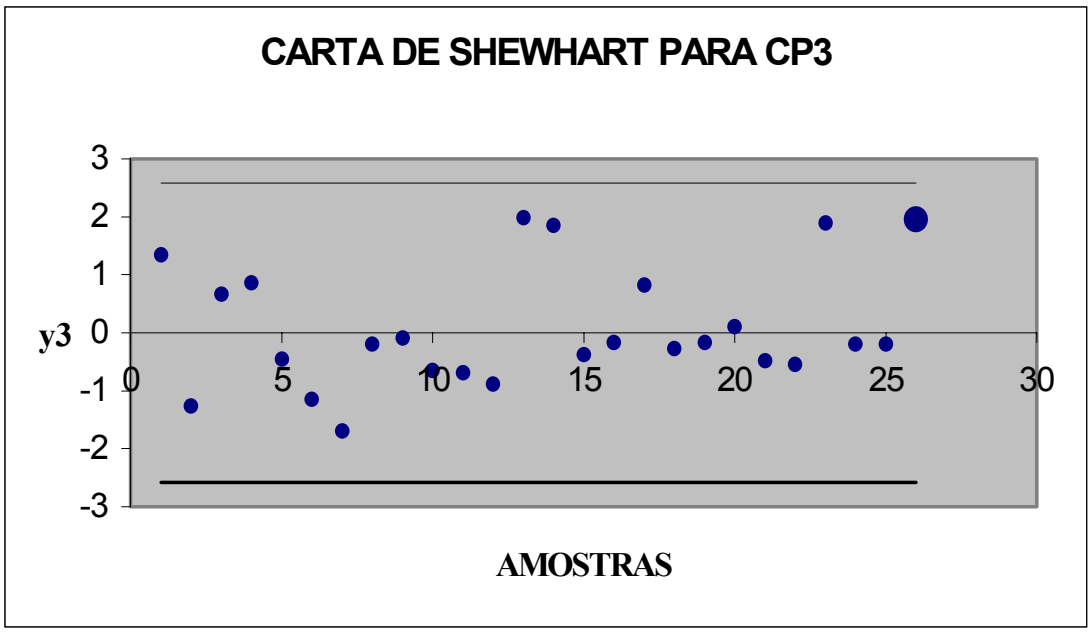

Figura 21. Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova amostra retirada do processo (amostra *26*).

Suponha-se agora que outra nova observação tenha sido recolhida do processo, representada pela amostra *27*; $\mathbf{x}'_{27}$. Através da carta de Hotelling (Fig. 22), verifica-se que não há sinal de variação demasiada nas variáveis de processo ($T^2_{3,27} <$ LSC). Entretanto, essa observação está mal representada pelo modelo construído a partir dos três CPs, pois um resíduo significativo ($Q_{27} >$ LSC) é apontado na carta Q (Fig. 23). Assim, existe evidência de que algum evento atípico está ocorrendo no processo, devendo-se buscar um diagnóstico para este sinal em função da análise destes resíduos.

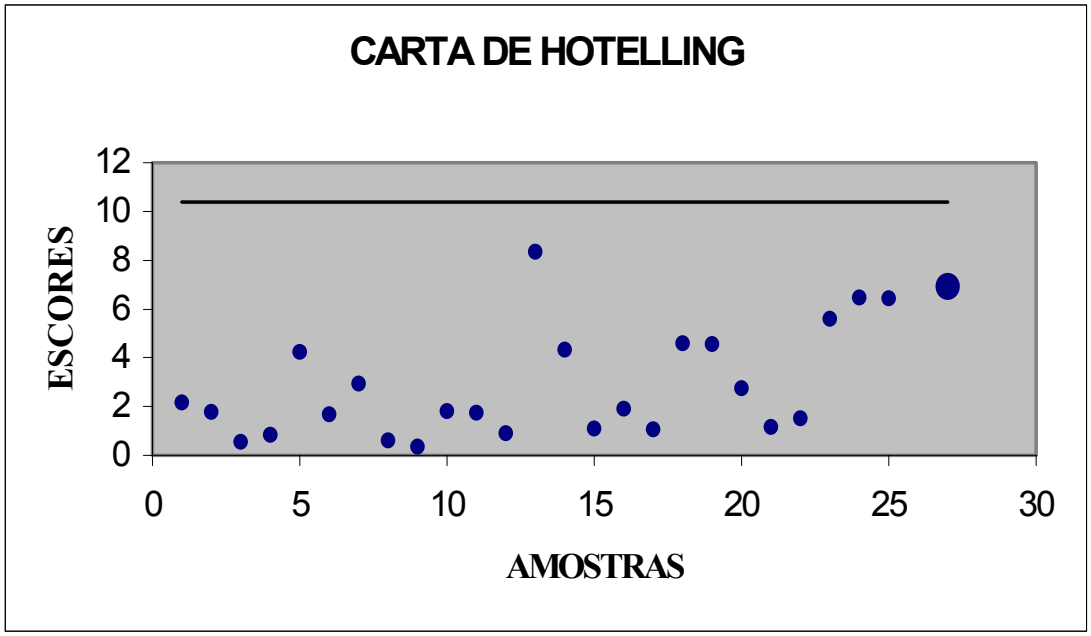

Figura 22. Carta de Hotelling em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); nova amostra retirada do processo (amostra *27*).

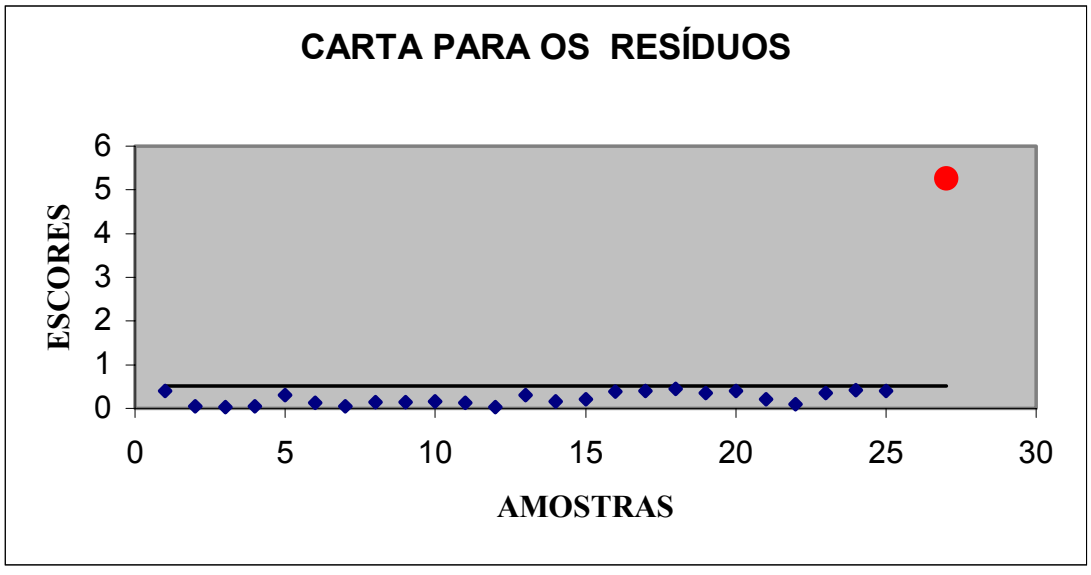

Figura 23. Carta *Q* em relação às 25 amostras preliminares; LSC obtido com significância de 5% (α=5%); nova amostra retirada do processo (amostra *27*).

3.4.4- Diagnósticos via Gráficos de Contribuição

No exemplo apresentado acima, simularam-se duas amostras futuras de 8 variáveis de processo, cada uma trazendo um sinal distinto de processo fora de controle. A primeira amostra (amostra *26*) apresentou um desvio demasiado no escore $y_{1,26}$, enquanto que a segunda amostra (amostra *27*) apresentou um resíduo Q_{27} significativamente alto. Deseja-se saber, em ambos os casos, quais variáveis contribuíram mais efetivamente para a ocorrência dos altos valores encontrados, tanto para $y_{1,26}$, quanto para Q_{27}. O objetivo é buscar um diagnóstico para esses sinais, isto é, encontrar a causa que levou o processo a sair do seu estado de controle.

Na seção 2.2.7, foram citados alguns métodos de diagnóstico de causas especiais de variação, dentre eles os Gráficos de Contribuição. Estas ferramentas de diagnóstico foram sugeridas por Miller *et al.* (1993) e MacGregor *et al*. (1994). Os Gráficos de Contribuição indicam as variáveis que possuem os maiores pesos no cálculo dos escores $y_{i,f}$ e Q_f.

Sabe-se que um escore $y_{i,f}$ (relativo ao CP_i), referente a uma amostra futura *f*, é obtido através de [expressão (41)]

$$y_{i,f} = \mathbf{u}_i^{*\prime}\,\mathbf{x}_f, \qquad \text{com} \quad \mathbf{u}_i^{*} = \mathbf{u}_i / \sqrt{\lambda_i}\,.$$

Fazendo:

$$\mathbf{u}_i^{*\prime} = \begin{bmatrix} u_{i,1} & u_{i,2} & .... & u_{i,p} \end{bmatrix} \qquad \text{e} \qquad \mathbf{x}_f = \begin{bmatrix} x_{1,f} \\ x_{2,f} \\ .... \\ x_{p,f} \end{bmatrix},$$

pode-se rescrever a expressão (41) da seguinte forma:

$$y_{i,f} = u_{i,1}x_{1,f} + u_{i,2}x_{2,f} + ... + u_{i,p}x_{p,f} = \sum_{j=1}^{p} u_{i,j}x_{j,f}\,,$$

onde a parcela $u_{i,j}\,x_{j,f}$ representa a contribuição da $j^{ésima}$ variável para o escore $y_{i,f}$. As maiores parcelas, conseqüentemente, possuem maior contribuição para a formação do valor final de $y_{i,f}$,

devendo assim ser investigadas. Kourti & MacGregor (1996) sugerem que apenas as parcelas maiores e com mesmo sinal em relação ao escore $y_{i,f}$ devam ser investigadas. Se, por exemplo, algum escore apresentar um valor significativo negativo, deve-se entender que as parcelas negativas ofereceram as maiores contribuições para este valor. Neste caso, a inclusão de uma parcela positiva na análise, mesmo que com valor relevante, não irá agregar informação na busca das causas para a variação espúria refletida neste escore.

Quando mais de um escore $y_{i,f}$, referente a uma mesma amostra *f*, apresentar valor significativo, pode-se obter a contribuição conjunta da $j^{ésima}$ variável para os escores $y_{i,f's}$ da seguinte forma:

$$\text{CONT}(j)=\sum_{i=1}^{a} y_{i,f} u_{i,j} x_{j,f} \quad ,$$

onde *a* representa o número de escores com valores significativos. Kourti & MacGregor (1996) sugerem que parcelas com sinal diferente em relação ao escore devem assumir valor zero no cálculo da contribuição conjunta.

A Fig. 24 ilustra as contribuições das 8 variáveis de processo na composição do valor de $y_{1,26}$ (escore referente à amostra *26*). Observa-se que as três primeiras variáveis de processo (X_1, X_2, X_3) apresentam as maiores contribuições. Dessa forma, devem-se concentrar esforços nestas variáveis na busca de uma causa para o alto valor encontrado em $y_{1,26}$.

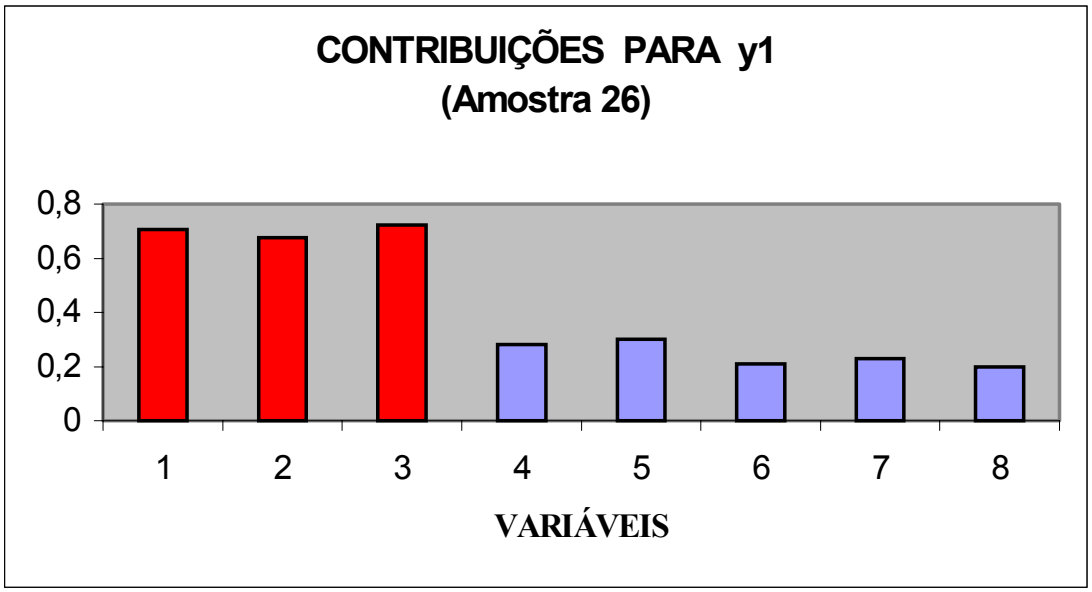

Figura 24. Contribuição das 8 variáveis de processo para o escore y_1 (CP_1); amostra *26*.

Adotando a mesma filosofia utilizada em relação aos escores $y_{i's}$, pode-se determinar as variáveis que mais contribuem para um alto valor de Q_f. Sabe-se que o resíduo total referente a uma amostra f é obtido através de

$$Q_f = \sum_{i=1}^{p} (x_{i,f} - \hat{x}_{i,f})^2 .$$

A parcela $x_{i,f} - \hat{x}_{i,f}$ representa o resíduo gerado a partir da $i^{ésima}$ variável. Novamente, deve-se entender que as maiores parcelas possuem maior contribuição para a formação do valor final de Q_f.

A Fig. 25 apresenta as contribuições das 8 variáveis de processo na composição do valor de Q_{27} (escore referente à amostra *27*). Observa-se que as variáveis 3, 7 e 8 (X_3, X_7, X_8) apresentam as maiores contribuições. Nesse caso, devem-se concentrar esforços nestas variáveis na busca de uma causa para o alto valor encontrado em Q_{27}.

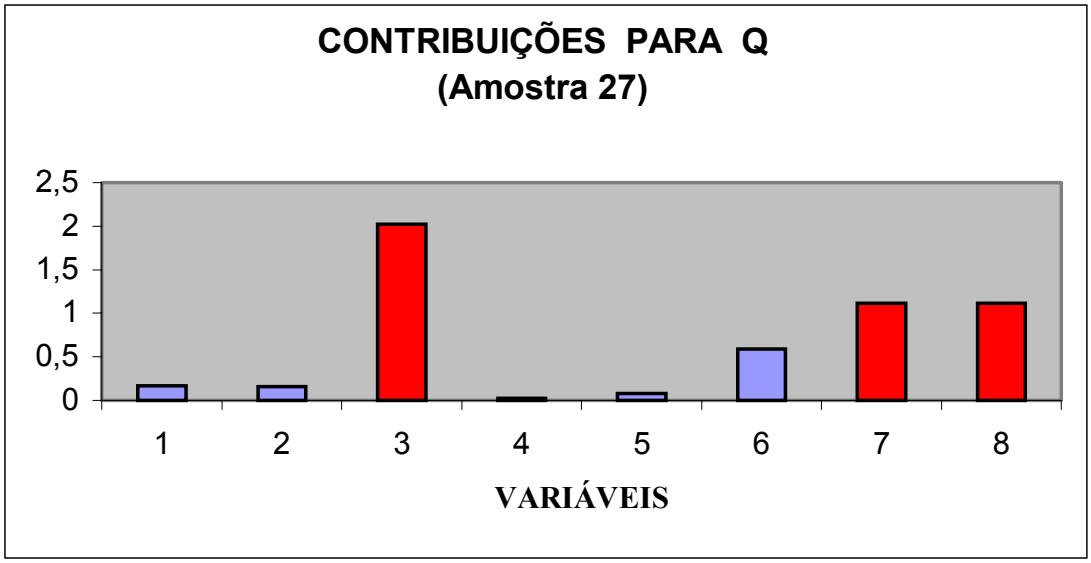

Figura 25. Contribuição das 8 variáveis de processo para o valor *Q*; amostra *27*.

CAPÍTULO 4

4- CARTAS DE CONTROLE MULTIVARIADAS VIA ANÁLISE DE COMPONENTES PRINCIPAIS MULTIDIRECIONAIS (ACPM)

4.1- Processos em batelada

O cenário de interesse neste trabalho são as indústrias químicas e de processos que conduzam processos em bateladas. Processos em batelada ou semi-bateladas são utilizados, com freqüência, na obtenção de produtos químicos, bioquímicos, farmacêuticos e alimentícios.

Um processo químico de produção por bateladas passa pelas seguintes etapas: (*i*) uma determinada combinação de matérias-primas é colocada num recipiente; (*ii*) durante a batelada, estas matérias-primas sofrem uma série de transformações, que são controladas através do

monitoramento de trajetórias específicas das variáveis de processo, tais como temperaturas, pressões e taxas de mistura; e (*iii*) ao término da batelada, o produto final é analisado em laboratório quanto às suas características de qualidade. A Fig. 26 apresenta os elementos genéricos de um processo em batelada. As variáveis de qualidade ***Z*** são obtidas a partir de medições de qualidade feitas sobre as matérias-primas. As variáveis de processo ***X*** são mensurações *on-line* do desempenho do processo, em pontos próximos no tempo. As variáveis de qualidade final do produto ***Y*** são mensuradas, no produto acabado, após o término da batelada.

Figura 26. Esquematização de uma batelada.

O objetivo dessa dissertação é a apresentação e a aplicação de cartas de controle para processos em bateladas, utilizando a ACPM. É importante ressaltar que esta abordagem considera somente dados acerca das variáveis de processo (***X***) na elaboração das cartas de controle, e, conseqüentemente, monitora o processo através de amostras futuras das variáveis de processo. A abordagem PELM, que se constitui na extensão da PEL (brevemente descrita nas seções 2.3.2 e 2.3.2.1) para processos em bateladas, é uma técnica alternativa à ACPM que permite a utilização das informações contidas em ***Z***, ***X*** e ***Y*** no monitoramento destes processos; para maiores detalhes, ver MacGregor *et al*. (1994).

Salienta-se também que o foco deste trabalho são os processos químicos em bateladas de tempo fixo. Em um processo químico de tempo variável, o tempo de realização de uma batelada depende do comportamento do processo, ou seja, dos eventos que ocorreram durante a batelada. Uma série de transformações químicas ocorrem durante a batelada até que se atinja um valor alvo para as variáveis de processo, dentro de um determinado tempo. Se a batelada transcorreu dentro da normalidade (presença apenas de causas comuns de variação incidindo no processo), então este tempo estará dentro do padrão. No processo em bateladas de tempo fixo, as variáveis seguem suas trajetórias específicas durante cada batelada (todas com mesmo tempo de duração), sendo que se o processo está sob controle, cada variável estará seguindo sua trajetória-padrão (ou

trajetória média). Sempre que as variáveis do processo apresentarem variações toleráveis em relação às suas trajetórias médias, a batelada resultante será considerada como conforme.

Variações entre bateladas são conseqüência, por exemplo, de uma combinação inadequada de matérias-primas utilizada no início da batelada, impurezas presentes no processo e desvios das variáveis de processo em relação às suas trajetórias médias (Nomikos & MacGregor, 1995). O objetivo principal no controle desses processos é manter uma variabilidade mínima entre as bateladas, gerando um produto final consistente. Em outras palavras, em cada batelada, deseja-se que a trajetória média de cada variável de processo não seja significativamente alterada. Se alguma causa especial de variação incidir sobre o processo, acarretará uma variabilidade excessiva entre bateladas, alterando a trajetória média de uma ou mais variáveis do processo. A conseqüência é a obtenção de um produto final fora das especificações de qualidade.

A estrutura dos dados em um processo em batelada apresenta diferenças em relação à estrutura dos dados obtidos de processos contínuos. Um processo contínuo é estacionário, isto é, se apenas causas comuns estão atuando, as variáveis apresentam pequenas flutuações em torno de suas médias fixas. A partir de sucessivas amostras periódicas (obtidas apenas quando causas comuns estiverem presentes), pode-se obter a distribuição de referência das variáveis de processo (estrutura de correlação das variáveis de processo) para realizar o monitoramento das médias, através de amostras futuras. Os processos em bateladas normalmente não são estacionários (Marsh & Tucker, 1991). Conseqüentemente, a média ou valor alvo das variáveis de processo não é constante, bem como seus limites de tolerância. Desta forma, nestes processos, obter a distribuição de referência das variáveis de processo implica modelar suas trajetórias médias (trajetórias realizadas durante bateladas bem sucedidas).

Durante cada batelada, inúmeras amostras consecutivas são obtidas das variáveis de processo. Ao final da batelada, dispõem-se de várias mensurações das variáveis de processo em pontos espaçados no tempo. Para construir-se uma distribuição de referência adequada e realizar um monitoramento eficiente das bateladas futuras, deve-se considerar não apenas a estrutura de correlação das variáveis de processo (como nos processos contínuos), mas também a estrutura de autocorrelação de cada variável dentro das bateladas. Para tal, necessita-se considerar o histórico dessas variáveis em bateladas passadas bem sucedidas, tratando cada batelada como uma amostra

completa do comportamento de cada variável em todos os instantes observados, dentro da batelada. Assim, cada batelada observada traz uma amostra das trajetórias de todas as variáveis.

A Fig. 27 mostra as trajetórias de três variáveis de processo, observadas em 20 instantes, dentro de cada batelada, durante 30 bateladas consecutivas bem sucedidas. Observa-se que cada variável apresenta uma flutuação em torno da sua média que, por sua vez, ocupa diferentes patamares a cada instante. As bandas amarelas, vermelhas e azuis representam a magnitude da variação de cada variável (observada através de 30 bateladas amostradas) em cada instante, em torno de suas médias (ou em torno de suas trajetórias médias). Isso caracteriza uma estrutura de autocorrelação de cada uma das três variáveis neste processo. Quando se observa a posição das três bandas em cada instante, pode-se visualizar a estrutura de correlação apresentada pelas variáveis em cada instante.

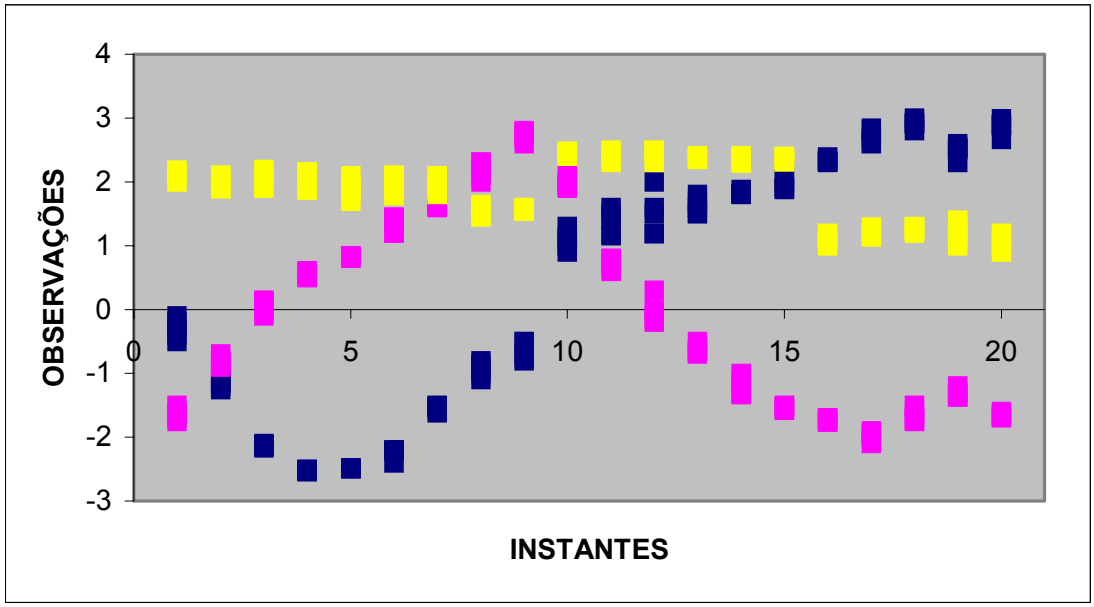

Figura 27.Trajetórias de três variáveis de processo observadas em 20 instantes, durante 30 bateladas consecutivas. Exemplo adaptado de Nomikos & MacGregor, 1995.

