

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**MÉTODO PARA CONTROLE ESTATÍSTICO
MULTIVARIADO DE PROCESSOS EM
BATELADA**

Ariane Ferreira Porto Rosa

Porto Alegre, 2001

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**MÉTODO PARA CONTROLE ESTATÍSTICO
MULTIVARIADO DE PROCESSOS EM BATELADA**

Ariane Ferreira Porto Rosa

Orientador: Professor Flávio Sanson Fogliatto, Ph.D.

Banca Examinadora:

**Luiz Paulo Luna de Oliveira, Dr.
Prof. Centro de Ciências Exatas e Tecnológicas / UNISINOS**

**Dinara Westphalen Xavier Fernandez, Dra.
Prof. Depto. Estatística / UFRGS**

**Carla Schwengber ten Caten, Dra.
Prof. PPGEPI / UFRGS**

**Dissertação submetida 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, Abril de 2001.

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção.

Prof. Flávio Sanson Fogliatto, Ph.D.

Universidade Federal do Rio Grande do Sul
Orientador

Prof. Luiz Afonso dos Santos Senna, Ph. D.

Coordenador PPGEP/UFRGS

Banca Examinadora:

Luiz Paulo Luna de Oliveira, Dr.

Prof. Centro de Ciências Exatas e Tecnológicas/ UNISINOS

Dinara Westphalen Xavier Fernandez, Dra.

Profa. Depto. Estatística / UFRGS

Carla Shwengber ten Caten, Dra.

Profa. PPGEP / UFRGS

“Se prestares atenção à sabedoria com o teu ouvido, para inclinares teu coração ao discernimento, se persistires a procurar isso como a prata e continuares a buscar isso como a tesouros escondidos, neste caso acharás o próprio conhecimento. ”

Livro de Provérbios 2: 2-5 (Provérbios de Salomão)

AGRADECIMENTOS

Após a conclusão desta jornada tenho o mais profundo sentimento de gratidão a expressar àquelas pessoas que participaram e contribuíram para o seu cumprimento e êxito.

Agradeço, em especial, ao Professor Flávio Sanson Fogliatto, orientador deste trabalho, que tornou claros os caminhos a serem percorridos e que, com sua dedicação e entusiasmo, serviu de exemplo e incentivo.

Agradeço aos professores do Programa de Pós-Graduação em Engenharia de Produção, por sua contribuição à minha formação profissional.

Agradeço aos colegas, com os quais compartilhei a experiência de aprender, e cuja amizade tornou mais agradável a jornada. Em especial, agradeço aos colegas Fernando Rezende Pellegrini e Rogério Royer, pela atenção e incentivo dispensados a mim.

Agradeço, também aos funcionários do Programa de Pós-Graduação em Engenharia de Produção.

Finalmente, dedico este trabalho aos meus pais, raízes profundas que me concederam a vida e que iluminam meus caminhos.

RESUMO

As cartas de controle estatístico têm sido amplamente utilizadas no monitoramento do desempenho de processos. Com a crescente informatização dos processos industriais, tem-se verificado um aumento sensível na quantidade de informações disponíveis sobre variáveis de processo. Via de regra, essas variáveis apresentam-se fortemente correlacionadas. Em casos especiais, como nos processos em batelada, tais variáveis descrevem um perfil de variação ao longo do tempo, caracterizando o comportamento normal do processo. Nessas condições especiais, as cartas de controle tradicionais não proporcionam um monitoramento eficaz sobre o processo.

Esta dissertação de mestrado apresenta uma alternativa para o monitoramento *on line* de processos em bateladas: a proposição de uma metodologia para implantação de cartas de controle multivariadas baseadas em componentes principais. A idéia central dessas cartas é monitorar simultaneamente diversas variáveis, controlando somente algumas poucas combinações lineares independentes delas; tais combinações são denominadas componentes principais.

O presente trabalho ilustra a metodologia proposta em um estudo de caso realizado na etapa de fermentação do processo de fabricação de cerveja de uma indústria de bebidas, localizada na região metropolitana de Porto Alegre.

ABSTRACT

Statistical control charts have been widely used for monitoring the performance of industrial processes. The current computer integration of manufacturing processes has led to an increasing amount of data available on process performance. Usually, such data are outcomes of strongly correlated variables. In special processes, such as batch processes, variables are expected to follow a profile under normal process operating conditions. Traditional control charts are not prepared to efficiently monitor such processes.

This dissertation presents an alternative to the on-line monitoring of batch processes: a multivariate control chart based on principal components. The key idea is to provide efficient control over several variables by monitoring only a few independent linear combinations of them, called principal components. The principal components control charts are particularly efficient for on-line monitoring of batch processes, as discussed in the dissertation.

The present work illustrates the method proposed in a case carried out in a brewing company located in the metropolitan area of Porto Alegre.

LISTA DE TABELAS

TABELA 2.1: Dados do exemplo de CCM	29
TABELA 4.1: Variabilidade total explicada pelos oito CP retidos no modelo ACPM elaborado.....	86

ÍNDICE

LISTA DE FIGURAS	IX
LISTA DE TABELAS	XI
RESUMO	XII
ABSTRACT	XIII
1. INTRODUÇÃO	1
1.1. Considerações iniciais	1
1.2. O Tema e sua Importância	3
1.3. Objetivos	4
1.3.1. Objetivo Principal	4
1.3.2. Objetivos Específicos	4
1.4. Método de Trabalho	5
1.5. Estrutura da Dissertação	5
1.6. Limitações	6
2. REVISÃO BIBLIOGRÁFICA	8
2.1. Métodos de Projeção Multivariada de Dados	8
2.1.1. Análise de Componentes Principais (ACP)	10
2.1.1.1. Seleção do Número de CP	13
2.1.1.2. Interpretação Geométrica dos CP	19
2.1.1.3. Análise dos Resíduos dos CP	20
2.1.2. Mínimos Quadrados Parciais (MQP)	21
2.2. Fundamentos das Cartas de Controle Estatístico de Processos	22
2.2.1. Principais Cartas de Controle para Variáveis	23
2.3. Cartas de Controle Multivariadas	25
2.3.1. Cartas de Hotelling ou cartas T^2	26
2.3.2. Carta de controle multivariada para soma acumulada (MCUSUM)	33
2.3.3. Carta de controle multivariada para média móvel exponencialmente ponderada (MEWMA)	39
2.3.4. Carta de Controle para Monitoramento da Variabilidade de Processos ...	42
2.3.5. Controle Estatístico de Processos para Dados de Processo Auto Correlacionados	44
2.4. Carta de Controle Multivariada do tipo Hotelling T^2 baseada em Métodos de Projeção Multivariada de Dados	46
2.4.1. Carta de Controle Multivariada Baseada em Componentes Principais para Processos em Batelada	50
2.4.2. Identificação das variáveis fora de controle em cartas multivariadas	61