4.2- ACPM

Considere-se um processo químico por batelada com i=1,...,p variáveis mensuradas em l=1,...,h instantes. Dados coletados acerca de k=1,...,m amostras (ou bateladas) preliminares deste processo são arranjados na matriz $\underline{\mathbf{X}}$ tridimensional (bateladas × variáveis × instantes), representada na Fig. 28.

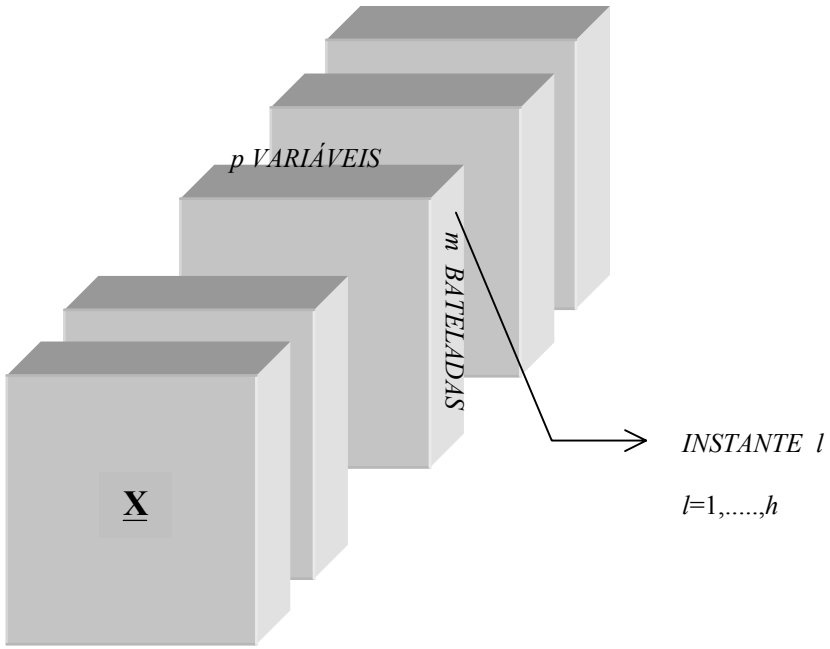

Figura 28. Representação dos dados num processo em bateladas, organizados na matriz $\underline{\mathbf{X}}$ tridimensional.

Pode-se observar através da matriz $\underline{\mathbf{X}}$ (Fig. 28) a estrutura dos dados num processo em bateladas. Cada fatia de $\underline{\mathbf{X}}$ apresenta os dados acerca de p variáveis de processo, em m bateladas, observadas em um instante l distinto. A Fig. 29 apresenta uma projeção bidimensional da matriz $\underline{\mathbf{X}}$, sugerida por Nomikos & MacGregor (1994). Na projeção, a matriz $\underline{\mathbf{X}}$, de dimensão $m \times p \times h$, é decomposta em uma matriz $\mathbf{X}$ bidimensional, de dimensão $m \times ph$, que contém cada uma das fatias representadas em $\underline{\mathbf{X}}$ arranjadas lado a lado, em ordem cronológica. Desta forma, cada linha da matriz $\mathbf{X}$ apresenta um vetor $\mathbf{x}_k'$, de dimensão $1 \times ph$, que representa uma observação

ph-variada. Cada observação traz o resultado das mensurações de *p* variáveis de processo em *h* instantes dentro da batelada, ou seja, nada mais é do que o resultado da realização completa de uma batelada.

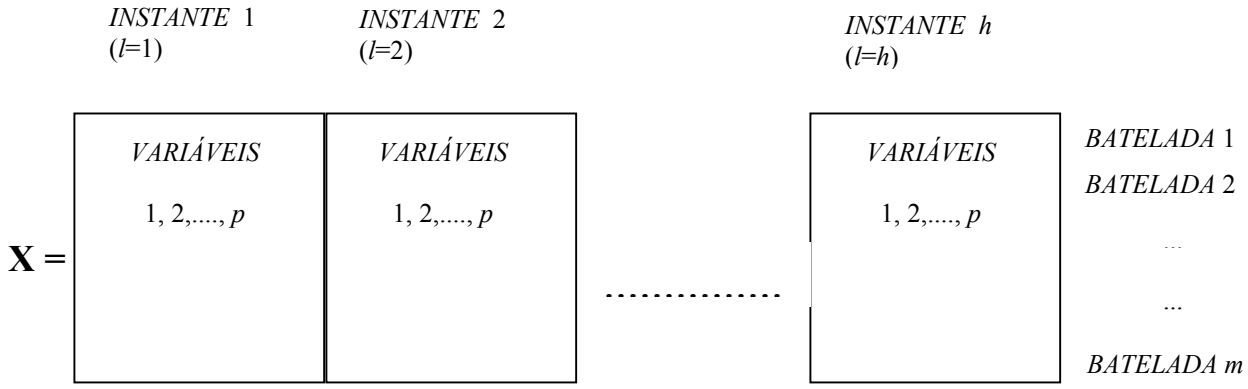

Figura 29. Representação dos dados num processo em bateladas, organizados na matriz **X** bidimensional.

A *Análise de Componentes Principais Multidirecionais* (ACPM) consiste na aplicação da *Análise de Componentes Principais* (ACP) na matriz **X**, tratando cada uma das *ph* colunas desta matriz como uma variável aleatória distinta. Na verdade, como cada linha desta matriz contém *h* replicações das mesmas *p* variáveis, tem-se *ph* variáveis fortemente correlacionadas. Isto acorre porque, em cada instante, as mensurações obtidas acerca das *p* variáveis de processo representam diferentes vizualizações de um mesmo grupo de eventos comuns, presentes no processo. Sabe-se, então, que essas *ph* variáveis devem conter uma boa parcela de informação redundante sobre o processo (ruído), dado que nele dificilmente se verificam *ph* eventos distintos ocorrendo simultaneamente. Desta forma, poucos CPs (geralmente dois ou três) irão descrever suficientemente a estrutura geral de correlação dos dados. Tal estrutura, neste caso, inclui a estrutura de correlação das *p* variáveis dentro de cada batelada e a estrutura de autocorrelação de cada variável durante as bateladas. Em outras palavras, poucos CPs capturam as principais direções de variabilidade do processo, descrevendo satisfatoriamente as trajetórias médias das suas *p* variáveis.

A partir da obtenção de dados preliminares acerca de m bateladas, estima-se a estrutura geral de correlação dos dados da matriz **X**, de dimensão $m \times ph$, através da matriz de covariâncias amostrais **S**, de dimensão $ph \times ph$. Os ph autovalores e respectivos autovetores podem ser calculados a partir da matriz **S**, utilizando as expressões (29), (30) e (31).

Os m escores relativos ao $i^{ésimo}$ CP podem ser obtidos através de

$$\mathbf{c}_i = \mathbf{X}\mathbf{u}_i ,$$

onde $\mathbf{u}_i$ representa o $i^{ésimo}$ autovetor obtido de **S** e contém as cargas referentes ao $i^{ésimo}$ CP. Sabe-se que os CPs são variáveis independentes, pois cada autovetor associado representa a projeção das ph variáveis originais na $i^{ésima}$ maior direção de variabilidade dos dados, ortogonal em relação às demais (ph-1) direções. Esta variabilidade é estimada pelo autovalor λ_i, que representa, então, a variância do $i^{ésimo}$ CP.

Pode-se reescrever a matriz **X** em função dos ph CPs obtidos [expressão (33)] da seguinte forma:

$$\mathbf{X} = \sum_{i=1}^{ph} \mathbf{c}_i \mathbf{u}_i'$$

Como as ph variáveis são fortemente correlacionadas, sabe-se que os q (< ph) primeiros CPs devem oferecer uma caracterização satisfatória da estrutura geral de correlação das variáveis iniciais. Desta forma, a matriz **X** de dados pode ser bem representada pelo modelo [expressão (36)]

$$\hat{\mathbf{X}} = \sum_{i=1}^{q} \mathbf{c}_i \mathbf{u}_i' + \mathbf{E} , \qquad (45)$$

com

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} .$$

A matriz $\hat{\mathbf{X}}$ contém os dados originais acerca das m bateladas amostradas, estimados em função dos q primeiros CPs, retidos segundo algum critério (geralmente através da *validação cruzada*;

seção 3.4.2). A matriz **E**, por sua vez, contém os resíduos obtidos do modelo (diferença entre o observado e o estimado através do modelo proposto).

O monitoramento de um processo em bateladas através da ACPM é realizado de maneira similar ao monitoramento de processos contínuos via ACP (seção 3.4). Monitoram-se as trajetórias das p variáveis de processo através de amostras de bateladas futuras, projetando a referente observação $\mathbf{x}'_f$, de dimensão $1 \times ph$, no plano definido pelos q CPs retidos obtendo os referentes escores $c_{i,f}$. Reescreve-se então a expressão (37):

$$c_{i,f} = \mathbf{u}'_i \mathbf{x}_f, \quad \text{para} \quad i=1,...,q \tag{46}$$

Deve-se observar que cada autovetor resume a variação das p variáveis de processo em cada instante da batelada em torno das suas trajetórias médias, sendo que suas cargas representam a contribuição de cada variável, em cada instante, na variabilidade encontrada nas k bateladas preliminares, utilizadas como referência no controle do processo. Desta forma, o escore $c_{i,f}$ traduz a variabilidade encontrada numa batelada futura em relação à variabilidade capturada nas m bateladas bem sucedidas, utilizadas na construção do modelo de referência (distribuição de referência dos dados).

Monitora-se efetivamente as trajetórias das variáveis de processo através da estatística de Hotelling, escrita em função dos q CPs retidos [(expressão (38)], ou seja:

$$T^2_{q,f} = \sum_{i=1}^{q} \frac{c^2_{i,f}}{\sqrt{\lambda_i}} = \sum_{i=1}^{q} y^2_{i,f}, \tag{47}$$

onde $y_{i,f}$ (= $c_{i,f} / \sqrt{\lambda_i}$) representa a padronização do escore $c_{i,f}$ [(expressão (41)].

O valor $T^2_{q,f}$ obtido a partir de uma observação $\mathbf{x}'_f$, referente a uma batelada futura, é comparado aos limites de controle apresentados na expressão (40). A estatística $T^2_{q,f}$ representa a distância de Mahalanobis, no plano definido pelos q CPs retidos no modelo, entre a posição da batelada futura (dada pelos escores $y_{i,f}$) e a origem que representa o ponto de mínima variação do processo. Assim, um valor elevado de $T^2_{q,f}$ significa que a batelada apresentou uma variabilidade muito além daquela tolerada, considerando a hipótese de apenas causas comuns presentes no

processo (ou seja, as variáveis apresentaram trajetórias com desvios significativos em relação às suas trajetórias médias). Os valores $T^2_{q,k}$ de Hotelling, referentes às bateladas preliminares $\mathbf{x}'_k$, também são comparados aos limites de controle, para excluir da distribuição de referência uma possível batelada que tenha apresentado variação além do tolerado. Nesta fase, o limite de controle LSC é obtido através da expressão (39) (Primeira fase do controle; ver seções 2.2.2.1 e 3.4.3).

Se o valor de $T^2_{q,f}$ for superior a LSC, a exemplo do que é feito na ACP em relação a processos contínuos, deve-se analisar os referentes escores $y_{i,f}$ na busca de um diagnóstico para este sinal. Estes escores devem ser comparados com os limites apresentados na expressão (35). Se algum escore $y_{i,f}$ estiver além dos limites de controle, então deve-se, através dos Gráficos de Contribuição (seção 3.4.4), analisar as cargas do autovetor $\mathbf{u}_i$ correspondente, para verificar qual variável teve maior contribuição na formação deste valor elevado. Pode-se assim ter uma idéia de quais variáveis apresentaram desvios significativos em torno de suas trajetórias médias e em quais instantes isso ocorreu dentro da batelada.

Além das utilização da carta de Hotelling e das cartas individuais para os escores, é aconselhável utilizar a carta de controle Q_f [expressão (42)] para monitorar os resíduos obtidos nas bateladas futuras analisadas. Dado um vetor $\mathbf{x}'_f = [x_{1,f},...,x_{ph,f}]$ referente a uma batelada futura, tem-se:

$$Q_f = \sum_{i=1}^{ph} (x_{i,f} - \hat{x}_{i,f})^2 \ . \qquad (48)$$

A observação $\mathbf{x}'_f$ é estimada através do vetor $\hat{\mathbf{x}}_f = \mathbf{U}_q \mathbf{c}_{q,f}$ (observação $\mathbf{x}'_f$ reescrita a partir do modelo de referência), onde $\mathbf{c}_{q,f}$ é o vetor, de dimensão $q \times 1$, de escores obtidos do modelo (cada um referente a uma dos CPs retidos), e $\mathbf{U}_q = [\mathbf{u}_1 \,|\, \mathbf{u}_2 \,|\, ... \,|\, \mathbf{u}_q]$ é a matriz dos q autovetores retidos. Utilizando-se os escores padronizados, a observação $\mathbf{x}'_f$ pode ser reescrita a partir do modelo [expressão (43)]

$$\hat{\mathbf{x}}_f = \mathbf{V}_q \mathbf{y}_{q,f} \ , \qquad (49)$$

onde $\mathbf{y}_{q,f}$ é o vetor, de dimensão $q \times 1$, de escores padronizados $y_{i,f}$ [expressão (41)], e $\mathbf{V}_q = [\mathbf{u}_1\sqrt{\lambda_1} \,|\, \mathbf{u}_2\sqrt{\lambda_2} \,|\, ... \,|\, \mathbf{u}_q\sqrt{\lambda_q}]$ é a matriz dos q autovetores multiplicados pela raiz quadrada dos seus devidos autovalores. Utiliza-se a matriz $\mathbf{V}_q$ no lugar da matriz $\mathbf{U}_q$ para gerar a observação estimada $\hat{\mathbf{x}}_f$ na mesma unidade de medida da observação original $\mathbf{x}'_f$, visto que utilizaram-se os escores padronizados $y_{i,f}$.

A estatística Q_f representa a distância perpendicular entre a batelada futura representada na sua dimensão original (isto é, no plano definido pelas *ph* variáveis originais) e o plano definido pelos q CPs retidos no modelo. O valor Q_f é comparado aos limites de controle apresentados na expressão (44) (Jackson & Mudholkar, 1979). Um valor elevado de Q_f indica que a estrutura de correlação apresentada pelas variáveis na batelada futura $\mathbf{x}'_f$ não está de acordo com a estrutura de correlação capturada pelo modelo de referência, construído a partir das *m* bateladas preliminares bem sucedidas. Neste caso, a exemplo do que foi dito nos processos contínuos, existem indícios de que algum evento atípico, não contemplado na distribuição de referência, esteja presente no processo.

Novamente utilizando os Gráficos de Contribuição (seção 3.4.4), pode-se determinar o grupo de variáveis que mais contribuíram para o alto resíduo gerado e em quais instantes dentro da batelada o modelo não representou satisfatoriamente estas variáveis na observação $\mathbf{x}'_f$. Os valores Q_k, referentes às bateladas preliminares $\mathbf{x}'_k$ também são comparados aos limites de controle, para excluir da distribuição de referência uma possível batelada em que as *ph* variáveis tenham apresentado estrutura de correlação diferente em relação às demais (*m*-1) bateladas selecionadas (Primeira fase do controle; ver seções 2.2.2.1 e 3.4.3).

Os resíduos Q_f referentes a uma batelada futura também podem indicar se alguma variável da processo apresenta desvio acentuado em relação à sua trajetória média. Visto que na abordagem ACPM, a cada instante *l*, *p* variáveis de processo são tratadas como *l* variáveis distintas (Fig. 29), sabe-se então que cada um dos *q* CPs representam uma combinação linear de *ph* variáveis. Após obter o modelo de referência a partir das *m* bateladas preliminares, uma variável em determinado instante pode apresentar cargas baixas nos *q* autovetores $\mathbf{u}_i$, indicando que no referente período o processo é bastante estável (pouca variabilidade das suas fontes de

variação). Desta forma, se uma batelada futura apresentar variabilidade demasiada neste período, o elevado desvio medido na variável correspondente não será refletido nos escores y_i's, pois será multiplicado por uma coeficiente baixo (carga baixa). No entanto esta observação não será bem descrita pelo modelo ACPM ($\hat{\mathbf{x}}_f = \mathbf{V}_q \mathbf{y}_{q,f}$), gerando assim um elevado resíduo Q_f.

Salienta-se novamente que a única suposição necessária para a utilização dos limites de controle para as cartas de Hotelling e Q [apresentados nas expressões (40) e (44), respectivamente], é que as *ph* variáveis sigam uma distribuição Multinormal (neste caso, uma Normal *ph*-variada). A exemplo do que foi dito para os processos contínuos na seção 3.4, esta suposição é bastante razoável, visto que se o processo está sob controle estatístico, as *m* bateladas preliminares coletadas representam desvios aleatórios das variáveis em relação às suas trajetórias médias. Como os CPs são combinações lineares dessas variáveis, eles também apresentam flutuações aleatórias em torno de suas médias (Horswell & Looney, *apud* Nomikos & MacGregor, 1995); desta forma, são bem descritos pelo modelo Normal. Por outro lado, independente da distribuição subjacente das *ph* variáveis iniciais, cada CP pode ser bem aproximado pelo modelo Normal, pois representam uma soma ponderada de variáveis aleatórias (*ph* variáveis); ver *teorema do limite central*, Mood *et al*. (1974). Os q CPs seguem então uma distribuição aproximadamente Normal q-variada, com vetor de médias **0** (já que dados originais foram centrados) e matriz de covariâncias amostrais **S** diagonal (pois os CPs são independentes), de dimensão ($q \times q$).

Resumindo o exposto nesta seção, o monitoramento de um processo em bateladas é realizado através de amostras $\mathbf{x}'_f$ de bateladas futuras, das quais obtêm-se, a partir dos escores $c_{i,f}$*'s* [expressão (46)], os escores padronizados $y_{i,f}$*'s* [expressão (41)], o devido valor $T^2_{q,f}$ (calculado em função dos q primeiros CPs retidos no modelo) e o resíduo Q_f [expressões (47), (49) e (48)]. Estes valores são comparados com os seus respectivos limites de controle [expressões (35), (40) e (44)]. Através da estatística $T^2_{q,f}$, verifica-se se a variabilidade de alguma das p variáveis em torno das suas médias, em qualquer instante dentro da batelada, está de acordo com a distribuição de referência obtida quando apenas causas comuns de variação incidem sobre o processo. Um elevado valor $T^2_{q,f}$ indica que alguma variável, em algum instante dentro da batelada, apresentou um desvio significativo em relação à sua trajetória média (provável presença

de causas especiais). Já através da estatística Q_f, verifica-se a estrutura geral de correlação das *ph* variáveis dentro da batelada futura está de acordo com a estrutura de correlação capturada no modelo de referência. Um elevado valor Q_f indica que o modelo de referência (obtido sob a presença apenas de causas comuns de variação) não descreveu satisfatoriamente esta observação. Isto sugere a provável presença de um evento atípico, não incluído no modelo de referência, evidenciando assim novamente a provável presença de causas especiais no processo.

4.3- Monitoramento on-line via ACPM

Na seção acima, foi apresentada a filosofia das cartas de controle via ACMP para monitoramento de processos por bateladas. Ao final de cada nova batelada, após o preenchimento completo do vetor $\mathbf{x}'_f$, projeta-se a observação no plano definido pelos q CPs retidos, monitorando os escores $c_{i,f}$'s , o valor $T^2_{q,f}$ e o Q_f. Desta forma, pode-se verificar se na batelada em questão, as p variáveis de processo apresentaram comportamento dentro do padrão desejado. No entanto se, por exemplo, algum evento atípico começa a incidir no processo no início de uma batelada (causa especial de variação presente), a provável sinalização sobre a presença desta causa especial será gerada pelas cartas apenas após a realização completa da batelada, pois os escores $c_{i,f}$'s e o resíduo Q_f correspondente só poderão ser obtidos quando o vetor $\mathbf{x}'_f$ estiver totalmente preenchido; para tanto, é necessário que a batelada esteja terminada, configurando um monitoramento *off-line*. Esta batelada certamente estará comprometida e, em conseqüência disto, o produto obtido estará fora das especificações de qualidade. Dependendo do tempo gasto para a identificação e eliminação desta causa especial, algumas bateladas seguintes poderão ainda gerar um produto final inadequado.

Para minimizar este problema, introduz-se a idéia do monitoramento *on-line* dos processos em bateladas. Busca-se uma ferramenta para o monitoramento da trajetória das p variáveis de processo a cada instante transcorrido durante a batelada. Em outras palavras, deseja-se avaliar o comportamento de uma batelada futura a cada nova mensuração obtida das p

variáveis de processo. A questão é: dado que uma nova batelada foi iniciada, como gerar em cada instante l os escores $c_{i,f}$'s e o correspondente resíduo Q_f, visto que o vetor $\mathbf{x}'_f$ será completado apenas ao término da batelada (ou seja, ao final do instante h)?

A primeira estratégia possível não utiliza a ACPM, trabalhando diretamente com as variáveis originais ao invés dos CPs. Assume-se que as ph variáveis originais envolvidas na matriz $\mathbf{X}$ seguem uma distribuição Normal multivariada e obtém-se, para cada instante dentro da batelada futura, o valor T_f^2 referente às pl variáveis originais mensuradas até o instante l (a cada instante percorrido numa nova batelada, tem-se o vetor $\mathbf{x}'_f$ correspondente de dimensão $1 \times pl$, para l=1,...,h). Desta forma, verifica-se, a cada instante l transcorrido, se as pl variáveis apresentam desvios significativos em torno das suas médias. Em cada instante l, o vetor de médias amostrais $\bar{\mathbf{x}}'$, de dimensão $1 \times pl$, e a matriz de covariâncias amostrais $\mathbf{S}$, de dimensão $pl \times pl$ (distribuição de referência), são estimados a partir dos dados obtidos das pl variáveis nas m bateladas preliminares analisadas. Esta estratégia não é muito aconselhada, pois a cada instante transcorrido, tem-se um número maior de variáveis fortemente correlacionadas e colineares (o que implica redundâncias na representação do processo), tornando difícil a inversão da matriz de covariâncias, de dimensão $pl \times pl$, para o cálculo da estatística de Hotelling (Kourti & MacGregor, 1996).

Uma outra estratégia consiste em construir h modelos de referência via ACPM. Transcorrido o instante l (l=1,...,h) numa batelada, obtêm-se os valores $c_{i,f}$'s $T_{q,f}^2$ e Q_f através do modelo ACMP, construído a partir dos dados obtidos das pl variáveis, nas m bateladas preliminares analisadas. Nesta estratégia, para cada instante l, obtêm-se via ACPM os q autovetores $\mathbf{u}_i$ (cada um de dimensão $pl \times 1$) que representam as projeções das pl variáveis envolvidas no plano definido pelos q CPs retidos. Embora seja o mais adequado, este procedimento requer um elevado aparato computacional, a menos que a batelada tenha duração curta (h pequeno) e que o número p de variáveis de processo envolvidas seja pequeno (Nomikos & MacGregor, 1995).