2.5. Interface entre Controle de Processos na Engenharia e Controle Estatístico de Processo.....	65
3. MÉTODO PROPOSTO.....	67
3.1. Identificação das Variáveis de Interesse.....	69
3.2. Elaboração de uma Distribuição de Referência para os Dados de Processo. 70	
3.2.1. Arranjo dos Dados das Bateladas.....	71
3.2.2. Normalização dos Dados da Matriz X.....	71
3.2.3. Aplicação da ACPM nos Dados Normalizados da Matriz X.....	72
3.2.4. Seleção de Bateladas no Conjunto de Dados Históricos.....	73
3.3. Elaboração da CCM de Hotelling Baseada em Componentes Principais.....	73
3.4. Procedimento de Monitoramento <i>on line</i> de Novas Bateladas.....	74
3.4.1. Preenchimento dos Dados que Faltam no vetor x_{Novo} entre o Tempo Corrente k e o Término da Batelada.....	74
3.4.2. Cálculos dos Escores e da Estatística de Hotelling para cada Instante K	75
3.4.3. Verificar se uma Nova Batelada Apresenta Desvios do Modelo ACPM Elaborado para o.....	77
3.5. Procedimento para Diagnóstico das Variáveis que Opcionaram Causa Especial.....	78
3.5.1. Identificação de Sinal Fora de Controle na CCM de Hotelling e/ou na Carta <i>QEP</i>	78
3.5.2. Análise das CC de Shewart dos Escores Normalizados dos CP.....	78
3.5.3. Cálculo da Contribuição das Variáveis.....	79
4. ESTUDO DE CASO.....	80
4.1. Identificação das Variáveis de Interesse.....	81
4.2. Elaboração de uma Distribuição de Referência para os dados de Processo . 83	
4.2.1. Arranjo dos Dados das Bateladas.....	84
4.2.2. Normalização dos Dados da Matriz X.....	84
4.2.3. Aplicação da ACPM nos Dados Normalizados da Matriz X	84
4.2.4. Seleção de Bateladas no Conjunto de Dados Históricos.....	87
4.3. Elaboração da CCM de Hotelling Baseada em Componentes Principais.....	88
4.4. Procedimento de Monitoramento <i>on line</i> de Novas Bateladas.....	89
4.4.1. Preenchimento dos Dados que Faltam no vetor x_{16}^t entre o Tempo Corrente k e o Término da Batelada.....	90
4.4.2. Cálculos dos Escores e da Estatística de Hotelling para cada Instante k .	90
4.4.3. Verificar se uma Nova Batelada apresenta desvios do Modelo ACPM Elaborado para o Processo.....	94
4.5. Procedimento para Diagnóstico das Variáveis que Opcionaram Causa Especial.....	96
4.5.1. Identificação de Sinal Fora de Controle na CCM de Hotelling e/ou na Carta <i>QEP</i>	96
4.5.2. Análise das CC de Shewart dos Escores Normalizados dos CP.....	96
4.5.3 Cálculo da Contribuição das Variáveis.....	97
4.6. Considerações Finais.....	99
5. CONCLUSÃO.....	101
5.1. Sugestões para Trabalhos Futuros.....	103

REFERÊNCIAS BIBLIOGRÁFICAS	105
ANEXO 1	113
ANEXO 2	117

CAPÍTULO 1

1. INTRODUÇÃO

1.1. Considerações iniciais

As cartas de controle estatístico de processos (CC) são importantes ferramentas utilizadas no gerenciamento da qualidade total. Estas cartas foram introduzidas por Shewhart em 1924, para o monitoramento de uma variável de qualidade ao longo de processos produtivos. Mais tarde a Western Electric Company (WECO) introduziu um conjunto de regras para identificar situações fora de controle em processos utilizando estas cartas. Desde então, uma variedade de cartas de controle tem sido desenvolvidas e usadas como ferramentas auxiliares no controle estatístico de processos, no controle de qualidade e no gerenciamento da qualidade total.

A estratégia de controle univariado de Shewhart, acrescida de algumas suposições, pode ser estendida para o caso multivariado. CC multivariadas são indicadas em situações onde variáveis de processo ou produto são monitoradas simultaneamente. Os resultados das CC multivariadas e univariadas são particularmente diferentes quando do monitoramento simultâneo de variáveis correlacionadas; nestes casos, o uso de CC multivariadas é fortemente recomendado. O tipo de CC multivariada de utilização mais comum é a carta de Hotelling, ou carta de T^2 (Hotelling, 1947). Outros tipos de CC multivariadas são apresentadas em Jackson (1991).

Algumas situações especiais dificultam ou impossibilitam a utilização das CC tradicionais. Por exemplo, as CC multivariadas (assim como as univariadas) pressupõem independência entre pontos amostrais. Com a crescente automatização de processos,

variáveis podem ser medidas em tempo real e em pontos temporais próximos; nessas condições, a suposição de independência dificilmente é verificada. Uma outra situação especial ocorre quando a variação normal de uma variável monitorada ao longo do tempo é descrita por um perfil; ou seja, a média da variável, em condições normais de processo, muda ao longo do tempo. As CC univariadas não podem ser aplicadas nessa situação, já que pressupõem variáveis com média constante.