A estratégia que será adotada para o monitoramento de uma batelada futura durante o seu desenvolvimento, baseia-se no mesmo modelo ACPM de referência, utilizado para o monitoramento *off-line*. A partir das m bateladas preliminares completadas, usam-se as

observações $\mathbf{x}'_f$, de dimensão $1 \times ph$, para a obtenção de uma única distribuição de referência para as ph variáveis envolvidas (isto é, para as p variáveis de processo, em h instantes amostrados).

Quando se realiza o monitoramento *on-line* de uma batelada a partir do modelo ACPM proposto, deve-se preencher de alguma forma a observação $\mathbf{x}'_f$ incompleta, pois necessitam-se dos q autovetores $\mathbf{u}_i$ completos (isto é, com dimensão $ph \times 1$) para que se possa obter a projeção dos dados da batelada no plano definido pelos CPs. A Fig. 30 ilustra a idéia do monitoramento *on-line* de uma batelada futura. Após transcorrido o $l^{ésimo}$ instante da batelada, tem-se a observação incompleta, de dimensão $1 \times pl$. Assim deve-se estimar $ph - pl$ dados que ainda não estão disponíveis para gerar o vetor de escores parciais $\mathbf{c}_{ql,f}$, $q \times 1$ (que contém o escore parcial c_{il} devido a cada um dos q CPs retidos no modelo ACPM, após l instantes transcorridos na batelada), o valor $T^2_{ql,f}$ de Hotelling e o resíduo $Q_{l,f}$, cada instante transcorrido.

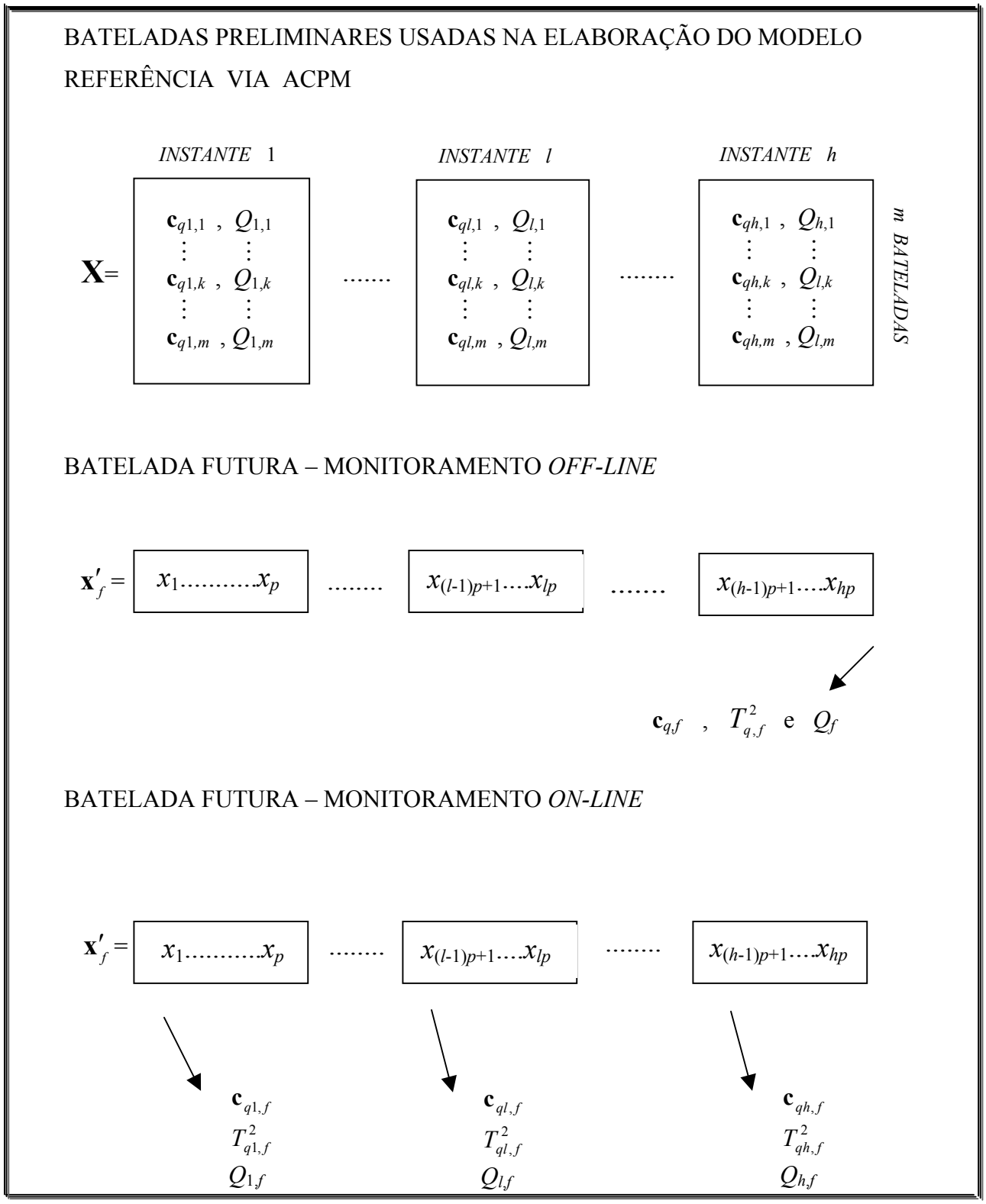

Figura 30. Esquema comparativo das estratégias de monitoramento *off-line* e *on-line* via ACPM

Nomikos & MacGregor (1995) apresentam três estratégias de preenchimento do vetor $\mathbf{x}'_f$ para monitoramento *on-line* da variabilidade das bateladas futuras através dos escores parciais $\mathbf{c}_{ql,f}$. Estas estratégias serão apresentadas a seguir. Estes autores também sugerem que o monitoramento *on-line* do comportamento da batelada através dos resíduos Q_f seja feito obtendo, a cada novo instante l transcorrido, o resíduo parcial $Q_{l,f}$, calculado apenas em função das mensurações mais recentes acerca das p variáveis de processo envolvidas (mensurações obtidas no último instante l transcorrido na batelada).

Seja $\mathbf{x}'_{l,f} = [x_{(l-1)p+1,f},....,x_{pl,f}]$ a parte da observação $\mathbf{x}'_f$ referente apenas ao instante l, para l=1,...,h. Reescreve-se a expressão (48) da seguinte forma:

$$Q_{l,f} = \sum_{i=(l-1)p+1}^{pl} (x_{i,f} - \hat{x}_{i,f})^2 , \qquad l=1,...,h , \tag{50}$$

onde $\hat{\mathbf{x}}'_{l,f} = [\hat{x}_{(l-1)p+1,f},....,\hat{x}_{pl,f}]$ representa parte da observação $\mathbf{x}'_f$ referente ao instante l, reescrita a partir do modelo de referência

$$\hat{\mathbf{x}}_f = \mathbf{U}_{ql}\mathbf{c}_{ql,f} . \tag{51}$$

A matriz $\mathbf{U}_{ql}$, de dimensão $l \times q$, contém em suas colunas as cargas dos q autovetores referentes apenas ao instante l e $\mathbf{c}_{ql,f}$ é o vetor, $q \times 1$, que contém os escores parciais c_{il}'s (referentes aos CPs retidos). No último instante l transcorrido da batelada, os q escores parciais c_{il} são obtidos através da expressão (46), utilizando o vetor $\mathbf{x}'_f$ completado nos $h - l$ instantes não ocorridos, segundo algum critério. A estatística $Q_{l,f}$ representa a distância perpendicular simultânea (isto é, considerando apenas a realização mais recente das p variáveis de processo) entre uma observação $\mathbf{x}_f$ ph-variada representada na sua dimensão original (ou seja, no plano ph-dimensional), e a sua representação no plano definido pelos q CPs retidos no modelo. A obtenção do resíduo parcial $Q_{l,f}$ utilizando os dados obtidos no instante l (informação mais recente), ao invés dos dados obtidos nos l instantes percorridos pela batelada, evita possíveis erros de avaliação sobre o comportamento da batelada, associados ao preenchimento do vetor $\mathbf{x}'_f$. Dessa forma, obtém-se uma estatística mais precisa para detectar a presença de um evento atípico no processo durante a batelada.

4.3.1- Preenchimento antecipado de uma observação $\mathbf{x}'_f$

Considere-se um processo químico por bateladas em que uma nova batelada esteja sendo monitorada através de medições periódicas das *p* variáveis de processo envolvidas. Dado que *l* instantes (para *l*=1,...,*h*) foram transcorridos, têm-se então *l* medições das *p* variáveis de processo, ou seja, uma observação $\mathbf{x}'_f$ de dimensão $1 \times pl$. Para realizar o monitoramento *on-line* desta batelada através do modelo ACPM para controle *off-line* (modelo de referência contendo as trajetórias médias das *p* variáveis de processo, obtido a partir das *m* observações preliminares $\mathbf{x}'_k$, $1 \times ph$, completas), precisa-se preencher $ph - pl$ dados na observação $\mathbf{x}'_f$. Para tanto, deve-se estimar, de alguma forma, os valores referentes às *p* variáveis nos $h - l$ instantes ainda não percorridos pela batelada. Três abordagens surgem como alternativa para resolução deste problema (Nomikos & MacGregor, 1995).

O primeira abordagem consiste em assumir que as *p* variáveis de processo seguirão suas trajetórias médias nos $h - l$ instantes futuros da batelada. Desta forma, deve-se preencher a observação $\mathbf{x}'_f$ com $ph - pl$ zeros (pois os dados estão *padronizados*, isto é, com média zero e desvio-padrão unitário); MacGregor & Nomikos, *apud* Nomikos & MacGregor (1995). Em outras palavras, assume-se que a batelada em estudo irá transcorrer exatamente dentro do padrão esperado até o seu término (isto é, nos $h - l$ instantes restantes). Através deste procedimento, pode-se detectar rapidamente a possível presença de um evento atípico no processo, pois tem-se uma carta $Q_{l,f}$ altamente sensível. Em contrapartida, os escores $\mathbf{c}_{ql,f}$ são poucos sensíveis à possível presença de uma variabilidade acentuada em alguma variável de processo, principalmente no início da batelada (Nomikos & MacGregor, 1995).

A segunda abordagem considera que as *p* variáveis de processo apresentarão, nos $h - l$ instantes restantes, desvios em relação às suas trajetórias médias semelhantes aos desvios encontrados no último instante verificado (isto é, no $l^{ésimo}$ instante). Assim, a cada novo instante *l* transcorrido, obtêm-se as medições *padronizadas* das *p* variáveis (cada valor subtraído da média e dividido pelo desvio-padrão da respectiva variável, ambos estimados a partir das *m* amostras preliminares das variáveis no $l^{ésimo}$ instante) e substitui-se os dados referentes às $h - l$

mensurações futuras destas variáveis pelos respectivos desvios obtidos no $l^{ésimo}$ instante; Nomikos & MacGregor (1994). Através deste procedimento, obtém-se uma carta $Q_{l,f}$ menos sensível, se comparada àquela do primeiro procedimento. Em contrapartida, através dos escores $\mathbf{c}_{ql,f}$ pode-se detectar rapidamente a possível presença de variabilidade acentuada na batelada (Nomikos & MacGregor, 1995).

A terceira abordagem considera que os dados futuros acerca dos $h - l$ instantes restantes da batelada, representam valores perdidos (*missing values*) de uma batelada utilizada na construção do modelo ACPM. Parte-se do princípio que os dados restantes seguirão o padrão apresentado nos l instantes observados, e que a estrutura de correlação das variáveis de processo, nos $h - l$ instantes restantes, estará de acordo com a estrutura de correlação capturada a partir das m bateladas preliminares. Desta forma, utilizam-se os q autovetores $\mathbf{u}_i$ (associados aos q CPs retidos no modelo), considerando apenas as pl primeiras cargas, associadas às p variáveis de processo nos l (l=1,...,h) instantes transcorridos, para estimar os escores parciais $\mathbf{c}_{ql,f}$. Para o $l^{ésimo}$ instante disponível, projeta-se a observação $\mathbf{x}'_f$, de dimensão $1 \times pl$, no plano definido pelos q CPs retidos, obtendo o vetor $\mathbf{c}_{ql,f}$ da seguinte forma:

$$\mathbf{c}_{ql,f} = (\mathbf{U}'_{q,l}\mathbf{U}_{q,l})^{-1}\mathbf{U}'_{q,l}\mathbf{x}_f,$$

onde a matriz $\mathbf{U}_{q,l}$, de dimensão $pl \times q$, contém em suas colunas as pl cargas referentes aos respectivos autovetores $\mathbf{u}_i$, para i=1,...,q. Esta abordagem parece superior às demais abordagens apresentadas quando mais que 10% da batelada já tiver transcorrido. No entanto este procedimento pode gerar escores parciais $\mathbf{c}_{ql,f}$ com variabilidade acentuada no início de uma nova batelada, pois existe pouca informação disponível para realizar as estimativas através do modelo proposto (Nomikos & MacGregor, 1995).

De uma maneira geral, a escolha da melhor abordagem, dentre as apresentadas acima, depende das características específicas do processo em estudo. Pode-se até mesmo usar mais de uma abordagem no monitoramento de uma batelada futura, combinando as vantagens apresentadas em cada método.

4.3.2- Limites de controle para os escores parciais $\mathbf{c}_{ql,f}$ e para os resíduos $Q_{l,f}$

Na seção 4.2, foram apresentados os limites de controle para os escores $\mathbf{c}_{q,f}$, a estatística de Hotelling $T^2_{q,f}$ e para o resíduo Q_f, utilizados no monitoramento *off-line* via ACPM de uma batelada futura. Evidenciou-se que estes limites eram os mesmos adotados no monitoramento via ACP de processos contínuos (Capítulo 3). No entanto, no monitoramento *on-line* de uma batelada futura, os limites de controle utilizados para os escores parciais $\mathbf{c}_{ql,f}$ e para o resíduo parcial $Q_{l,f}$ são diferentes em relação ao monitoramento *off-line*, pois devem ser obtidos a partir da distribuição de referência das p variáveis em cada instante l da batelada.

Sabe-se que cada um dos q escores c_{il} pertencentes ao vetor $\mathbf{c}_{ql,f}$ de escores parciais representa um combinação linear de ph variáveis aleatórias, com $ph - pl$ variáveis estimadas segundo alguma das abordagens apresentadas acima. Desta forma, através do *teorema do limite central*, admite-se que a distribuição destes escores é bem aproximada pela distribuição Normal, exceto no início de uma nova batelada. Neste momento, a maior parte da observação $\mathbf{x}'_f$ está sendo estimada de alguma forma, provocando desvios nos referentes escores $c_{il,f}$'s em relação ao modelo Normal (Nomikos & MacGregor, 1995). Admitindo uma distribuição Normal para os escores, obtêm-se os limites de controle para um escore $c_{il,f}$ (para i=1,...,q e l=1,...,h) da seguinte forma:

$$\text{LSC} = + \boldsymbol{T}_{m\text{-}1,\alpha/2}\, s_{il}\, \sqrt{(1+1/m)}\ ,$$

$$\text{LC} = 0\ ,$$

$$\text{LIC} = - \boldsymbol{T}_{m\text{-}1,\alpha/2}\, s_{il}\, \sqrt{(1+1/m)}\ , \qquad (52)$$

onde $\boldsymbol{T}_{m\text{-}1,\alpha/2}$ representa o percentil de uma distribuição de Student, com m-1 graus de liberdade, correspondente à α de alarme falso adotada. O valor s_{il} representa o desvio-padrão do escore $c_{il,f}$ (referente ao $l^{ésimo}$ instante), estimado a partir dos m escores $c_{il,k}$, obtidos a partir das m bateladas preliminares completadas através de alguma das abordagens na seção 4.3.1 (ver Fig. 30).

Na prática, faz-se o monitoramento utilizando os escores padronizados. Assim, dado uma observação futura $\mathbf{x}'_f$, cada escore $c_{il,f}$ (pertencente ao vetor $\mathbf{c}_{ql,f}$) é subtraído de sua média (que é zero) e dividido pelo seu desvio-padrão $[s_{il}\sqrt{(1+1/m)}]$. Os limites de controle para o escore padronizado

$$y_{il,f} = \frac{c_{il,f}}{s_i\sqrt{1+1/m}}, \tag{53}$$

com $c_{il,f} = \mathbf{u}'_i\mathbf{x}_f$ [expressão (46) utilizando a batelada $\mathbf{x}'_f$ completada através de algum critério] são dados por:

$$\text{LSC} = + \boldsymbol{T}_{m\text{-}1,\alpha/2} \ ,$$

$$\text{LC} = 0 \ ,$$

$$\text{LIC} = - \boldsymbol{T}_{m\text{-}1,\alpha/2} \ . \tag{54}$$

Para manter a probabilidade de alarme falso próxima de α, obtêm-se os limites de controle para cada uma das q cartas substituindo $\alpha/2$ por $\alpha/2q$ (limites de Bonferroni, seção 2.2.7.1), na expressão (48).

Os valores parciais $T^2_{ql,f}$ de Hotelling são calculados, através da expressão (47), em função dos escores parciais $\mathbf{c}_{ql,f}$ ao invés dos escores padronizados $\mathbf{y}_{q,f}$ [expressão (41)] ou $\mathbf{y}_{ql,f}$ [expressão (53)], isto é

$$T^2_{ql,f} = \sum_{i=1}^{q} \frac{c^2_{il,f}}{\lambda_i} \ . \tag{55}$$

Isto ocorre pelo fato de que a estatística de Hotelling utiliza a matriz de covariâncias estimadas em função de m bateladas preliminares completas. Escrevendo esta estatística em função dos q CPs, sabe-se que a referente matriz de covariâncias é diagonal com λ_i representando a variância estimada do $i^{ésimo}$ CP. Desta forma, o desvio-padrão do $i^{ésimo}$ CP, estimado através das m bateladas preliminares completadas bem sucedidas, é dado por $\sqrt{\lambda_i}$ (raíz quadrada do $i^{ésimo}$

autovalor). Em contrapartida, quando se utiliza os valores parciais $T^2_{ql,f}$, obtidos em função dos escores parciais $\mathbf{c}_{ql,f}$, o desvio-padrão do $i^{ésimo}$ CP estimado através das m bateladas preliminares consideradas até o $l^{ésimo}$ instante (l=1,...,h) é dado por $s_{il}\sqrt{(1+1/m)}$. Desta forma, obtêm-se os valores $T^2_{ql,f}$ em função dos escores originais $\mathbf{c}_{ql,f}$.

Os limites de controle usados na carta de Hotelling em função dos escores parciais $\mathbf{c}_{ql,f}$ são os mesmos utilizados no monitoramento *off-line*, sendo assim obtidos através da expressão (40).

Os limites de controle apresentados na expressão (44), utilizados no monitoramento *off-line* dos resíduos Q_f [expressão (48)], foram derivados da distribuição das *formas quadráticas* (Jackson & Mudholkar, 1979). Sob suposição de uma distribuição Normal multivariada para as ph variáveis envolvidas na batelada (p variáveis em h instantes distintos), obtêm-se os limites de controle aproximados para a estatística Q_f baseados nesta distribuição. Os limites de controle utilizados no monitoramento *on-line* dos resíduos parciais $Q_{l,f}$, também baseiam-se nos resultados obtidos da distribuição das formas quadráticas. No entanto, supondo distribuição Normal multivariada das ph variáveis envolvidas, a distribuição dos resíduos parciais $Q_{l,f}$, em cada instante l, é bem aproximada pela distribuição do Qui-Quadrado ponderada; Box, *apud* Nomikos & MacGregor (1995). Desta forma, os limites de controle para a carta $Q_{l,f}$, para l=1,...,h, são os seguintes:

$$\text{LSC} = a\chi^2_{\alpha,b} \quad ,$$

$$\text{LC} = 0 \ ,$$

onde $\chi^2_{\alpha,b}$ representa o percentil de uma distribuição do Qui-Quadrado, com b graus de liberdade e parâmetro de ponderação a, correspomdente à probabilidade α de alarme falso adotada. Os parâmetros a e b são obtidos a partir dos autovetores da matriz $\boldsymbol{\Sigma_0}$ de covariâncias das ph variáveis originais envolvidas. Entretanto existe uma maneira simples de estimar estes parâmetros, combinando o 1° e 2° momentos da distribuição $a\chi^2_{\alpha,b}$ e da distribuição amostral dos resíduos parciais $Q_{l,k}$ (k=1,...,m), referentes ao $l^{ésimo}$ instante das m bateladas preliminares (distribuição de referência ilustrada na Fig. 30). A distribuição $a\chi^2_{\alpha,b}$ possui média $\mu = ab$ e

variância $\sigma^2 = a^2 b$. Estes valores são igualados respectivamente à média w e variância v dos resíduos parciais $Q_{l,k}$, obtidos em cada instante l. Desta forma, os limites de controle para a carta Q_l assumem a seguinte configuração (Nomikos & MacGregor, 1995):

$$\text{LSC} = \frac{v}{2w}\chi^2_{\alpha,2w^2/v} \quad ,$$

$$\text{LC} = 0 \; . \tag{56}$$

Salienta-se que os valores de a e b, estimados em cada instante, trazem informação sobre a natureza da distribuição dos resíduos parciais $Q_{l,k}$ durante as m bateladas preliminares conformes amostradas. Valores estimados baixos para b (isto é, limites de controle Qui-Quadrado estreitos na carta dos resíduos parciais) indicam que, nos instantes correspondentes, a distribuição está dominada pela elevada variabilidade apresentada por algumas variáveis de processo em torno de suas trajetórias médias, refletida na variabilidade acentuada dos escores $\mathbf{c}_{ql,f}$ (limites de controle alargados das cartas para os escores). Em contrapartida, valores estimados elevados para b (limites de controle Qui-Quadrado alargados) ocorrem em períodos mais estáveis dentro das bateladas (isto é, pouca variação nos escores $\mathbf{c}_{ql,f}$); nestes instantes, pequenos desvios nas variáveis em torno das suas trajetórias refletem diretamente nos resíduos $Q_{l,k}$ correspondentes. O valor a é utilizado apenas para obter os limites de controle aproximados da distribuição do Qui-Quadrado em função dos resíduos parciais.

Em algumas situações, dispõe-se de um número m reduzido de bateladas preliminares na distribuição de referência, acarretando uma variabilidade acentuada dos limites de controle para os escores parciais e resíduos parciais, devido ao pequeno número de observações na distribuição de referência. No entanto, num processo em bateladas, durante o transcurso de uma batelada conforme as p variáveis de processo geralmente não apresentam mudanças bruscas nas suas trajetórias médias em instantes próximos. Para tratar tal problema, sugere-se que os limites de controle para $\mathbf{c}_{ql,f}$ e $Q_{l,f,}$ sejam estimados não apenas em função dos m valores $\mathbf{c}_{ql,k}$ e Q_{lkf} disponíveis no $l^{ésimo}$ instante, mas também em função dos valores disponíveis em instantes imediatamente anteriores e posteriores (Nomikos & MacGregor, 1995). Desta forma, obtêm-se limites de controle mais estáveis em cada instante. Por exemplo, pode-se obter os limites de controle para os valores parciais $\mathbf{c}_{ql,f}$ e $Q_{l,f}$ no $l^{ésimo}$ instante, considerando os valores parciais $\mathbf{c}_{\boldsymbol{ql,k}}$ e Q_{lkf} obtidos no instante l e nos 2 instantes anteriores e posteriores. Neste caso, utilizando uma

janela móvel de 5 instantes, calcula-se, em cada instante l, os limites de controle para os escores e resíduos em função de $5m$ valores parciais.

Realiza-se, então, o monitoramento *on-line* de uma batelada futura obtendo os escores parciais $\mathbf{c}_{ql,f}$ [expressão (46) com $\mathbf{x}'_f$ completado através de algum critério], e, comparando, a cada instante l transcorrido, os referentes q escores padronizados parciais $\mathbf{y}_{ql,f}$, o valor $T^2_{ql,f}$ de Hotelling e o resíduo parcial $Q_{l,f}$ calculados [expressões (53), (55) e (50), respectivamente], com os respectivos limites de controle apresentados nas expressões (54), (40) e (56). Um elevado valor de $T^2_{ql,f}$ indica que a variabilidade apresentada na batelada, no instante l, está além do esperado. Através das cartas de controle para os q escores parciais $\mathbf{y}_{ql,f}$ e dos Gráficos de Contribuição (seção 3.4.4), pode-se buscar um diagnóstico para este sinal, determinando quais as variáveis que, naquele instante, apresentam desvios significativos em relação às suas trajetórias médias. Um alto resíduo $Q_{l,f}$ indica que o modelo ACPM não está mais descrevendo satisfatoriamente a estrutura de correlação das p variáveis de processo, no $l^{ésimo}$ instante, o que indica a provável presença de um evento atípico no processo. Novamente, através dos Gráficos de Contribuição, busca-se diagnosticar este sinal determinando o grupo de variáveis que mais contribuíram para o resíduo elevado encontrado.

4.3.3- Controle do erro tipo I (probabilidade de alarme falso) no monitoramento *on-line*

Um questão importante no monitoramento *on-line* é estabelecer o valor adequado para o erro tipo I. Dada uma nova observação $\mathbf{x}'_f$, em cada instante, os q escores parciais $\mathbf{c}_{ql,f} = [c_{1l,f},..., c_{il,f},..., c_{ql,f}]'$, e os resíduos $Q_{l,f}$ seguem uma distribuição subjacente distinta, através da qual os respectivos limites de controle são obtidos. Considerando-se apenas um instante l, admite-se que a probabilidade de alarme falso associada às cartas de controle para $\mathbf{c}_{ql,f}$ e $Q_{l,f}$ seja α. No entanto, quando monitora-se uma batelada obtendo-se os valores $\mathbf{c}_{ql,f}$ e $Q_{l,f}$ em h instantes sucessivos, não é correto admitir que α represente a chance total de que uma determinada carta

gere um alarme falso durante a realização completa da referente batelada (ver Montgomery, 1996; Rencher, 1995, entre outros).

Admitindo-se que os h valores parciais obtidos de uma batelada representam medições de variáveis independentes, então $1-(1-\alpha)^h$ representa a probabilidade total de alarme falso associada à respectiva carta de controle. Entretanto, como essa suposição não é verdadeira, a efetiva probabilidade total de alarme falso só poderia ser encontrada conhecendo-se a distribuição conjunta de $\mathbf{c}_{ql}$ e Q_l, em relação a todos os instantes verificados, o que não é, via de regra, o caso.

Assim, visando obter uma medida global de erro tipo I mais apropriada para as cartas de controle dos valores parciais $\mathbf{c}_{ql,f}$ e $Q_{l,f}$, adota-se um procedimento que utiliza os escores parciais $\mathbf{c}_{ql,k} = [c_{1,k},...,c_{i,k},...,c_{q,k}]'$ e $Q_{l,k}$ obtidos a partir das m bateladas preliminares conformes, utilizadas na distribuição de referência das variáveis de processo (Fig. 30). Em cada instante l (para l=1,...,h), compara-se os valores parciais $\mathbf{c}_{ql,f}$ (referentes ao vetor $\mathbf{x}'_{k,f}$, preenchido de acordo com algum critério) e $Q_{l,k}$ (para k=1,...,m) com os respectivos limites de controle [expressões (52) e (56)], conta-se a freqüência de valores além dos limites de controle, e divide-se o total obtido pelo número mh total de valores. Os resultados obtidos representam estimativas da probabilidade de alarme falso para as q cartas de controle, referentes aos escores parciais $\mathbf{c}_{ql,f}$, e para a carta de controle dos resíduos parciais $Q_{l,f}$. O erro total tipo I para a carta dos resíduos parciais estimado através deste critério está próximo do valor nominal α, tanto para α=5%, quanto para α=1% (valores de significância mais utilizados na prática). Em relação aos q escores $c_{i,f}$, as probabilidades de alarme falso associadas às cartas de controle estão próximas do valor nominal α, quando α=5%, porém, e afastadas, quando adota-se α=1% (Nomikos & MacGregor, 1995).

CAPÍTULO 5

5- ESTUDO DE CASO

5.1 – Apresentação do Exemplo

Neste capítulo, será realizada uma aplicação da ACPM no monitoramento de uma operação em batelada integrante do sub-processo de fabricação de borracha para cobertura de pneus. Para tanto, dispõe-se de dados que descrevem as trajetórias de duas variáveis desta operação: *Energia Integrada* (X_1) e *Temperatura da massa* (X_2), durante 22 bateladas completas. Em cada batelada, foram obtidas medições simultâneas de ambas as variáveis, em intervalos fixos de 20 segundos, durante 5 minutos (tempo de duração das bateladas), perfazendo um total de 15 medições consecutivas de cada variável.

Através das Fig. 31 e 32, pode-se observar, respectivamente, as trajetórias das variáveis X_1 e X_2 nas 22 bateladas preliminares disponíveis. Ressalta-se que para aplicar a ACPM, deve-se

considerar, em cada instante, cada uma das duas variáveis de processo envolvidas como sendo uma variável aleatória distinta. Como cada variável possui 15 replicações dentro da batelada, tem-se então 30 variáveis fortemente correlacionadas. Aplicando a ACP nestas variáveis, captura-se, a partir das trajetórias descritas nas Fig. 31 e 32, a estrutura de correlação e autocorrelação dessas variáveis ao longo das bateladas preliminares bem sucedidas, obtendo assim a trajetória média de cada variável.

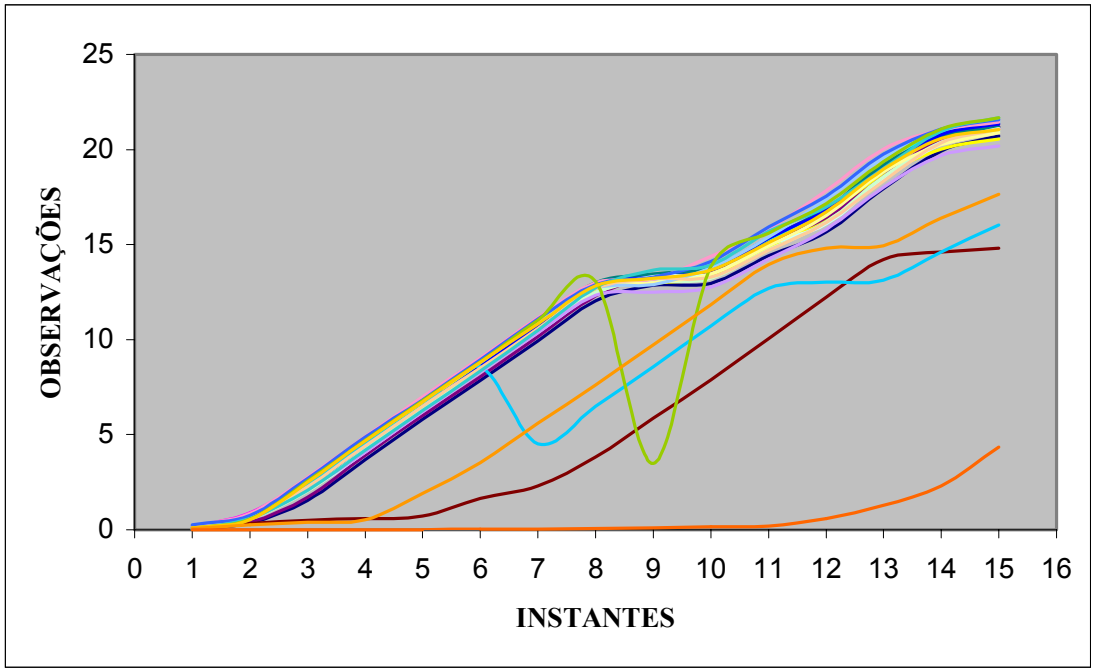

Figura 31. Trajetórias da variável X_1 observadas em 15 instantes, nas 22 bateladas preliminares.

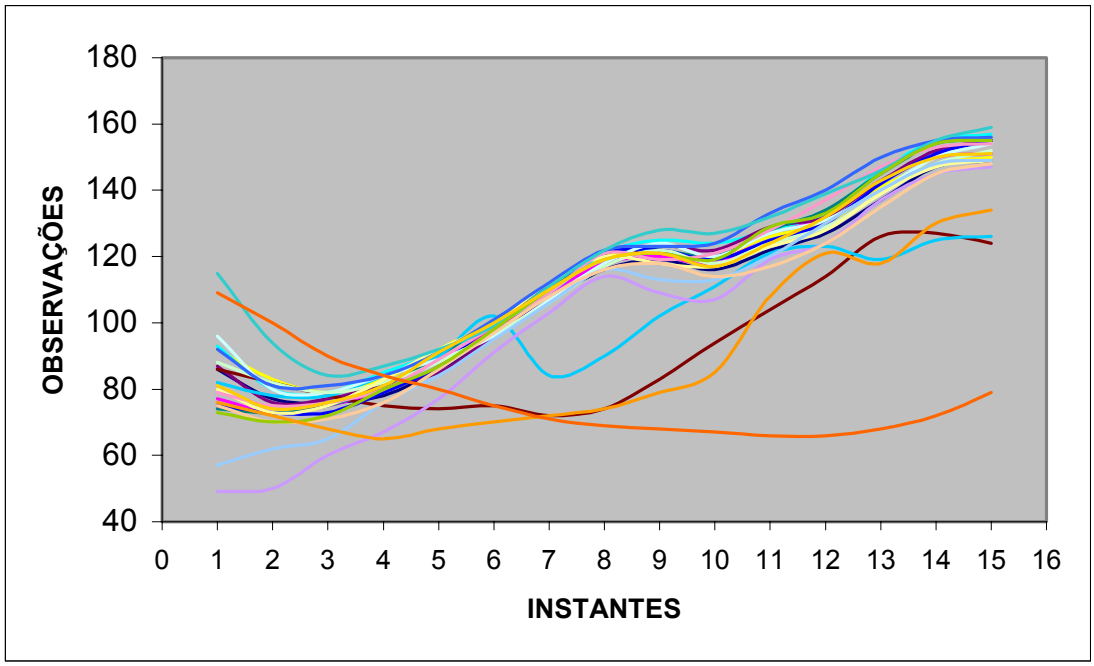

Figura 32. Trajetórias da variável X_2 observadas em 15 instantes, nas 22 bateladas preliminares.

5.2 – Construção do modelo ACPM de referência

Aplicando a idéia do CEP tradicional, na Primeira fase do controle, deve-se usar a ACPM para selecionar a distribuição de referência para as variáveis envolvidas, isto é, encontrar as bateladas nas quais as variáveis X_1 e X_2 apresentam variabilidade em torno das suas trajetórias médias devido apenas a causas comuns de variação.

Inicialmente, tem-se então a matriz $\underline{\mathbf{X}}$, $m \times p \times h$, tridimensional (22 bateladas × 2 variáveis × 15 instantes) que é desdobrada na matriz $\mathbf{X}$, $m \times ph$, bidimensional (22 bateladas × 30 variáveis). Cada linha da matriz $\mathbf{X}$ contém uma observação $\mathbf{x}_k'$ que representa a realização completa das variáveis X_1 e X_2 na batelada k (k=1,...,m).

Após a padronização dos dados da matriz $\mathbf{X}$ (seções 3.4.1 e 2.1.2), selecionam-se a partir do critério da *validação cruzada* (seção 3.4.2.2) os quatro primeiros CPs. Considera-se então que

estes CPs oferecem uma caracterização suficiente da estrutura de correlação das 30 variáveis envolvidas (ou seja, das variáveis X_1 e X_2 nos quinze instantes).

Utilizando as expressões (28), (29) e (30) calculam-se então os quatro maiores autovalores (λ_1, λ_2, λ_3 e λ_4) e os autovetores associados ($\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ e $\mathbf{u}_4$). Estes autovalores explicam juntos 96% da variabilidade existente nos dados preliminares. A partir das expressões (46) e (41), obtêm-se os escores $y_{i,k}$ ($y_{1,k}$, $y_{2,k}$,$y_{3,k}$ e $y_{4,k}$) para cada uma das 22 bateladas preliminares.

Calculado o valor da estatística de Hotelling [expressão (47)] em relação as 22 bateladas, e o limite de controle LSC através da expressão (39), obtém-se então a carta de Hotelling, em relação aos quatro CPs retidos, ilustrada na Fig. 33. Observa-se que as bateladas 6, 21 e 22 apresentam escores acima do LSC (pontos vermelhos). Isto indica que, a partir da projeção destas bateladas no plano definido pelos quatro CPs, existem evidências de que a variação ocorrida nas variáveis de processo, nestas bateladas, esteve acima do tolerado, dado que apenas causas comuns estão presentes no processo. Em outras palavras, deve-se entender que, nestas bateladas, as trajetórias das variáveis X_1 e/ou X_2 tiveram variabilidade acentuada, apresentando desvios significativos em relação às suas respectivas trajetórias conforme dadas pela distribuição de referência.

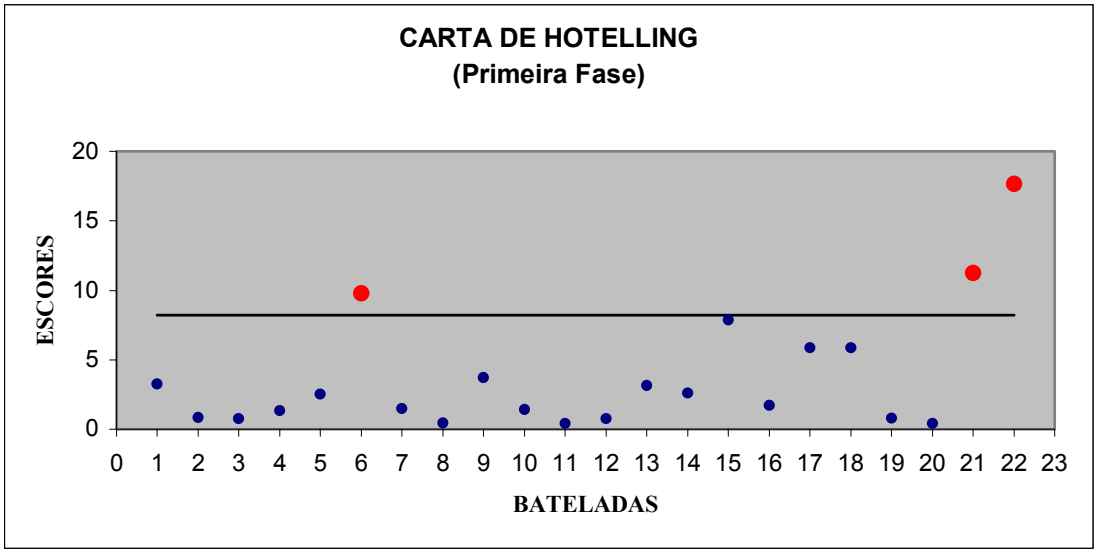

Figura 33. Carta de Hotelling em relação às 22 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

Através da carta Q ilustrada na Fig. 34 [expressões (49) e (48)], observa-se que as bateladas 9 e 19, embora dentro do padrão na carta de Hotelling, apresentam escores acima do LSC [expressão (44)]. Estes resíduos significativos indicam que o modelo construído em função dos quatro primeiros CPs não oferece uma representação suficiente do comportamento das variáveis envolvidas dentro das bateladas 9 e 19. Dessa forma, existem evidências de que algum evento incomum ocorreu no processo durante essas bateladas ou de que as variáveis X_1 e/ou X_2 apresentaram variabilidade acentuada num determinado instante onde existe pouca variabilidade no processo. Esta última situação faz com que desvios acentuados sejam multiplicados por autovalores com cargas baixas, gerando assim um valor T_k^2 não significativo (ver seção 4.2).

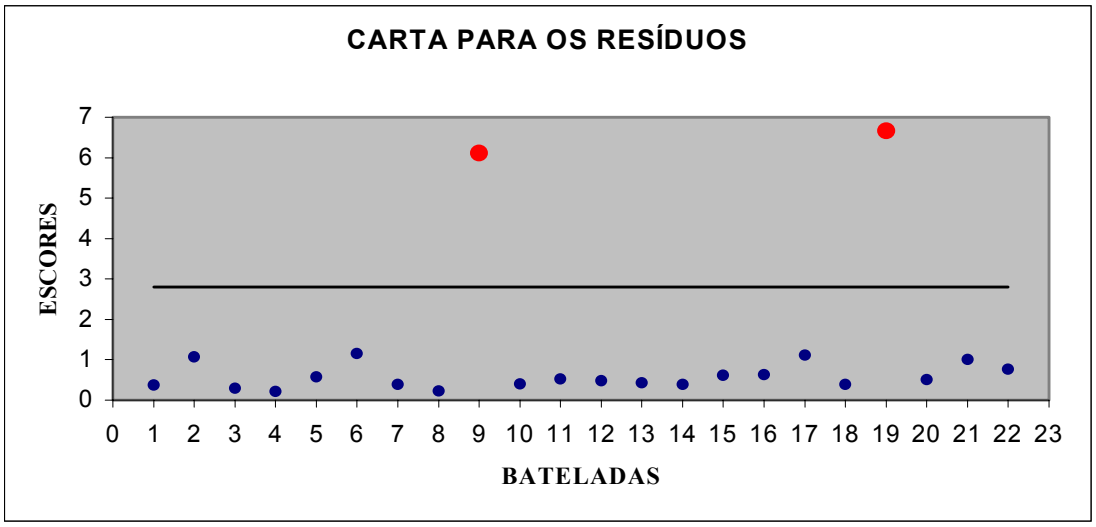

Figura 34. Carta Q em relação às 22 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

Após esta análise, considera-se que o processo esteve fora de controle estatístico durante a realização das bateladas 6, 9, 19, 21 e 22, pois estas bateladas não representam o padrão de variação das variáveis X_1 e X_2. Assim, essas bateladas serão excluídas da distribuição de referência por não caracterizarem suficientemente a estrutura de correlação e autocorrelação das duas variáveis envolvidas.

Realiza-se novamente a ACP na matriz **X**, de dimensão 17 × 30, de dados referentes às 17 bateladas preliminares restantes (com os dados novamente padronizados). Obtêm-se

novamente os quatro maiores autovalores e autovetores associados. Salienta-se que estes autovalores reunidos explicam 94% da variabilidade existente nos dados preliminares.

A partir da carta de Hotelling ilustrada na Fig. 35, observa-se que a batelada 15 apresenta escore acima de LSC. Conclui-se que esta batelada também apresentou variabilidade acima do tolerado, dado que apenas causas comuns de variação incidiam sobre o processo.

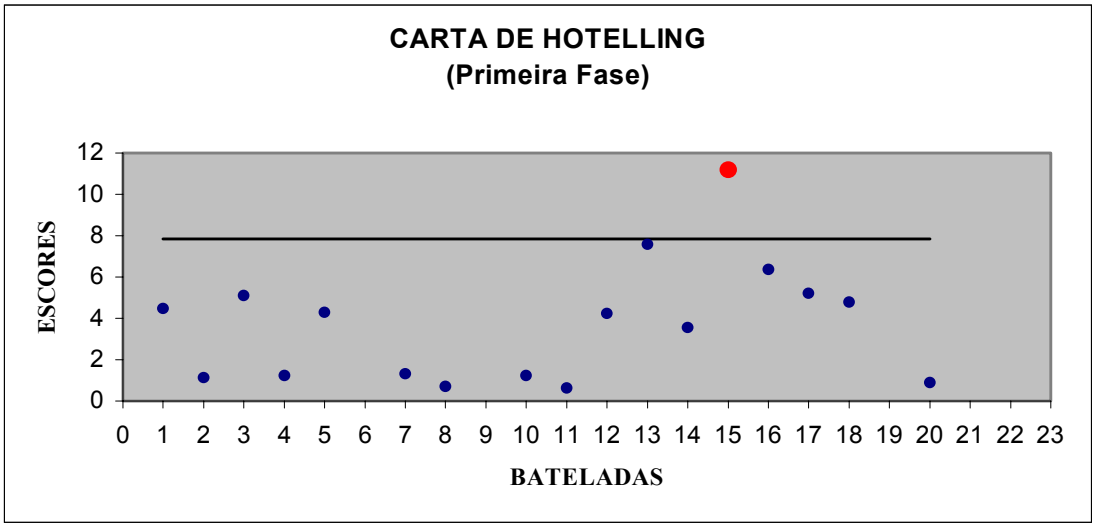

Figura 35. Carta de Hotelling em relação às 17 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

Observando a Fig. 36, conclui-se que nenhuma das 17 bateladas restantes apresenta resíduos significativos. Dessa forma, retira-se a batelada 15 da distribuição de referência e procede-se novamente a ACP em relação às 16 bateladas restantes (dados padronizados).

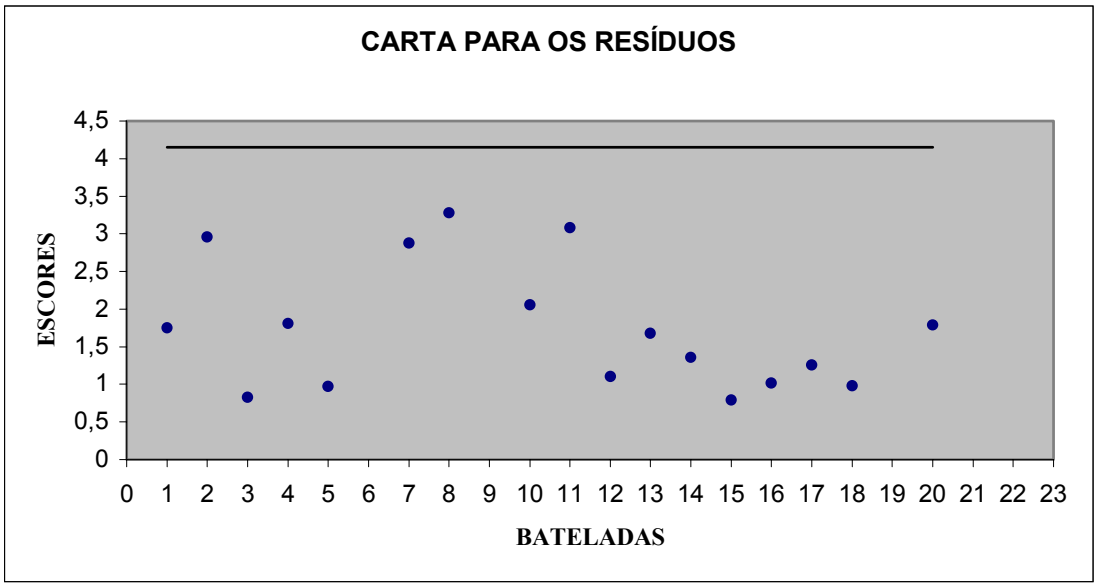

Figura 36. Carta *Q* em relação às 17 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

Mais uma vez, retêm-se quatro CPs para compor a distribuição de referência; estes CPs explicam 93% da variabilidade existente nos dados preliminares. A partir da carta de Hotelling ilustrada na Fig. 37, observa-se que a batelada 13 apresentou variabilidade acima do tolerado, devendo assim também ser descartada. Já a carta *Q* ilustrada na Fig. 38 não fornece resíduos significativos em relação a estas bateladas.