Nesta dissertação, apresenta-se uma proposta metodológica para o monitoramento, simultâneo e em tempo real, de variáveis determinantes da qualidade de produtos e processos: as cartas de controle multivariadas baseadas em componentes principais (CCP). Essas cartas são operacionalizadas utilizando a análise de componentes principais e as CC multivariadas de Hotelling (CCM). As CCP admitem pontos amostrais sucessivos dependentes, além de permitirem o monitoramento de variáveis que não apresentam valores fixos de média ao longo do tempo. As CCP são particularmente eficientes no monitoramento *on line* de processos em batelada, onde um grande número de variáveis de processo esteja envolvido.

A análise de componentes principais tem por objetivo reduzir a dimensionalidade de conjuntos de dados multivariados. Em outras palavras, busca-se representar de maneira reduzida um conjunto de variáveis aleatórias, sem que isso acarrete perda significativa de informações. Tal representação reduzida é obtida utilizando componentes principais. Matematicamente, os componentes principais são combinações lineares das variáveis aleatórias, com a propriedade especial de independência.

A idéia-chave das CCP consiste em utilizar uma estratégia de controle similar àquela das CCM utilizando, ao invés das variáveis de processo, componentes principais delas derivados. Desta forma, simplifica-se a tarefa de monitoramento de processos, já que este passa a ser feito sobre um número menor de variáveis (ou seja, sobre os componentes principais). Como os componentes principais são independentes entre si, as CCP podem ser utilizadas no monitoramento de processos com as características especiais descritas anteriormente.

Neste trabalho apresenta-se o desenvolvimento de uma metodologia para a aplicação de cartas de controle multivariadas no controle *on-line* de processos em batelada, baseada em uma abordagem proposta por Nomikos e Macgregor (1995). Cartas de controle multivariadas para processos em batelada constituem-se em desenvolvimento teórico recente, iniciado há pouco mais de uma década e, conseqüentemente, ainda pouco explorado.

Muitos produtos industriais são provenientes de processos de produção realizados em batelada, como exemplo, pode-se citar a produção de resinas poliméricas, produtos bioquímicos e farmacêuticos. Um processo de produção em batelada constitui-se das seguintes etapas: as matérias-primas são colocadas em um recipiente, sendo processadas durante um determinado tempo, geralmente fixo, até que o produto final seja obtido. A trajetória de uma série de variáveis de processo, tais como temperaturas e pressões, podem ser medidas durante a etapa de processamento. Finalmente, o produto é analisado quanto às suas principais características de qualidade para verificar se está dentro do padrão de qualidade desejado, após a conclusão de cada batelada.

1.2. O Tema e sua Importância

As cartas de controle multivariadas baseadas em componentes principais foram escolhidas como tema desta dissertação de mestrado por apresentarem-se como uma alternativa para o monitoramento *on line* de processos. A idéia central dessas cartas é monitorar simultaneamente diversas variáveis de processo controlando somente algumas poucas combinações lineares independentes delas; tais combinações são denominadas componentes principais. O projeto enfatiza a aplicação das cartas de componentes principais no monitoramento de processos em batelada.

O tema proposto neste projeto de dissertação tem sua relevância fundamentada em alguns fatores , apresentados a seguir.

A utilização de processos em bateladas na produção e no processamento de matérias-primas e produtos acabados no meio industrial tem apresentado incremento significativo nas últimas décadas. Isto se deve, em grande parte, à crescente demanda por produtos customizados (isto é, com características definidas pelos clientes) (Davis, 1989). A customização de produtos leva à produção em pequenos lotes, característica da maioria dos processos em batelada. Neste sentido, a investigação de práticas de controle de qualidade para tais processos tem sua importância plenamente justificada.

Finalmente, é importante observar que a aplicação de cartas de controle estatístico de processos nos processos em bateladas tem sido muito limitada. A maioria dos métodos de controle estatístico de processo utilizam apenas medidas de qualidade do produto, obtidas no final de cada batelada e, por isso, monitoram apenas as variações entre bateladas (MacGregor, 1995). O monitoramento da variação da qualidade dentro das bateladas aumenta as chances de obtenção de produtos acabados dentro das especificações. Desta forma, reduzem-se as

perdas por refugo e retrabalho, bem como os custos associados às análises laboratoriais de produtos acabados.