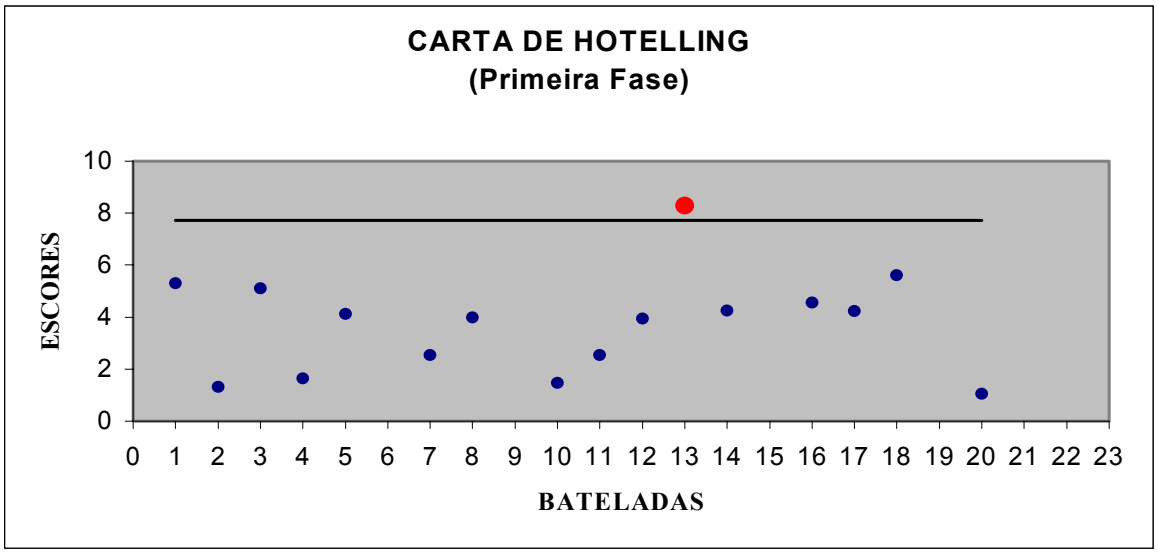

Figura 37. Carta de Hotelling em relação às 16 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

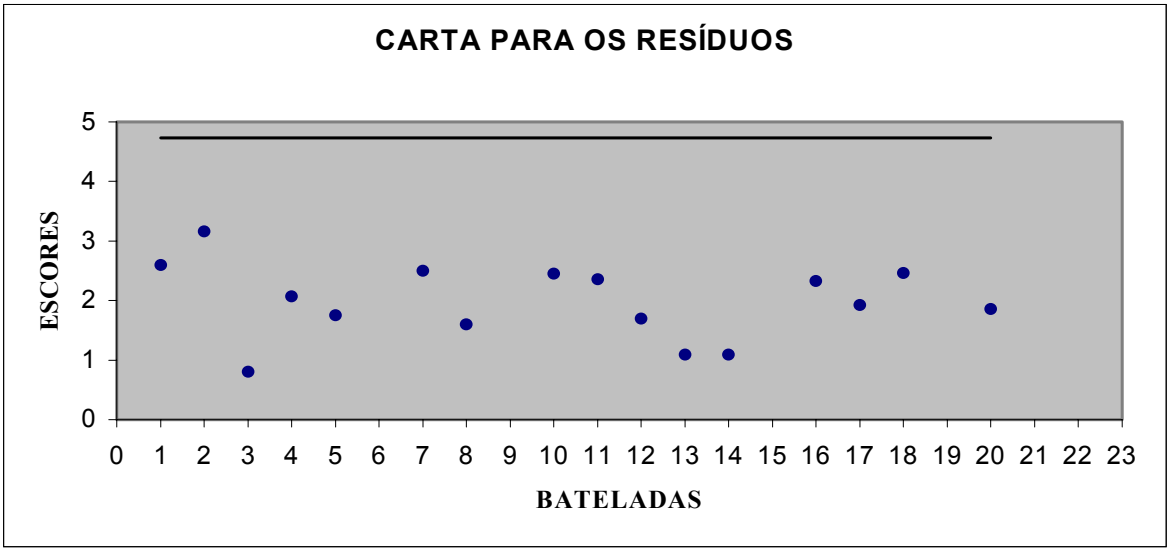

Figura 38. Carta Q em relação às 16 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

Finalmente, aplicando a ACP na matriz **X**, de dimensão 15 × 30, de dados padronizados referentes às 15 bateladas preliminares restantes, observa-se através da carta de Hotelling (Fig. 39) e da carta Q (Fig. 40) que nenhuma das bateladas apresentou variabilidade ou resíduo acima do tolerado.

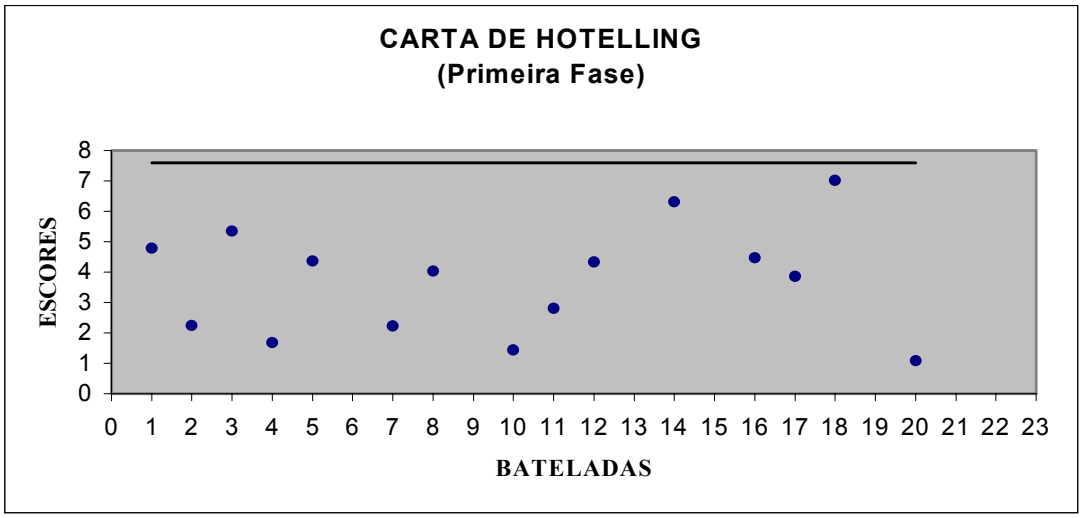

Figura 39. Carta de Hotelling em relação às 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

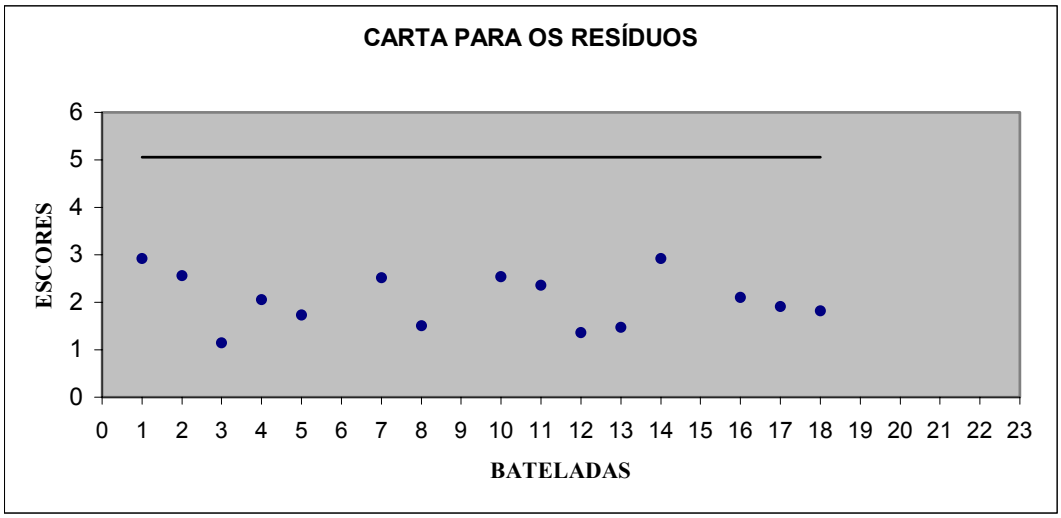

Figura 40. Carta *Q* em relação às 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%).

Conclui-se que estas 15 bateladas oferecem uma caracterização satisfatória da distribuição de referência das 30 variáveis de processo, isto é, apresentam as trajetórias padrão das variáveis X_1 e X_2. Observando as Fig. 41 e 42, constata-se que nas 15 bateladas retidas, as variáveis apresentam trajetórias consistentes se comparadas às suas respectivas trajetórias descritas nas 22 bateladas iniciais (Fig. 31 e 32).

Os quatro primeiros autovalores (λ_1, λ_2, λ_3 e λ_4) e os autovetores associados ($\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ e $\mathbf{u}_4$), obtidos a partir das 15 bateladas com variáveis apresentadas nas Fig. 41 e 42, serão usados no monitoramento de futuras bateladas deste processo (Segunda fase). Deve-se entender que as cargas destes autovetores refletem a estrutura de correlação e autocorrelação das variáveis X_1 e X_2 no processo sob controle estatístico, compondo assim o modelo ACPM de referência [expressões (45) e (49)]. Ressalta-se que os quatro autovalores obtidos explicam, em conjunto, 92,6% da variabilidade existente nos dados preliminares.

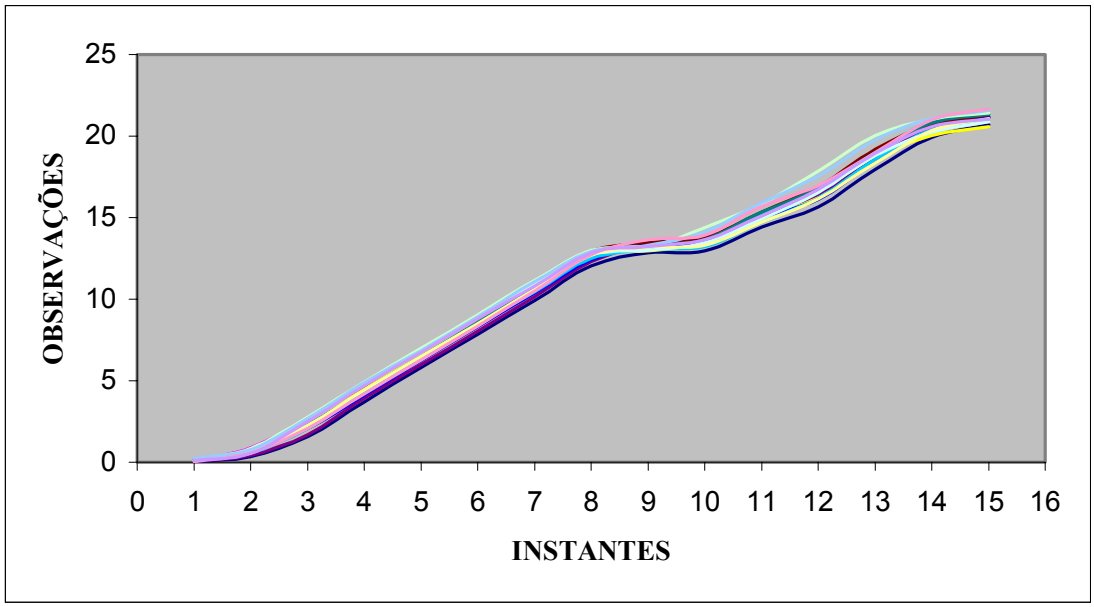

Figura 41. Trajetória-padrão da variável X_1 observadas em 15 instantes, durante 15 bateladas preliminares.

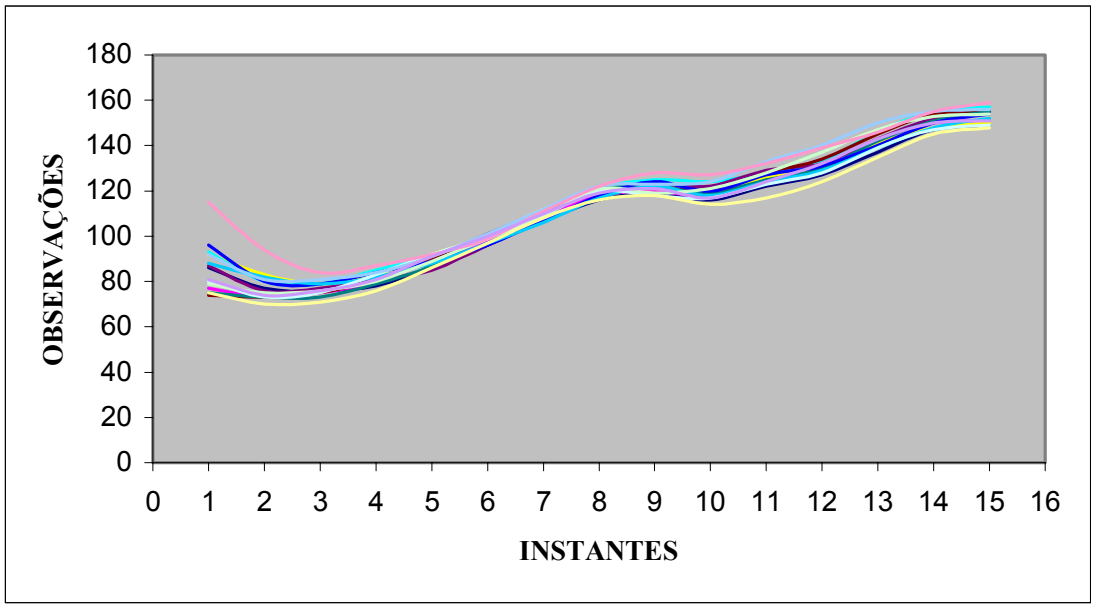

Figura 42. Trajetória-padrão da variável X_2 observadas em 15 instantes, durante 15 bateladas preliminares.

Nesta fase, os limites de controle LSC para as carta de Hotelling e para a carta Q [expressões (47) e (48)] são obtidos respectivamente através das expressões (40) e (44) (Fig. 43 e 44). Destaca-se que o modelo ACPM de referência foi obtido apenas na quarta triagem. Isto ocorreu porque algumas bateladas apresentaram variabilidade muito além do padrão observado

nas bateladas restantes (observar Fig. 31 e 32). Em outras palavras, deve-se entender que as variáveis X_1 e X_2 apresentaram trajetórias muito discrepantes em relação às suas respectivas trajetórias, conforme observadas nas demais bateladas observadas. Esta variabilidade acentuada gera um LSC elevado para a carta de Hotelling, fazendo com que uma determinada batelada com variabilidade não tão acentuada mais ainda indesejável não apresente escore acima deste limite. Na medida em que as bateladas com variabilidade mais elevada vão sendo descartadas, os LSC subseqüentes assumem valores menores, tornando assim a carta de Hotelling mais sensível diante de uma batelada com variabilidade ainda significativa.

5.3 - Monitoramento off-line de uma batelada futura

Na Primeira fase do monitoramento, 7 bateladas foram excluídas para obtenção da distribuição de referência das variáveis X_1 e X_2. Admite-se agora que uma das bateladas descartadas seja uma batelada futura. Considere-se, por exemplo, a batelada 6 como sendo uma nova batelada recém finalizada. O modelo ACPM, construído em função de 15 bateladas preliminares, será então utilizado para monitorar a variabilidade apresentada nesta batelada. Dado que existem 15 bateladas na distribuição de referência e assumindo que a batelada 6 é a primeira batelada obtida na fase de controle, esta será assim representada por *batelada 16*.

Projetando a observação $\mathbf{x}'_{16}$ no plano definido pelos quatro primeiros CPs, observa-se através da carta de Hotelling e a carta Q (Fig. 43 e 44, respectivamente) que a referente batelada apresentou variabilidade e resíduo além do tolerando (isto é, $T^2_{4,16}$ > LSC e Q_{16} > LSC). Conclui-se que as variáveis apresentaram desvios acentuados em relação às suas trajetórias médias durante a batelada. Existem também evidências de que algum evento incomum esteve presente no processo devido ao elevado resíduo obtido. Dessa forma, essa batelada é considerada não conforme.

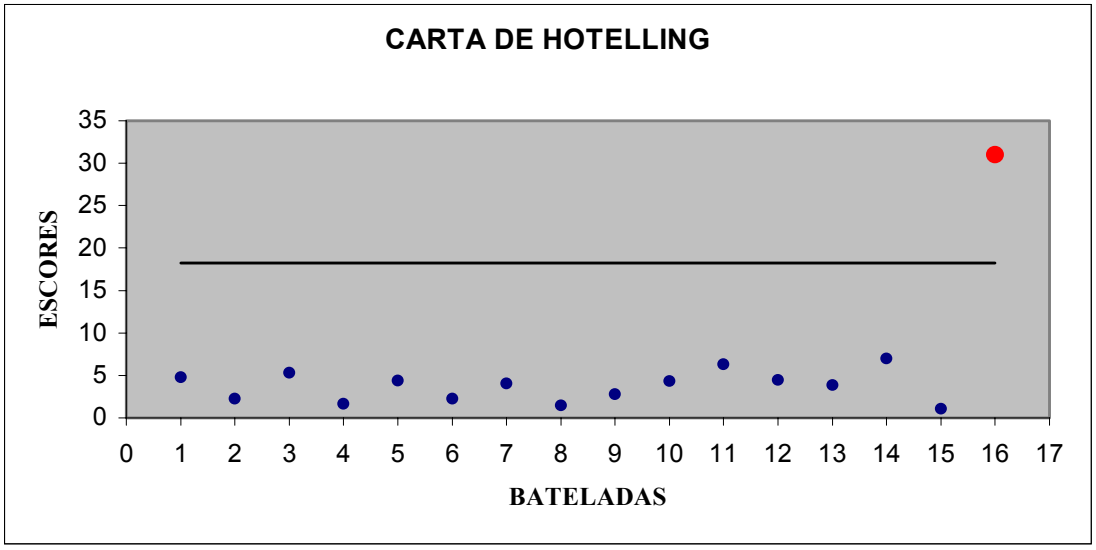

Figura 43. Carta de Hotelling construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); nova baletada observada no processo (batelada *16*).

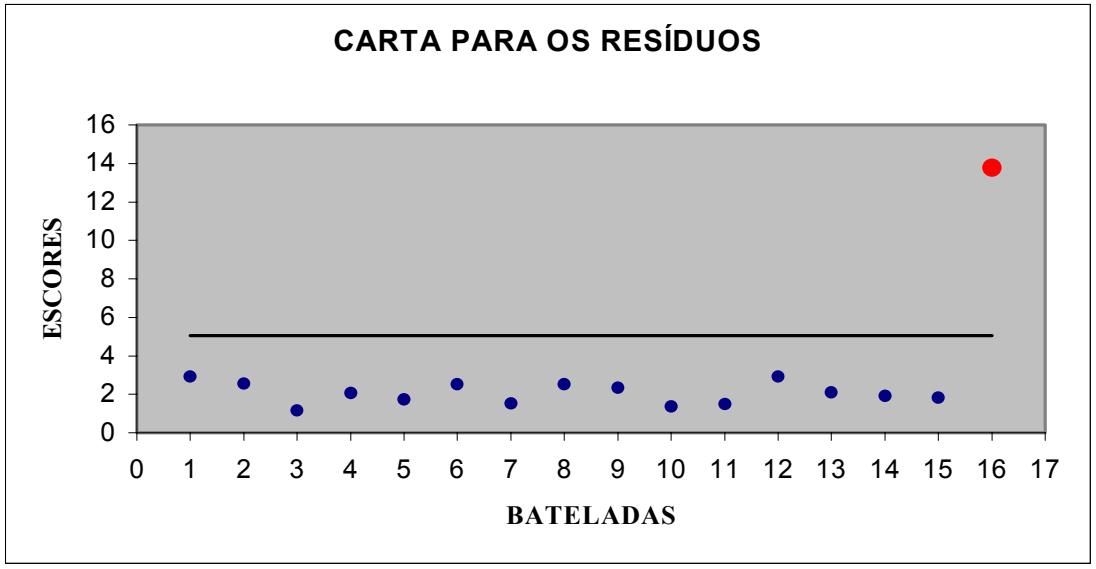

Figura 44. Carta Q construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); nova baletada observada no processo (batelada *16*).

Observando as cartas individuais para os CPs (Fig. 45, 46, 47 e 48), pode-se notar que os escores padronizados referentes ao primeiro e ao quarto CPs [$y_{1,16}$ e $y_{4,16}$ obtidos através da expressões (46) e (41)] estão além dos limites de controle da carta [expressão (35)]. Assim, deve-

se concentrar esforços nestes CPs na busca de um diagnóstico para o sinal de processo fora de controle, gerado a partir da carta de Hotelling.

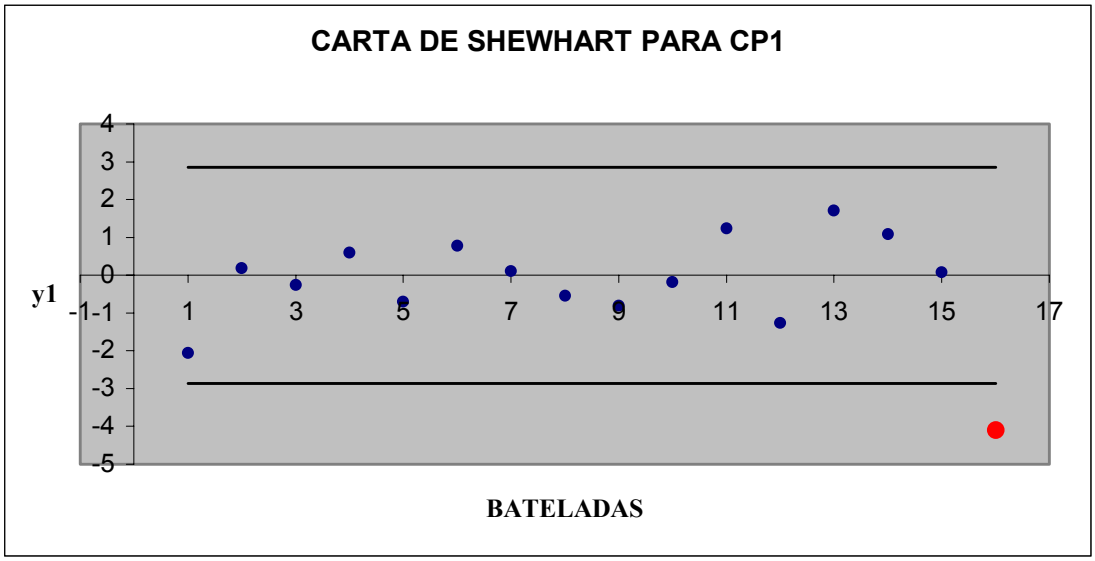

Figura 45. Carta de Shewhart para o primeiro Componente Principal; LIC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*).

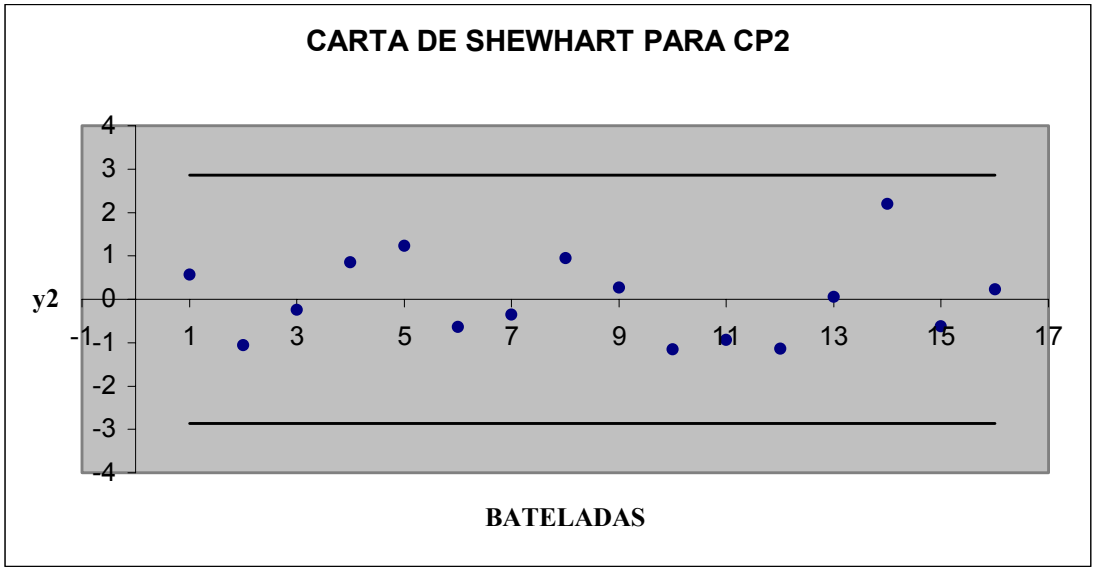

Figura 46. Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*).