1.3. Objetivos

Nesta Dissertação propõe-se o estudo das cartas multivariadas de controle de processo baseadas em componentes principais (CCP). O estudo compreende um levantamento dos aspectos teóricos referentes a estas cartas, bem como a proposição de uma metodologia para o monitoramento de processos em batelada. O projeto traz uma proposta de estudo de caso com o objetivo principal de aplicar a CCP em um processo industrial em batelada.

1.3.1 Objetivo Principal

O objetivo principal desta dissertação é apresentar o desenvolvimento de uma metodologia para implantação das Cartas de Controle Multivariadas Baseadas em Componentes Principais, para monitoramento de processos em batelada. Os passos metodológicos incluem etapas relacionadas à coleta e organização dos dados, bem como sua análise estatística.

Para tanto, é necessário um desenvolvimento teórico detalhado sobre a ferramenta de projeção multivariada de dados escolhida: a Análise de Componentes Principais (ACP) e também o desenvolvimento teórico das CC multivariadas de Hotelling. O monitoramento *on line* de produtos e processos em batelada é realizado através das CC de Hotelling, utilizando como dados de entrada nas CC as novas variáveis definidas na ACP.

1.3.2 Objetivos Específicos

Nesta dissertação, apresentam-se como objetivos específicos:

- Revisar a bibliografia existente sobre Cartas de Controle Multivariadas e sobre os Métodos de Projeção Multivariada de dados.
- Desenvolver uma metodologia para o monitoramento, simultâneo e em tempo real, de variáveis determinantes da qualidade de produtos e de processos utilizando como método de projeção multivariada de dados a Análise de Componentes Principais e as CC multivariadas de Hotelling, constituindo as cartas de controle multivariadas baseadas em componentes principais (CCP);

- Elaborar e realizar um estudo de caso constituindo-se na aplicação da metodologia elaborada em um processo industrial em batelada, utilizando dados obtidos na etapa de fermentação no processo de produção de cerveja.

1.4. Método de Trabalho

O método de pesquisa científica, adotado nesta dissertação, enquadra-se na categoria de pesquisa aplicada aliada a um estudo de caso.

A pesquisa aplicada consiste na aplicação da pesquisa básica a problemas do mundo real. Este método refere-se à discussão de um problema e à proposição de soluções para o mesmo, utilizando-se um referencial teórico. A pesquisa aplicada considera uma preocupação teórica, no sentido de alcançar o refinamento ou mesmo o desenvolvimento de uma nova teoria ou método para intervenção e solução de problemas genéricos. O tema da pesquisa aplicada deve ser tão generalizável no tempo e no espaço quanto possível, mas limitado ao contexto de sua aplicação (Pantton *apud* Roesch, 1994).

O estudo de caso é uma metodologia de pesquisa indicada quando: (i) questionamentos “como?” e “por que?” estão sendo colocados, (ii) os investigadores possuem pouco controle sobre o evento estudado e (iii) o foco é sobre um fenômeno contemporâneo dentro de um contexto da vida real. Entre as muitas situações em que o estudo de caso é utilizado, inclui-se a condução de dissertações de mestrado e teses de doutorado no meio acadêmico (Yin, 1994).

A metodologia de pesquisa constitui-se na elaboração e aplicação de uma modelagem matemática para controle estatístico multivariado de processos industriais. A verificação da metodologia teórica proposta será efetuada através da aplicação do mesmo em um estudo de caso, utilizando-se para tanto um processo industrial em batelada.

1.5. Estrutura da Dissertação

O trabalho proposto está estruturado em cinco capítulos, os quais estão descritos a seguir.

No capítulo 1, é apresentado o tema abordado, as justificativas para a escolha do mesmo, os objetivos a serem alcançados os métodos para alcançá-los e as limitações deste trabalho.

No capítulo 2 é realizada uma revisão bibliográfica dos assuntos pertinentes ao tema principal da dissertação. Através desta revisão procura-se apresentar, de forma clara e detalhada, o conhecimento consolidado e as pesquisas realizadas até o momento na área de Controle Estatístico Multivariado de Processos.

No capítulo 3 é proposta, de forma genérica, uma metodologia para o controle estatístico multivariado baseado em componentes principais para processos em batelada.