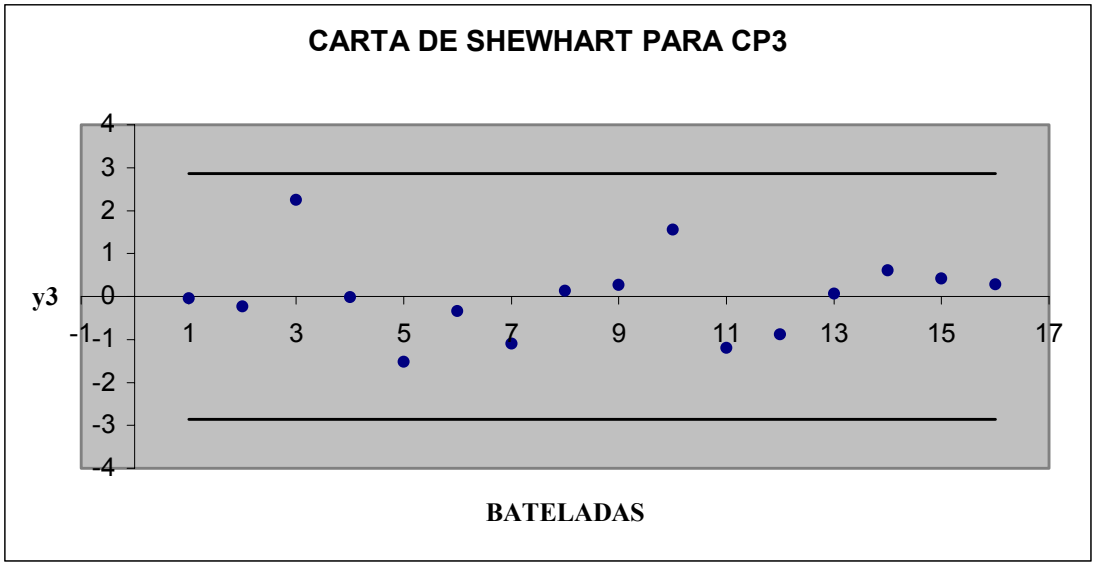

Figura 47. Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*).

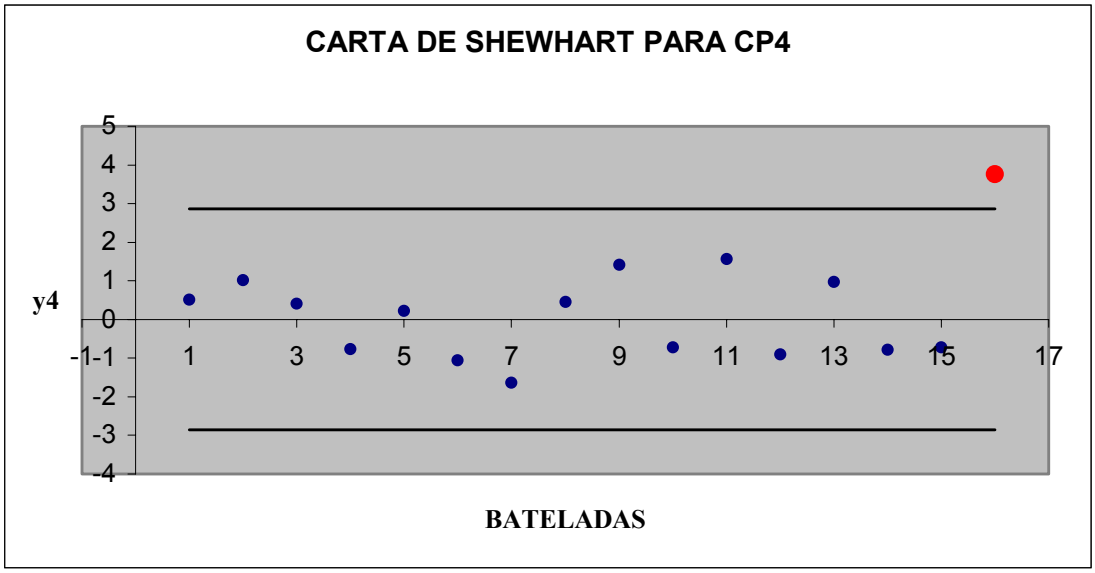

Figura 48. Carta de Shewhart para o quarto Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); nova batelada observada no processo (batelada *16*).

Deseja-se saber quais variáveis contribuíram mais efetivamente para a ocorrência dos altos valores de $y_{1,16}$, $y_{4,16}$ e de Q_{16}. O objetivo é buscar um diagnóstico para estes sinais, isto é, encontrar a causa que levou o processo a sair do seu estado de controle. A rápida detecção da

causa especial, seguida da devida ação corretiva, pode evitar que bateladas subseqüentes ainda apresentem variabilidade indesejada.

A Fig. 49 ilustra as contribuições conjuntas das 30 variáveis de processo para os valores significativos de $y_{1,16}$ e $y_{4,16}$ encontrados na batelada *16*. Nota-se que as variáveis 12, 14, 15, 16, 17 e 18 apresentaram as maiores contribuições. Na verdade, como estas variáveis representam replicações das duas variáveis de processo (X_1 e X_2) durante a batelada, deve-se entender que as variáveis 15 e 17 representam aferições da variável X_1 nos instantes 8 e 9, respectivamente, dentro da batelada. De maneira análoga, entende-se que as variáveis 12, 14, 16 e 18 representam aferições da variável X_2 nos instantes 6, 7, 8 e 9, respectivamente. Dessa forma, deve-se concentrar esforços nestas variáveis dentro destes períodos na busca de uma causa para os valores elevados encontrados nos escores correspondentes.

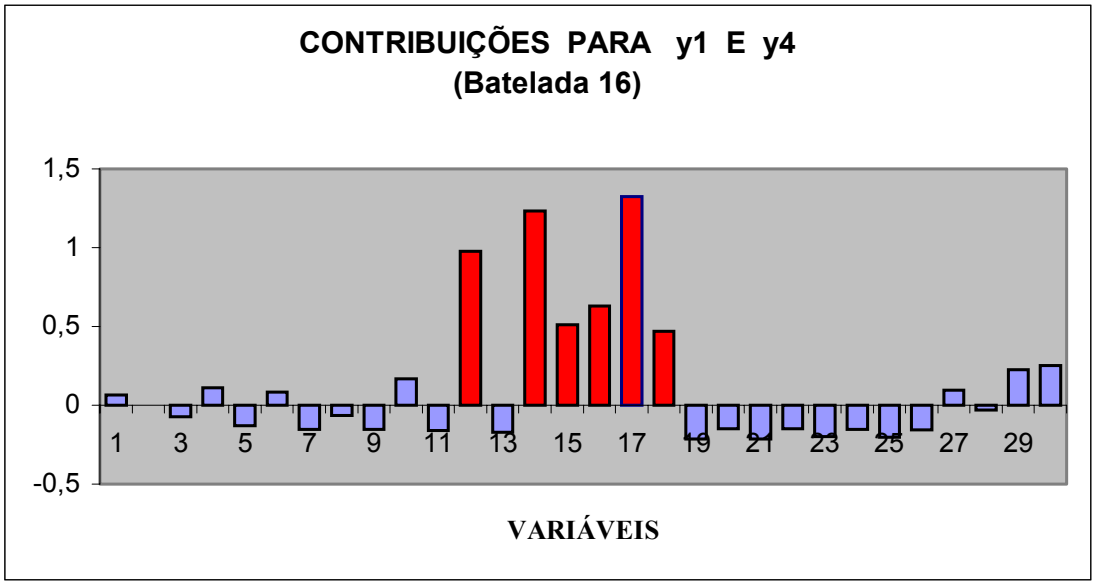

Figura 49. Contribuições conjuntas das 30 variáveis de processo para os escores y_1 (CP_1) e y_4 (CP_4); batelada *16*.

A Fig. 50 apresenta também as contribuições das 30 variáveis de processo para o valor significativo de Q_{16} encontrado na batelada *16*. Nota-se que as variáveis 1, 9, 11 e 18 apresentam as maiores contribuições. Analogamente, conclui-se que o modelo ACPM de referência construído em função dos quatro primeiros CPs gerou um elevado resíduo para a variável X_1 nos instantes 1, 5 e 6. Da mesma forma, entende-se que o modelo gerou um resíduo elevado para a

variável X_2 no instante 9. Deve-se, então, também concentrar esforços nestas variáveis, durante estes períodos, na busca de uma causa para o alto resíduo encontrado.

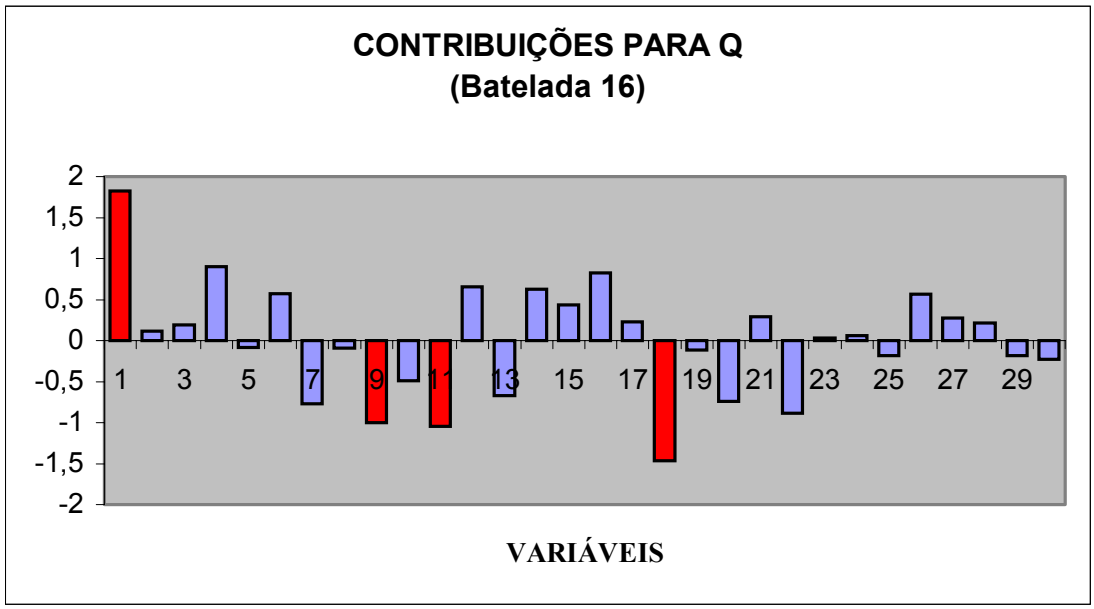

Figura 50. Contribuição das 30 variáveis de processo para o valor Q; batelada *16*.

5.3 - Monitoramento on-line de uma batelada futura

Na seção anterior, utilizou-se o modelo ACPM de referência para o monitoramento *off-line* da batelada *16*. Após finalizada a batelada, os valores de $y_{1,16}$, $y_{2,16}$, $y_{3,16}$, $y_{4,16}$ [expressões (46) e (41)], $T^2_{4,16}$ [expressão (47)] e Q_{16} [expressões (49) e (48), foram comparados com os seus respectivos limites de controle. Através dos Gráficos de Contribuição, buscou-se a obtenção de um diagnóstico para as possíveis causas para a variabilidade acentuada encontrada nas variáveis X_1 e X_2. Nesta seção, apresenta-se o monitoramento *on-line* da batelada *16*. Dessa forma, a partir do modelo ACPM, obtêm-se a cada novo instante l transcorrido na batelada, os valores parciais $y_{1l,16}$, $y_{2l,16}$, $y_{3l,16}$, $y_{4l,16}$ [expressões (46) e (53)], $T^2_{4l,16}$ [expressão (55)] e $Q_{l,16}$, para l=1,...,15,

[expressões (51) e (50)], e compara-se este valores em relação aos seus respectivos limites de controle.

É importante destacar que a batelada *16* será projetada no plano definido pelos quatro primeiros CPs, realizando-se o preenchimento gradativo da referente observação $\mathbf{x}'_{16}$ através do segundo critério apresentado na seção 4.3.1. Calcula-se, então, os escores $c_{1l,16}$, $c_{2l,16}$, $c_{3l,16}$, $c_{4l,16}$ [expressão (46)], completando-se o vetor $\mathbf{x}'_{16}$ com os resultados padronizados de X_1 e X_2 obtidos no último instante transcorrido na batelada.

A Fig. 51 apresenta os valores parciais da estatística de Hotelling obtidos a cada instante dentro da batelada *16* [expressão (55)] e o LSC da carta [expressão (40)]. Observa-se que, a partir do instante 4, a batelada começa a apresentar uma variabilidade significativa (isto é, $T^2_{4l,16}$ > LSC, para $l \geq 4$). Deve-se entender que a partir deste instante as trajetórias das variáveis X_1 e/ou X_2 começam a apresentar desvios significativos em relação às suas respectivas trajetórias médias.

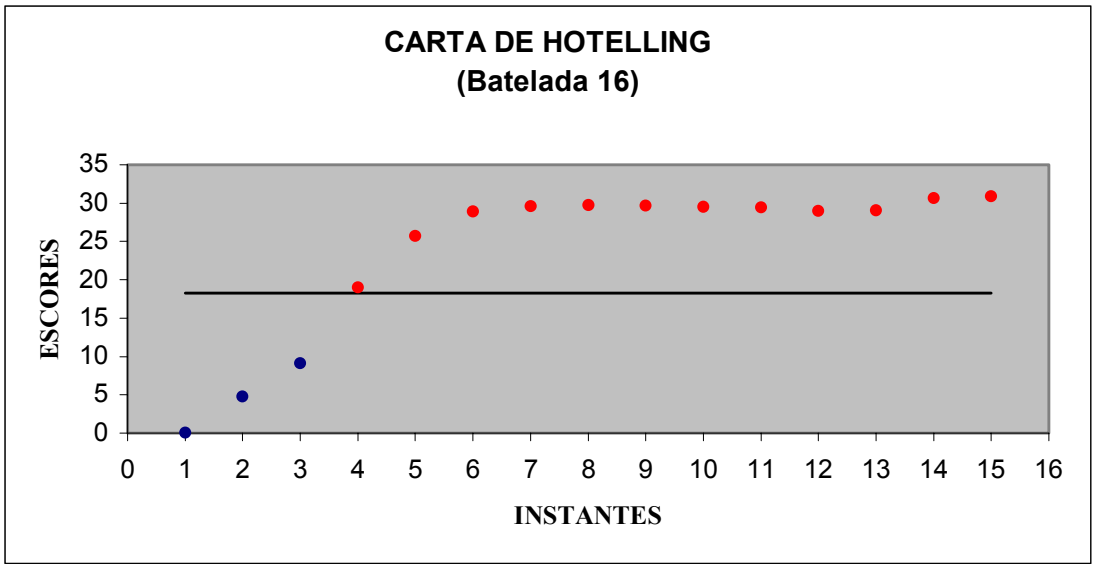

Figura 51. Carta de Hotelling construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*).

As Fig. 52, 53, 54 e 55 apresentam, respectivamente, as cartas de Shewhart para os quatro CPs retidos. Observa-se que a partir do instante 4 os escores padronizados referentes ao primeiro

e ao quarto CPs [$y_{14,16}$ e $y_{44,16}$ obtidos através da expressões (46) e (53)] apresentam valores inferiores ao LIC e superiores ao LSC, respectivamente [expressão (54)]. Isto indica que os valores parciais significativos encontrados na carta de Hotelling são reflexo dos escores elevados encontrados nestes CPs. Deve-se entender que, a partir do instante 4, a variabilidade demasiada verificada na batelada foi provocada pelas fontes de variabilidade do processo descritas por estes CPs. Concentram-se esforços nestes CPs na busca de um diagnóstico *on-line* para o sinal de processo fora de controle, gerado a partir deste instante pela carta de Hotelling.

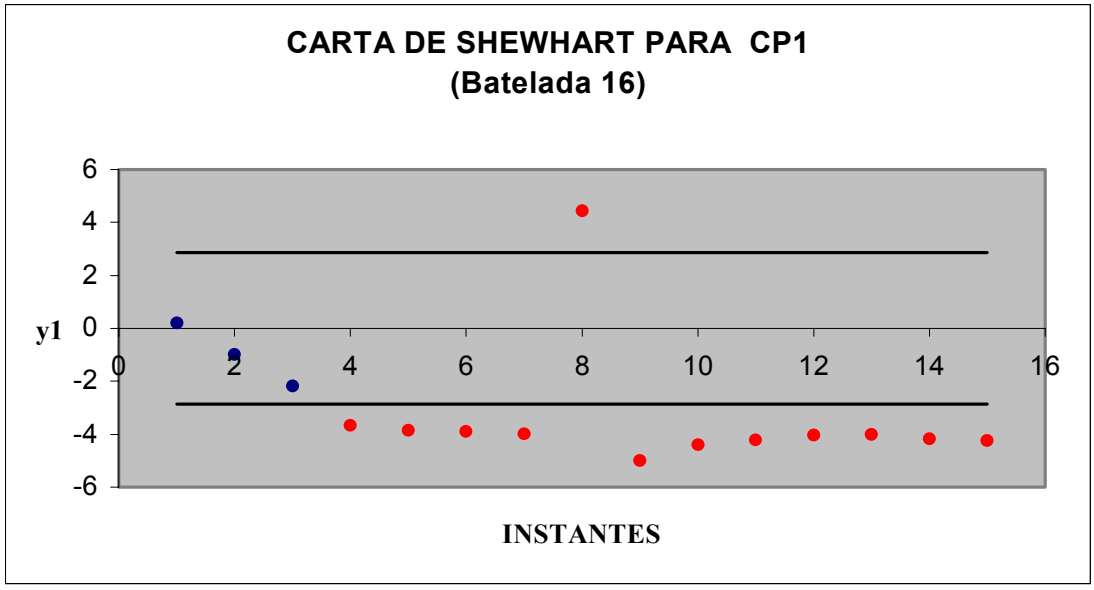

Figura 52. Carta de Shewhart para o primeiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*).

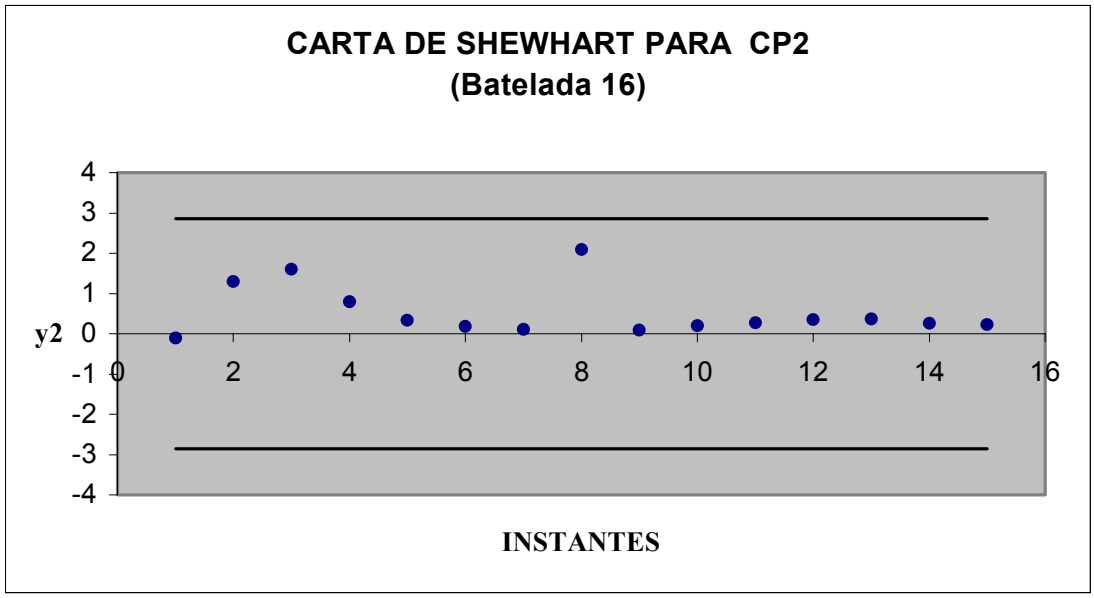

Figura 53. Carta de Shewhart para o segundo Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*).

Figura 54. Carta de Shewhart para o terceiro Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*).

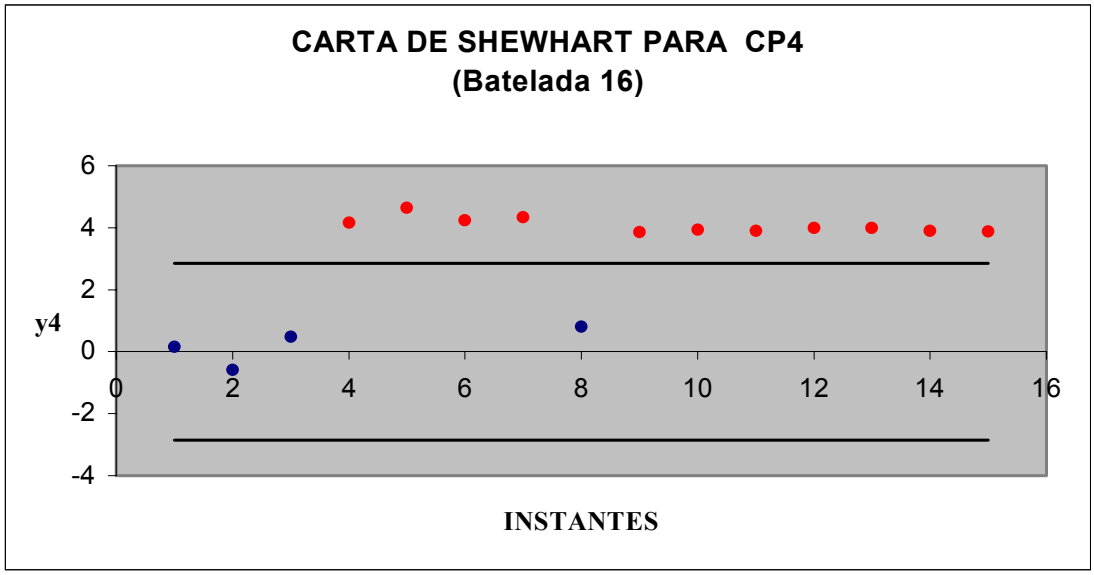

Figura 55. Carta de Shewhart para o quarto Componente Principal; LSC e LIC calculados com significância de 5%, utilizando o critério de Bonferroni (limites de Bonferroni); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*).

A Fig. 56 apresenta as contribuições conjuntas das variáveis X_1 e X_2 para os valores significativos de $y_{14,16}$ e $y_{44,16}$. Estes valores são encontrados utilizando a expressão descrita na seção 3.4.4, considerando apenas as cargas dos autovetores $\mathbf{u}_1$ e $\mathbf{u}_4$ referentes as variáveis X_1 e X_2 no instante 4 (isto é, $u_{1,7}$ e $u_{1,8}$ para $\mathbf{u}_1$, e $u_{4,7}$ e $u_{4,8}$ para $\mathbf{u}_4$). Nota-se que a variável X_1 apresenta maior contribuição em relação à variável X_2. Dessa forma, deve-se, a partir deste momento, concentrar esforços nesta variável na busca de uma causa para os valores elevados encontrados nos referentes escores parciais.

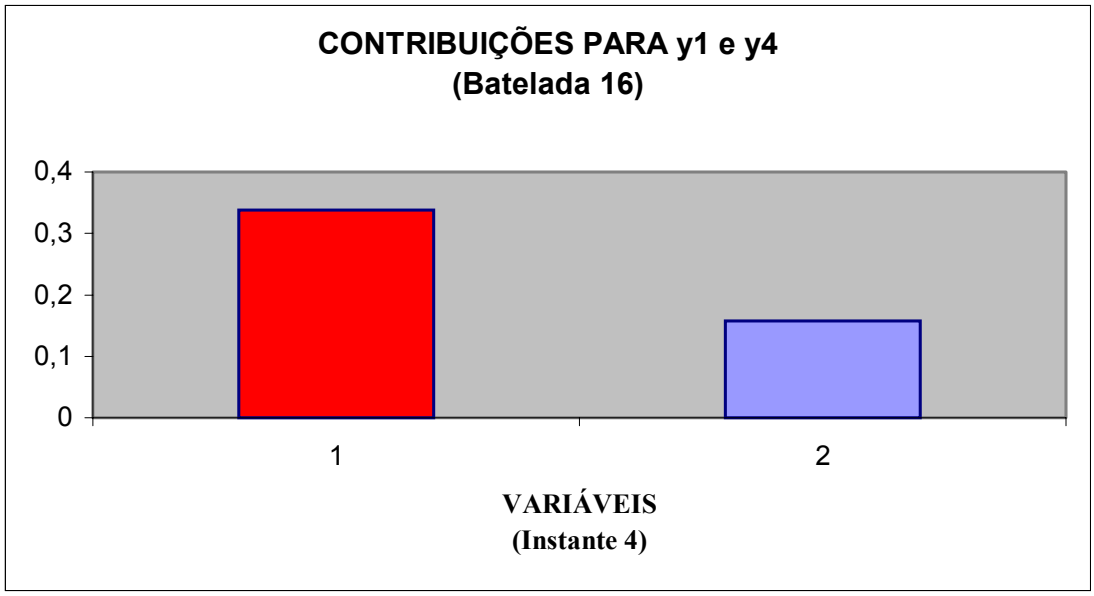

Figura 56. Contribuições conjuntas das variáveis X_1 e X_2, no instante 4 da batelada *16*,para os escores parciais $y_{12,16}$ e $y_{42,16}$.

A Fig. 57 apresenta os resíduos parciais $Q_{l,16}$ obtidos a cada instante dentro da batelada *16* [expressões (51) e (50)] e o LSC da carta [expressão (56)]. Observa-se que no instante 2 a batelada começa a apresentar um resíduo significativo (ou seja, $Q_{2,16} >$ LSC, para $2 \leq l \leq 11$). Conclui-se que, a partir deste instante, o modelo ACPM de referência construído em função dos quatro primeiros CPs não está descrevendo suficientemente o comportamento das variáveis X_1 e X_2, dentro da batelada *16*.

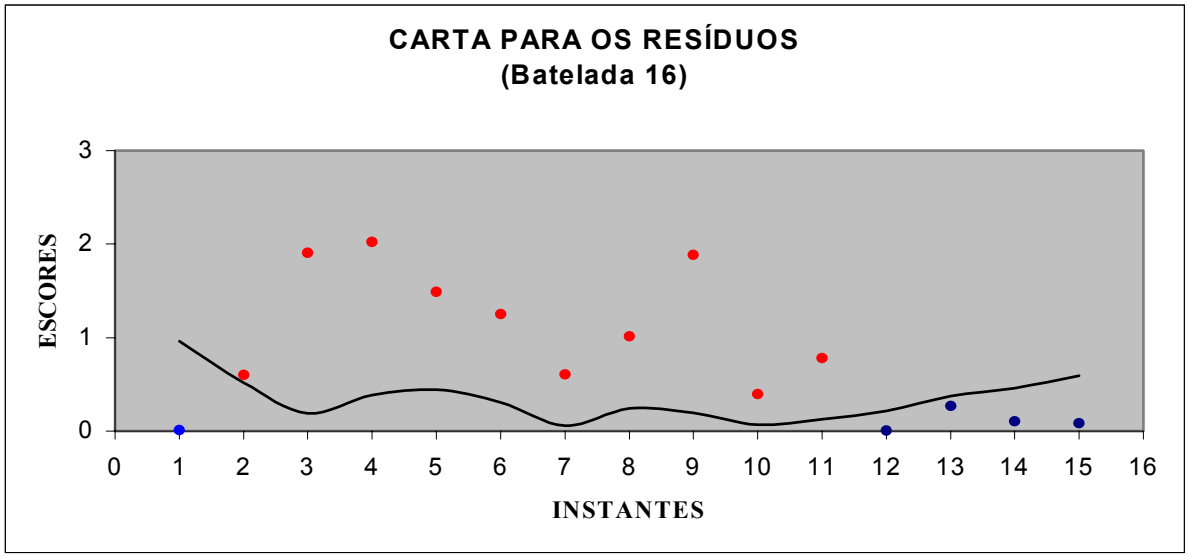

Figura 57. Carta Q construída a partir de 15 bateladas preliminares; LSC obtido com significância de 5% (α=5%); monitoramento *on-line* em 15 instantes da nova baletada observada no processo (batelada *16*).

A Fig. 58 apresenta as contribuições conjuntas das variáveis X_1 e X_2 para o valor significativo de $Q_{2,16}$ (seção 3.4.4). Estas contribuições são obtidas a partir do cálculo das diferenças $x_{3,16} - \hat{x}_{4,16}$ para a variável X_1, e $x_{3,16} - \hat{x}_{4,16}$ para a variável X_2. Observa-se que a variável X_2 apresenta maior contribuição em relação à variável X_1. Isto indica que, no instante 2 da batelada *16*, o modelo ACPM de referência não está descrevendo de maneira suficiente o comportamento da variável X_2, principalmente. Em conseqüência disto, deve-se, a partir deste momento, concentrar esforços nesta variável na busca de uma causa para o resíduo parcial elevado encontrado.

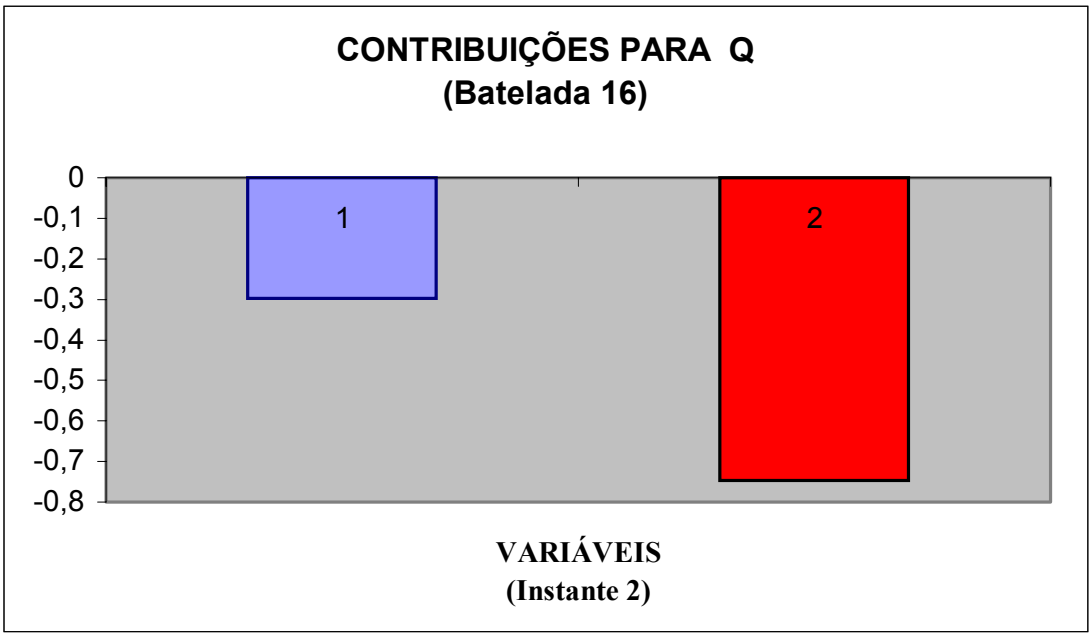

Figura 58. Contribuições conjuntas das variáveis X_1 e X_2, no instante 2 da batelada *16*, para o resíduo parcial $Q_{2,16}$.

5.4 – Considerações finais

É importante destacar, inicialmente, que as 22 bateladas disponíveis no processo apresentado neste capítulo foram consideradas conformes a partir da realização *a posteriori* de uma análise laboratorial. Em outras palavras, deve-se entender que estas bateladas geraram um produto final dentro das especificações de qualidade ***Y*** (ver Fig. 26). Em contrapartida, analisando os dados acerca das duas variáveis de processo envolvidas (dados da matriz **X**) através da abordagem ACPM, observou-se que 7 bateladas apresentaram variação demasiada nas trajetórias destas variáveis, sendo, assim, excluídas do modelo ACPM de referência. Isto ocorre pelo fato de que em processos químicos é difícil se obter, através das variáveis ***Y***, uma caracterização completa da qualidade do produto final. A análise ACPM considera aspectos da qualidade que não refletem em tais variáveis, fornecendo informações adicionais sobre a qualidade do produto final. Combinando os resultados obtidos da análise das variáveis ***Y*** e os

resultados obtidos na análise ACPM das variáveis $\boldsymbol{X}$, obtém-se, assim, uma caracterização mais precisa da qualidade do produto final. Dessa forma, bateladas que apresentam, através da análise ACPM, variação demasiada nas variáveis de processo e/ou produto final com qualidade fora das especificações, devem ser excluídas do modelo ACPM de referência para o monitoramento do desempenho de bateladas futuras.

Ressalta-se também que a abordagem ACPM não é um modelo de causa e efeito. Quando uma causa especial de variação é sinalizada através do monitoramento *on-line* de uma batelada futura, isto significa que, neste momento, a estrutura de correlação das variáveis de processo não está de acordo com a estrutura de correlação padrão. Apenas reajustando de forma deliberada as variáveis afetadas, fazendo com que o modelo gere nos instantes seguintes escores parciais $y_{il,f}$ e/ou os resíduos parciais $Q_{l,f}$ dentro dos limites de controle, não garante que o produto final obtido esteja dentro dos padrões de qualidade. Em outras palavras, dada a estrutura de correlação entre as variáveis, ajustes independentes em uma ou mais variáveis não irão necessariamente garantir a qualidade do produto final produzido nesta batelada e nas bateladas seguintes. Dessa forma, seguindo a filosofia do CEP tradicional, quando uma causa especial está presente, deve-se usar a ACPM em conjunto com os Gráficos de Contribuição para a identificação da causa especial, seguida de uma ação corretiva sobre o processo.

É importante observar que a ACPM é uma ferramenta eficiente na detecção da presença de eventos atípicos que reflitam alterações nas variáveis de processo. No entanto, em algumas situações, eventos importantes para a qualidade final do produto não impactam sobre as variáveis de processo aferidas. Nestes casos, deve-se obter uma caracterização mais apurada deste processo, adicionando para a análise mensurações acerca de novas variáveis de processo.

Uma questão igualmente importante para realização da ACPM reside no número de aferições das variáveis de processo realizadas durante a batelada. A freqüência de amostragem deve ser a mais adequada, permitindo a obtenção de informações relevantes sobre a trajetória das variáveis no processo. Quando existe conhecimento *a priori* de que, em determinados instantes da batelada, ocorrem transformações no processo que afetam significativamente a qualidade do produto final, deve-se então aumentar a freqüência de aferições das variáveis nestes instantes. Desta forma, obtém-se um modelo ACPM mais acurado para detectar desvios excessivos, nestes

instantes, nas trajetórias médias das variáveis, pois cada autovetor $\mathbf{u}_i$ associado terá um número maior de cargas $u_{i,j}$ correspondentes a estes períodos.

Sempre que o processo em estudo sofrer mudanças estruturais, deve-se reconstruir o seu modelo ACPM de referência, pois, neste caso, novos eventos comuns podem aparecer, provocando uma alteração na estrutura de correlação padrão das variáveis envolvidas. Através da obtenção de dados históricos acerca das variáveis de processo após esta mudança, gera-se um modelo que incorpora estes eventos, trazendo, assim, a nova estrutura de correlação das variáveis de processo.

Finalmente, ressalta-se que a ACPM constitui-se em um modelo empírico construído a partir de dados históricos acerca das trajetórias das variáveis de processo. Evidenciou-se a habilidade da ACPM em discriminar bateladas "conformes" e "não conformes", considerando apenas a estrutura de correlação das variáveis, obtida a partir de bateladas ocorridas no processo em controle estatístico.

CAPÍTULO 6

6- CONCLUSÃO

As indústrias estão buscando um controle cada vez mais eficaz de seus processos, disponibilizando inúmeros dados através de sistemas de controle informatizados. Particularmente, as indústrias químicas e de processos que utilizam processamento em bateladas (foco principal deste trabalho), possuem sensores, ligados *on-line* com a produção, que fornecem medições freqüentes das variáveis de processo, durante o transcurso de cada batelada.

Esta dissertação de mestrado teve como objetivo principal a apresentação de uma ferramenta estatística para controle multivariado de processos em bateladas. De forma detalhada, descreveu-se a base teórica para a elaboração de cartas de controle multivariadas baseadas na Análise de Componentes Principais Multidirecionais (ACPM).

A dissertação também trouxe uma revisão sobre as estratégias tradicionais de controle multivariado de processos. As principais cartas de controle univariadas (Cartas de Shewhart, CUSUM e EWMA), bem como suas respectivas extensões multivariadas (Cartas de Hotelling, MCUSUM, MEWMA) foram detalhadas. Também foi apresentada uma revisão sobre as

estratégias tradicionais de diagnóstico para as causas de distúrbios ocorridos no processo. Realizou-se uma discussão sobre a necessidade da utilização de ferramentas estatísticas mais robustas para modelar o comportamento de variáveis dentro de processos com grande quantidade de dados disponíveis. Finalmente, foi feito um apanhado das cartas de controle baseadas em métodos de projeção de dados, dentre as quais a ACPM faz parte.

O trabalho apresentou uma descrição detalhada da ferramenta estatística multivariada em que se baseia a estratégia ACPM, a Análise de Componentes Principais (ACP). A ACP é uma técnica estatística bastante utilizada em processos onde existem informações sobre um número muito grande de variáveis fortemente correlacionadas. Esta técnica nos permite reduzir essas variáveis para um número menor de variáveis independentes denominadas Componentes Principais (CPs). Os CPs trazem grande parte da informação contida nas variáveis iniciais. Através de dados simulados oriundos de um processo contínuo, foi apresentada uma aplicação da ACP na elaboração das cartas de controle simples de serem apresentadas e fáceis de serem interpretadas. Também foi apresentada uma estratégia eficaz de diagnóstico para as causas de distúrbios ocorridos no processo, os Gráficos de Contribuição. Viu-se que tais gráficos constituem-se numa ferramenta de auxilio na identificação das causas para os distúrbios no processo, apontados pelas cartas baseadas nos CPs.

Apresentou-se, também, a estrutura de um processo em bateladas. Viu-se que nestes processos, a média das variáveis de processo ocupam diferentes patamares, descrevendo trajetórias não lineares durante a batelada. Dessa forma, discutiu-se sobre a necessidade da utilização de uma estratégia de controle que monitore as trajetórias médias de tais variáveis. Forneceu-se o embasamento teórico para a elaboração das cartas de controle para monitoramento de processos desta natureza através da ACPM.

As cartas de controle baseadas em ACPM são ilustradas através de um estudo de caso, apresentando uma aplicação das cartas em um sub-processo em bateladas de produção de borrachas para cobertura de pneus. A abordagem ACPM mostrou ser eficiente para incorporar as informações relevantes trazidas deste processo, gerando uma distribuição de referência consistente das variáveis envolvidas. Através do modelo ACPM de referência, fez-se o monitoramento *on-line* do comportamento destas variáveis em uma batelada futura simulada. As cartas de controle construídas forneceram sinalizações *on-line* sobre o estado do processo, em

vários instantes da batelada, permitindo, assim, a correção dos distúrbios apresentados durante a batelada em tempo real.

6.1 – Sugestões para pesquisas futuras

Neste trabalho, apresentou-se o referencial teórico para a elaboração de cartas de controle, via ACPM, para monitoramento de processos em bateladas. Utilizando esta abordagem, realizou-se, através de uma situação simulada, o monitoramento *on-line* do processo em bateladas apresentado no capítulo 5. Entretanto, para viabilizar a efetiva execução desse controle, necessita-se de *software* que incorpore a teoria proposta. São necessárias ferramentas computacionais que forneçam aos operadores, em tempo real, informações sobre o processo através da geração de cartas e gráficos de controle. A criação de *software* que viabilize a implantação dessas cartas de controle se constitui num tema extremamente interessante para um trabalho futuro.

Ressalta-se que a ACPM é uma estratégia que se baseia na projeção dos dados oriundos das variáveis de processo, via ACP, num plano de dimensões reduzidas. É importante destacar que esta abordagem considera somente a estrutura de correlação das variáveis de processo $\boldsymbol{X}$ na elaboração das cartas de controle e, conseqüentemente, monitora o processo através de amostras futuras das variáveis de processo. A abordagem PELM (Projeção de Estruturas Latentes Multidirecionais) é uma técnica alternativa à ACPM para monitoramento de processos em bateladas, que permite a utilização simultânea das informações contidas nas variáveis de qualidade das matérias-primas $\boldsymbol{Z}$, nas variáveis de processo $\boldsymbol{X}$, e nas variáveis de qualidade do produto final $\boldsymbol{Y}$. O detalhamento desta técnica, bem como sua aplicação em uma indústria brasileira que conduza seu processo em bateladas, também se constitui num tema propício para um trabalho futuro.

Finalmente, destaca-se que as estratégias MPCA e MPEL constituem-se em modelos lineares para monitoramento de processos. No entanto existem processos industriais complexos, cujos dados apresentam fortes características não lineares que não podem ser monitorados via

métodos lineares. Para tratar deste problema, técnicas não lineares vêm sendo desenvolvidas, incluindo PCA não linear, PLS não linear, análise de correlação não linear e procedimentos envolvendo redes neurais. A realização de um trabalho que ilustre a utilização de um modelo não linear para monitoramento de tais processos parece pertinente.

ANEXO 1

Tabela 4. Glossário.

CC	Carta e Controle
CQ	Característica de Qualidade
CEP	Controle Estatístico do Processo
CEMP	Controle Estatístico Multivariado do Processo
CUSUM	Soma Acumulada
MCUSUM	Soma Acumulada Multivariada
MCU	Múltiplas Cartas Univariadas
CCUSUM	Soma Acumulada Multivariada de Crosier
MC1	Soma Acumulada Multivariada de Pignatiello e Runger
EWMA	Média Móvel Exponencialmente Ponderada
MEWMA	Média Móvel Exponencialmente Ponderada Multivariada
ARL	Tamanho Médio de Corrida
ACP	Análise de Componentes Principais
CP	Componente Principal
ACPM	Análise de Componentes Principais Multidirecionais
PEL	Projeção de Estruturas Latentes
PEL-MB	Projeção de Estruturas Latentes via Múltiplos Blocos
PELM	Projeção de Estruturas Latentes Multidirecionais

ANEXO 2

Tabela 5. Dados simulados, relativos a 30 amostras de 8 variáveis de processo, utilizados para o exemplo apresentado no Capítulo 3.

AMOSTRAS	VARIÁVEIS							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	0	0,7	-0,09333	0,35	-0,3217	-0,2657	1,007	1,232
2	0,4	-0,2	0,306667	-0,05	0,0783	0,0343	0,307	0,232
3	-0,3	0	-0,29333	0,05	-0,0217	0,0043	-0,193	-0,768
4	-0,3	0,1	-0,39333	-0,15	0,1783	0,1343	-0,093	0,132
5	1,7	1,5	1,806667	1,35	-1,3217	-1,2657	1,807	2,032
6	1	0,8	2,906667	1,75	-1,7217	-1,5657	0,007	0,232
7	-1,3	-1,2	-2,49333	-2,25	2,2783	2,2343	-0,993	-0,768
8	-0,5	-0,7	-0,19333	-0,45	0,4283	0,2343	-0,243	0,032
9	0,1	-0,6	0,006667	-0,25	0,2283	0,0343	0,107	0,432
10	-0,4	-0,4	-0,49333	-0,55	0,5283	0,4343	-0,093	0,132
11	0,5	0,4	0,406667	-0,05	0,0283	0,0343	0,247	0,732
12	-0,8	-1	-0,79333	-0,45	0,3283	0,1343	-0,893	-0,668
13	-0,8	-1	-0,79	-0,45	0,32	0,1343	-0,8	-0,6
14	0,1	-0,2	0,106667	-0,25	0,2283	0,0343	0,007	0,232
15	-1,5	-0,8	-1,59333	-1,65	1,0283	2,0343	-1,193	-0,768
16	0	0,7	-0,09333	0,75	-0,7217	-0,6657	0,007	0,232
17	0,4	-0,2	-0,09333	0,75	-0,6717	-0,5657	0,107	0,332
18	-0,3	0	-0,19333	-1,05	1,0283	0,5343	-0,193	0,032
19	-0,3	0,1	-0,29333	-0,45	0,4283	0,3343	-0,593	-0,368
20	1,7	1,5	1,406667	0,75	-0,5717	-0,6657	2,007	2,232
21	1,7	1,5	1,4	0,8	-0,57	-0,6	2	2
22	-1,3	-1,2	-1,49333	-1,25	1,0283	1,0343	-0,993	-1,768
23	-0,5	-0,7	-0,49333	-0,35	0,0283	0,0343	-0,793	-0,668
24	0,1	-0,6	0,006667	-0,25	0,2783	0,2843	0,007	-2,768
25	-0,4	-0,4	-0,49333	-0,55	0,5783	0,4343	-0,093	-2,768
26	0,5	0,4	0,606667	-0,25	0,2783	0,1843	0,507	0,732
27	-0,8	-1	-0,99333	-1,25	-0,9717	-0,9657	-0,993	-0,868
28	1,3	1,6	0,906667	1,75	-0,8717	-0,7657	1,007	1,532
29	0,1	-0,2	-0,09333	1,5	-1,5	-1,6	0,007	0,472
30	-1,5	-0,8	-1,09333	-2,25	2,2783	2,2343	-0,993	-0,268

ANEXO 3

Tabela 6. Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas.

l	Var	BATELADAS										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	X_1	0,02	0,13	0,1	0,06	0,03	0,12	0,19	0,11	0,1	0,04	0,13
	X_2	86	77	92	93	87	86	74	75	82	96	88
2	X_1	0,35	0,89	0,56	0,6	0,41	0,32	0,76	0,6	0,65	0,61	0,58
	X_2	77	73	83	80	76	82	73	73	78	80	82
3	X_1	1,56	2,68	2,42	2,26	1,7	0,48	2,64	2,3	2,37	2,07	2,06
	X_2	76	75	79	79	77	78	75	73	78	79	79
4	X_1	3,68	4,77	4,54	4,4	3,92	0,61	4,7	4,4	4,52	4,1	4,27
	X_2	78	80	83	85	81	75	81	79	84	84	82
5	X_1	5,8	6,92	6,64	6,47	6	0,74	6,82	6,5	6,5	6,25	6,33
	X_2	86	89	92	91	85	74	90	87	90	89	88
6	X_1	7,85	8,97	8,73	8,57	8,04	1,66	8,82	8,56	8,57	8,3	8,41
	X_2	96	98	100	99	96	75	101	100	102	96	97
7	X_1	9,94	11,07	10,85	10,7	10,15	2,31	11	10,74	4,52	10,39	10,45
	X_2	107	109	110	109	107	72	111	110	84	107	106
8	X_1	12,03	12,87	12,68	12,8	12,29	3,81	12,98	12,85	6,5	12,44	12,5
	X_2	116	118	120	121	117	74	121	121	90	118	117
9	X_1	12,86	13,12	12,96	13,39	13,35	5,88	13,45	13,28	8,57	13,32	13,03
	X_2	119	120	121	125	123	83	122	122	102	124	122
10	X_1	12,96	13,64	13,34	13,68	13,53	7,87	13,78	13,64	10,73	13,46	13,22
	X_2	116	118	120	124	122	94	120	119	111	120	118
11	X_1	14,41	15,04	14,88	15,26	15,09	10,05	15,35	15,35	12,69	14,99	14,75
	X_2	122	127	126	128	129	104	128	125	121	127	124
12	X_1	15,67	16,86	16,4	16,76	16,43	12,23	16,9	16,82	13,02	16,11	16,21
	X_2	127	132	130	133	132	114	134	130	123	131	129
13	X_1	17,95	19,09	18,64	19,05	18,57	14,21	19,2	19,01	13,14	18,41	18,58
	X_2	137	143	141	146	143	126	145	142	119	140	139
14	X_1	19,92	20,54	20,02	20,95	20,57	14,61	20,71	20,79	14,6	20,33	20,38
	X_2	147	150	148	155	152	127	154	151	125	150	148
15	X_1	20,71	21,11	20,56	21,44	21,2	14,81	21,16	21,35	16,02	21,11	20,99
	X_2	149	151	150	157	155	124	155	155	126	154	152

Tabela 7. Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas. (Continuação da tabela 6).

l	Var	BATELADAS										
		12	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
1	X_1	0,12	0,18	0,2	0,17	0,07	0,25	0,02	0,09	0,11	0,05	0
	X_2	80	57	79	49	75	92	115	73	81	76	109
2	X_1	0,63	0,7	0,85	0,6	0,65	0,76	0,58	0,57	0,53	0,26	0
	X_2	73	62	75	50	70	81	94	70	74	72	100
3	X_1	2,66	2,65	2,78	2,16	2,28	2,69	2,09	2,57	2,51	0,4	0,01
	X_2	75	65	76	60	71	81	84	72	76	68	90
4	X_1	4,77	4,8	4,93	4,08	4,43	4,87	4,2	4,73	4,69	0,53	0,01
	X_2	83	76	80	67	76	84	87	80	81	65	84
5	X_1	6,78	6,84	7	6,19	6,51	6,83	6,27	6,81	6,73	1,9	0,01
	X_2	92	84	89	77	86	91	92	87	91	68	80
6	X_1	8,85	8,81	9,04	8,28	8,51	8,94	8,33	8,84	8,8	3,53	0,03
	X_2	100	95	99	91	97	101	99	98	100	70	75
7	X_1	10,94	10,9	11,16	10,36	10,6	11,08	10,5	11	10,82	5,6	0,04
	X_2	112	106	109	103	108	112	111	110	110	72	71
8	X_1	12,74	12,71	13,01	12,32	12,71	12,91	12,72	13,06	12,84	7,63	0,07
	X_2	120	116	121	114	116	122	122	119	119	74	69
9	X_1	13	12,91	13,18	12,51	13,19	13,3	13,67	3,49	13,23	9,74	0,11
	X_2	118	113	119	109	118	123	128	121	121	79	68
10	X_1	13,46	13,78	14,37	12,76	13,29	14,13	13,93	13,84	13,66	11,83	0,15
	X_2	117	113	121	107	114	124	127	119	117	85	67
11	X_1	14,94	15,41	15,83	14,25	14,72	15,92	15,7	15,6	15,11	13,94	0,21
	X_2	123	120	128	119	117	133	132	129	124	108	66
12	X_1	16,52	17,37	17,84	15,81	16,16	17,55	16,97	17,14	16,73	14,8	0,6
	X_2	128	130	137	124	124	140	139	133	132	121	66
13	X_1	18,82	19,57	20,02	18,01	18,35	19,76	19,01	19,35	18,96	14,93	1,27
	X_2	139	140	147	137	135	150	146	145	143	118	68
14	X_1	20,25	21,06	21,05	19,69	20,36	21,08	20,99	21,08	20,58	16,38	2,3
	X_2	147	148	153	145	145	155	155	154	150	130	72
15	X_1	20,88	21,52	21,44	20,18	21,06	21,57	21,67	21,67	21,04	17,65	4,34
	X_2	149	149	154	147	148	156	159	155	151	134	79

ANEXO 1

Tabela 4. Glossário.

CC	Carta e Controle
CQ	Característica de Qualidade
CEP	Controle Estatístico do Processo
CEMP	Controle Estatístico Multivariado do Processo
CUSUM	Soma Acumulada
MCUSUM	Soma Acumulada Multivariada
MCU	Múltiplas Cartas Univariadas
CCUSUM	Soma Acumulada Multivariada de Crosier
MC1	Soma Acumulada Multivariada de Pignatiello e Runger
EWMA	Média Móvel Exponencialmente Ponderada
MEWMA	Média Móvel Exponencialmente Ponderada Multivariada
ARL	Tamanho Médio de Corrida
ACP	Análise de Componentes Principais
CP	Componente Principal
ACPM	Análise de Componentes Principais Multidirecionais
PEL	Projeção de Estruturas Latentes
PEL-MB	Projeção de Estruturas Latentes via Múltiplos Blocos
PELM	Projeção de Estruturas Latentes Multidirecionais

ANEXO 2

Tabela 5. Dados simulados, relativos a 30 amostras de 8 variáveis de processo, utilizados para o exemplo apresentado no Capítulo 3.

AMOSTRAS	VARIÁVEIS							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	0	0,7	-0,09333	0,35	-0,3217	-0,2657	1,007	1,232
2	0,4	-0,2	0,306667	-0,05	0,0783	0,0343	0,307	0,232
3	-0,3	0	-0,29333	0,05	-0,0217	0,0043	-0,193	-0,768
4	-0,3	0,1	-0,39333	-0,15	0,1783	0,1343	-0,093	0,132
5	1,7	1,5	1,806667	1,35	-1,3217	-1,2657	1,807	2,032
6	1	0,8	2,906667	1,75	-1,7217	-1,5657	0,007	0,232
7	-1,3	-1,2	-2,49333	-2,25	2,2783	2,2343	-0,993	-0,768
8	-0,5	-0,7	-0,19333	-0,45	0,4283	0,2343	-0,243	0,032
9	0,1	-0,6	0,006667	-0,25	0,2283	0,0343	0,107	0,432
10	-0,4	-0,4	-0,49333	-0,55	0,5283	0,4343	-0,093	0,132
11	0,5	0,4	0,406667	-0,05	0,0283	0,0343	0,247	0,732
12	-0,8	-1	-0,79333	-0,45	0,3283	0,1343	-0,893	-0,668
13	-0,8	-1	-0,79	-0,45	0,32	0,1343	-0,8	-0,6
14	0,1	-0,2	0,106667	-0,25	0,2283	0,0343	0,007	0,232
15	-1,5	-0,8	-1,59333	-1,65	1,0283	2,0343	-1,193	-0,768
16	0	0,7	-0,09333	0,75	-0,7217	-0,6657	0,007	0,232
17	0,4	-0,2	-0,09333	0,75	-0,6717	-0,5657	0,107	0,332
18	-0,3	0	-0,19333	-1,05	1,0283	0,5343	-0,193	0,032
19	-0,3	0,1	-0,29333	-0,45	0,4283	0,3343	-0,593	-0,368
20	1,7	1,5	1,406667	0,75	-0,5717	-0,6657	2,007	2,232
21	1,7	1,5	1,4	0,8	-0,57	-0,6	2	2
22	-1,3	-1,2	-1,49333	-1,25	1,0283	1,0343	-0,993	-1,768
23	-0,5	-0,7	-0,49333	-0,35	0,0283	0,0343	-0,793	-0,668
24	0,1	-0,6	0,006667	-0,25	0,2783	0,2843	0,007	-2,768
25	-0,4	-0,4	-0,49333	-0,55	0,5783	0,4343	-0,093	-2,768
26	0,5	0,4	0,606667	-0,25	0,2783	0,1843	0,507	0,732
27	-0,8	-1	-0,99333	-1,25	-0,9717	-0,9657	-0,993	-0,868
28	1,3	1,6	0,906667	1,75	-0,8717	-0,7657	1,007	1,532
29	0,1	-0,2	-0,09333	1,5	-1,5	-1,6	0,007	0,472
30	-1,5	-0,8	-1,09333	-2,25	2,2783	2,2343	-0,993	-0,268

ANEXO 3

Tabela 6. Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas.

l	Var	BATELADAS										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	X_1	0,02	0,13	0,1	0,06	0,03	0,12	0,19	0,11	0,1	0,04	0,13
	X_2	86	77	92	93	87	86	74	75	82	96	88
2	X_1	0,35	0,89	0,56	0,6	0,41	0,32	0,76	0,6	0,65	0,61	0,58
	X_2	77	73	83	80	76	82	73	73	78	80	82
3	X_1	1,56	2,68	2,42	2,26	1,7	0,48	2,64	2,3	2,37	2,07	2,06
	X_2	76	75	79	79	77	78	75	73	78	79	79
4	X_1	3,68	4,77	4,54	4,4	3,92	0,61	4,7	4,4	4,52	4,1	4,27
	X_2	78	80	83	85	81	75	81	79	84	84	82
5	X_1	5,8	6,92	6,64	6,47	6	0,74	6,82	6,5	6,5	6,25	6,33
	X_2	86	89	92	91	85	74	90	87	90	89	88
6	X_1	7,85	8,97	8,73	8,57	8,04	1,66	8,82	8,56	8,57	8,3	8,41
	X_2	96	98	100	99	96	75	101	100	102	96	97
7	X_1	9,94	11,07	10,85	10,7	10,15	2,31	11	10,74	4,52	10,39	10,45
	X_2	107	109	110	109	107	72	111	110	84	107	106
8	X_1	12,03	12,87	12,68	12,8	12,29	3,81	12,98	12,85	6,5	12,44	12,5
	X_2	116	118	120	121	117	74	121	121	90	118	117
9	X_1	12,86	13,12	12,96	13,39	13,35	5,88	13,45	13,28	8,57	13,32	13,03
	X_2	119	120	121	125	123	83	122	122	102	124	122
10	X_1	12,96	13,64	13,34	13,68	13,53	7,87	13,78	13,64	10,73	13,46	13,22
	X_2	116	118	120	124	122	94	120	119	111	120	118
11	X_1	14,41	15,04	14,88	15,26	15,09	10,05	15,35	15,35	12,69	14,99	14,75
	X_2	122	127	126	128	129	104	128	125	121	127	124
12	X_1	15,67	16,86	16,4	16,76	16,43	12,23	16,9	16,82	13,02	16,11	16,21
	X_2	127	132	130	133	132	114	134	130	123	131	129
13	X_1	17,95	19,09	18,64	19,05	18,57	14,21	19,2	19,01	13,14	18,41	18,58
	X_2	137	143	141	146	143	126	145	142	119	140	139
14	X_1	19,92	20,54	20,02	20,95	20,57	14,61	20,71	20,79	14,6	20,33	20,38
	X_2	147	150	148	155	152	127	154	151	125	150	148
15	X_1	20,71	21,11	20,56	21,44	21,2	14,81	21,16	21,35	16,02	21,11	20,99
	X_2	149	151	150	157	155	124	155	155	126	154	152

Tabela 7. Medições das variáveis *energia integrada* (X_1) e *temperatura da massa* (X_2) obtidas em 15 instantes, durante 11 bateladas. (Continuação da tabela 6).

l	Var	BATELADAS										
		12	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
1	X_1	0,12	0,18	0,2	0,17	0,07	0,25	0,02	0,09	0,11	0,05	0
	X_2	80	57	79	49	75	92	115	73	81	76	109
2	X_1	0,63	0,7	0,85	0,6	0,65	0,76	0,58	0,57	0,53	0,26	0
	X_2	73	62	75	50	70	81	94	70	74	72	100
3	X_1	2,66	2,65	2,78	2,16	2,28	2,69	2,09	2,57	2,51	0,4	0,01
	X_2	75	65	76	60	71	81	84	72	76	68	90
4	X_1	4,77	4,8	4,93	4,08	4,43	4,87	4,2	4,73	4,69	0,53	0,01
	X_2	83	76	80	67	76	84	87	80	81	65	84
5	X_1	6,78	6,84	7	6,19	6,51	6,83	6,27	6,81	6,73	1,9	0,01
	X_2	92	84	89	77	86	91	92	87	91	68	80
6	X_1	8,85	8,81	9,04	8,28	8,51	8,94	8,33	8,84	8,8	3,53	0,03
	X_2	100	95	99	91	97	101	99	98	100	70	75
7	X_1	10,94	10,9	11,16	10,36	10,6	11,08	10,5	11	10,82	5,6	0,04
	X_2	112	106	109	103	108	112	111	110	110	72	71
8	X_1	12,74	12,71	13,01	12,32	12,71	12,91	12,72	13,06	12,84	7,63	0,07
	X_2	120	116	121	114	116	122	122	119	119	74	69
9	X_1	13	12,91	13,18	12,51	13,19	13,3	13,67	3,49	13,23	9,74	0,11
	X_2	118	113	119	109	118	123	128	121	121	79	68
10	X_1	13,46	13,78	14,37	12,76	13,29	14,13	13,93	13,84	13,66	11,83	0,15
	X_2	117	113	121	107	114	124	127	119	117	85	67
11	X_1	14,94	15,41	15,83	14,25	14,72	15,92	15,7	15,6	15,11	13,94	0,21
	X_2	123	120	128	119	117	133	132	129	124	108	66
12	X_1	16,52	17,37	17,84	15,81	16,16	17,55	16,97	17,14	16,73	14,8	0,6
	X_2	128	130	137	124	124	140	139	133	132	121	66
13	X_1	18,82	19,57	20,02	18,01	18,35	19,76	19,01	19,35	18,96	14,93	1,27
	X_2	139	140	147	137	135	150	146	145	143	118	68
14	X_1	20,25	21,06	21,05	19,69	20,36	21,08	20,99	21,08	20,58	16,38	2,3
	X_2	147	148	153	145	145	155	155	154	150	130	72
15	X_1	20,88	21,52	21,44	20,18	21,06	21,57	21,67	21,67	21,04	17,65	4,34
	X_2	149	149	154	147	148	156	159	155	151	134	79

REFERÊNCIAS BIBLIOGRÁFICAS

AKEY, T. M., GREEN, S. B. & SALKIND, N. J. 2ª Ed. **Using SPSS for Windows: analysing and understanding data.** Prentice Hall, Upper Saddle River, New Jersey, 2000.

ALT, F. B. **Multivariate Quality Control, in Encyclopedia of Statistical Sciences.** John Wiley & Sons, Inc., New York, 1985.

BROOK, D. & EVANS, D. A. An approach to the probability distribuition of cusum run length. *Biometrika*, v.59, n° 3, p. 539-544, 1972.

CROWDER, V. C. Design of Exponentially Weighted Moving Average Schemes. *Journal of Quality Technology*, v. 21, n° 3, p. 155-162, 1989.

CROSIER, R. B. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, v. 30, n° 3, p. 291-303, 1988.

DOGANAKSOY, N., FALTIN, F. W. & TUCKER, W. T. Identification of out of control quality characteristics in a multivariate manufacturing environment. *Communications in Statistics – Theory and Methods*, v. 20, p. 1775-2790, 1991.

DUNCAN, A. J. **Quality Control and Industrial Statistics.** 5ª Ed. Irwin, Chicago, 1986.

GELADI, P. & KOWALSKI, B. R. An example of 2-Block Predictive Partial Last-Squares regression with simulated data. *Analytica Chimica Acta*, v. 185, p. 19-32, 1986b.

GELADI, P. & KOWALSKI, B. R. Partial Least Squares regression: a Tutorial. *Analytica Chimica Acta*, v. 185, p.1-17, 1986a.

GNANADESIKAN, R. **Methods for Statistical Data Analysis of Multivariate Observation.** 2ª Ed. John Wiley & Sons, Inc., New York, 1997.

HAWKINS, D. M. Multivariate Quality Control Based no Regression-Adusted Variables. *Technometrics*, v.33, n°1, p. 61-75, 1991.

HAWKINS, D. M. Regression Adjustment for Variables in Multivariate Quality Control. *Journal of Quality Technology*, v. 25, n° 3, p. 170-182, 1993.

HEALY, J. D. A Note on Multivariate CUSUM Procedures. *Technometrics*, v. 29, n° 4, p. 409-412, 1987.

HOLMES, D. S. & MERGEN, A. E. Improving the Performance of the T^2 Control Chart. *Quality Engineering*, v.5, p.619-625, 1993.

HOSKULDSSON, A. PLS Regression Methods. *Journal of chemometrics*, v. 2, p. 211-228, 1988.

JACKSON, J. E. A **User's Guide to Principal Components.** John Wiley & Sons, Inc., New York, 1991.

JACKSON, J. E. Principal Component and factor Analysis: Part I – Principal Components. *Journal of Quality Technology*, v. 12, n° 4, p. 201-213, 1980.

JACKSON, J. E. & MUDHOLKAR, G. S. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, v. 21, n° 3, p. 341-349, 1979.

JOHNSON, R.A. & WICHERN, D. W. **Applied Multivariate Statistical Analysis.** 5ª Ed. Prentice Hall, Englewood Cliffs, New Jersey, 1992.

KOURTI, T. & MACGREGOR, J. F. Multivariate SPC Methods for Process and Product Monitoring. *Journal of Quality Technology*, v 28, n° 4, p. 409-428, 1996

KOURTI, T., NOMIKOS, P. & MACGREGOR, J. F. Analysis, monitoring and faut diagnosis of batch processes using multiblock and multiway PLS. *Journal of process Control*, v. 5, n° 4, 1995.

KRESTA, J., MACGREGOR, J.F. & MARLIN, T.E. Multivariate Statistical Monitoring of Process Operating Performance. *Canadian Journal of Chemical Engineering*, v. 69, p. 35-47, 1991.

LORBER, A. WANGEN, L. E. & KOWALSKI, B. R. A theoretical Fundation for the PLS Algorithm. *Journal of Chemometrics*, v. 1, p. 19-31, 1987.

LOWRY, C. A. & MONTGOMERY, D. C. A Revew of Multivariate Control Charts. *IIE Transactions*, v. 27, p. 800-810, 1995.

LOWRY C. A. & WOODALL, W. H. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, v. 34, n° 1, p. 46-53, 1992.

LUCAS, J. M. The Design and Use of Cumulative Sum Quality Control Schemes. *Journal of Quality Technology*, v. 8, n° 1, p. 1-12, 1976.

LUCAS, J. M. & SACCUCCI, M. S. Exponentially Weighted Moving Average Control Schemes: properties and enhancements. *Technometrics*, v. 32, n° 1, p. 1-12, 1990.

MACGREGOR, J.F. Using On-line Process Data to Improve Quality. *39ª Annual Fall Conference*, p. 277-284, 1985.

MACGREGOR, J.F., JAECKLE, C., KIPARISSIDES, C. & KOUTOUDI, M. Process Monitoring and Diagnosis by Multi-Block PLS Methods. *Journal of the American Institute of Chemical Engineers*, v. 40, n° 5, p. 826-838, 1994.

MACGREGOR, J. F. & KOURTI, T. Statistical Process Control of Multivariate Processes. *Control Engeneering Practice*, v. 3, n° 3, p. 403-414, 1995.

MARSH, C. E. & TUCKER, T. W. Application of SPC Techniques to Batch Units. *ISA Transactions*, v. 30, n° 1, p. 39-47, 1991.

MARTIN, E. B. & MORRIS, A. J. Na overview of multivariate statistical process control in continous and batch process performance monitoring. *Trans Inst MC*, v. 18, n° 1, p. 51-60, 1996.

MARTIN, E. B., MORRIS, A. J. & ZHANG, J. Process performance monitoring using multivariate statistical process control. *IIE Control Theory Applied*, v.143, n° 2, p. 132-143, 1996.

MASON, R. L., CHAMP, C. W., TRACY, N. D., WIERDA, S. J. & YOUNG, J. C. Assessment of Multivariate Process Control Techniques. *Journal of Quality Technology*, v. 29, n° 2, p. 140-143, 1997.

MASON, R. L., TRACY, N. D. & YOUNG, J. C. Decomposition of T^2 *for Multivariate Control Chart Interpretation*, v. 27, n° 2, p. 99-108, 1995.

MILLER, P., SWANSON, R. E. & HECKLER, C. E. Contribution Plots: The Missing Link in Multivariate Quality Control. *Unpublished manuscript to be submitted to Journal of Quality Technology*, 1993.

MONTGOMERY, D. C. **Introduction to Statistical Quality Control.** 3ª Ed. John Wiley & Sons, Inc., New York, 1996.

MONTGOMERY, D. C. The Use of Statistical Process Control and Design of Experiments in Product and Process Improvement. *IIE Transactions*, v. 24, n° 5, p. 4-17, 1992.

MONTGOMERY, D. C. & MASTRANGELO, C. M. Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology*, v. 23, n° 3, p. 179-193, 1991.

MOOD, A. M., GRAYBILL, F. A. & BOES, D. C. **Introduction to the Theory of Statistics.** MacGraw-Hill Publishing Company, New York, 1974.

MURPHY, B. J. Selecting out of control variables with the T^2 multivariate quality control procedure. *The Statistician*, v. 36, p. 571-583, 1987.

NELSON, L. S. Control Charts for Individual Measurements. *Journal of Quality Technology*, v. 14, n° 2, p. 172-173, 1982.

NOMIKOS, P. & MACGREGOR, J. F. Monitoring Batch Processes Using Multiway Principal Component Analysis. *Journal of the American Institute of Chemical Engineers*, v. 40, n° 8, p. 1361-1375, 1994.

NOMIKOS, P. & MACGREGOR, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, v. 37, n° 1, p. 41-59, 1995.

PALM, A. C., RODRIGUEZ, R. N., SPIRING, F. A. & WHEELER, D. J. Some Perspectives and Challenges for Control Chart Methods. *Journal of Quality Technology*, v.29, n° 2, p. 122-127, 1997.

PIGNATIELLO, J. J. Jr. & RUNGER, G. C. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, v. 22, n° 3, p. 173-186, 1990.

RENCHER, A. C. **Methods of Multivariate Analysis.** John Wiley & Sons, Inc., New York, 1995.

RENCHER, A. C. The Contribuition of Individual Variables to Hotelling's T^2, Wilks' Λ, and R^2. *Biometrics*, v. 49, p. 479-489, 1993.

RYAN, T. P. **Statistical Methods for Quality Improvement.** 1ª Ed. John Wiley & Sons, Inc., New York, 1989.

SULLIVAN, J. H. & WOODALL, W. H. Comparison of Multivariate Control Charts for Individual observations. *Journal of Quality Technology*, v.28, n° 4, p. 398-408, 1996.

TRACY, N. D., YOUNG, J. C. & MASON, R. L. Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, v. 24, n° 2, p. 88-95, 1992.

WADE, M. R. & WOODALL, W. H. A Revew and Analysis of Cause-Selecting Control Charts. *Journal of Quality Technology*, v.25, n° 3, p. 161-169, 1993.

WOLD, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics*, v. 20, n° 4, p. 397-405, 1978.

WOODALL, W. H. & NCUBE, M. M. Multivariate CUSUM Quality Control Procedures. *Technometrics*, v. 27, n° 3, p. 285-292, 1985.

YASHCHIN, E. Some aspects of the theory of statistical control schemes. *IBM Journal of Research and Development*, v. 31, p. 199-205, 1987.

YIN, R. K. **Case study Research – design and methods.** 2ª Ed. Stage Publications, London, 1994.