No capítulo 4 é apresentado um estudo de caso realizado em uma empresa do ramo de bebidas.

O capítulo 5 é reservado para as conclusões e para a proposição de possíveis extensões a este trabalho.

1.6. Limitações

As limitações da metodologia proposta nesta dissertação são apresentadas a seguir.

A primeira limitação refere-se ao fato de que a abordagem de controle proposta só pode ser implantada em processos em batelada informatizados. Isto se deve à necessidade durante a realização de cada batelada, de dispor de medições *on line* das variáveis do processo. Essas informações do processo serão usadas na construção de uma distribuição de referência robusta para as variáveis de processo.

Outra limitação do presente trabalho refere-se ao fato de que a metodologia proposta aplica-se somente aos processos em bateladas que se realizem com tempo fixo de duração. O monitoramento de processos industriais realizados em bateladas com tempo variável de duração não foi abordado no trabalho.

Ressalta-se que a metodologia proposta na dissertação utiliza a ACPM (Análise de Componentes Principais *Multiway*), que considera somente a estrutura de correlação das variáveis de processo X na elaboração das cartas de controle e, conseqüentemente, monitora o processo através de amostras futuras das variáveis de processo. Como técnica alternativa à ACPM para monitoramento de processos em bateladas tem-se a PELM (Projeção de Estruturas Latentes *Multiway*), que permite a utilização simultânea das informações contidas nas variáveis de qualidade das matérias primas Z , nas variáveis de processo X , e nas variáveis de qualidade do produto final Y . A exploração teórica detalhada desta técnica e sua aplicação prática não foram contempladas nesta dissertação.

Além disso, outra limitação de ordem teórica deste trabalho, refere-se aos processos industriais complexos cujos dados apresentam fortes características não-lineares, que não podem ser monitorados através de métodos lineares. Para o efetivo monitoramento destes processos, técnicas não lineares vêm sendo apresentadas na literatura, incluindo ACP não-linear, PEL não-linear, análise de correlação não-linear e procedimentos envolvendo redes neurais. O desenvolvimento teórico dessas técnicas não constitui o objetivo desta dissertação, sendo uma de suas limitações.

Finalmente, no estudo de caso, apresentado no capítulo 4 desta dissertação, desenvolveu-se a carta de controle para o monitoramento *on line* do processo em bateladas em estudo. Mas, para viabilizar a implantação real desse controle de processos, são necessárias ferramentas computacionais. Estas ferramentas computacionais devem incorporar a teoria proposta de forma a fornecer aos operadores, em tempo real, informações sobre o processo gerando cartas e gráficos de controle. O desenvolvimento dessas ferramentas computacionais não constitui o objetivo desta dissertação, sendo também uma de suas limitações.

CAPÍTULO 2

2. REVISÃO BIBLIOGRÁFICA

Neste capítulo, faremos uma introdução aos métodos de projeção multivariada de dados e ao controle estatístico de processos. Apresentaremos uma revisão dos métodos de projeção multivariada de dados e das cartas de controle multivariadas mais difundidas na literatura.

2.1. Métodos de Projeção Multivariada de Dados

No controle estatístico de processos (CEP), tanto as variáveis de qualidade do produto (Y) como as variáveis de processo (X) podem ser monitoradas. As variáveis de qualidade do produto (viscosidade e dureza, por exemplo), são obtidas a partir de amostras coletadas do processo e analisadas no laboratório de controle de qualidade. As variáveis de processo (temperaturas, pressões e fluxos, por exemplo), são obtidas a partir de amostras coletadas diretamente do processo.

A maioria dos procedimentos de CEP baseia-se no monitoramento de um pequeno número de variáveis, de qualidade ou de processo. Isto se torna inadequado para os processos industriais modernos onde, devido a automação industrial, grandes quantidades de dados são coletados rotineiramente no processo (Caten *et al.*, 2000). Além disso, as variáveis podem ser medidas em tempo real e em pontos temporais próximos; nessas condições, a suposição de independência amostral, necessária para CEP univariado, dificilmente é verificada.

Em processos industriais, tanto variáveis de qualidade como variáveis de processo possuem uma natureza multivariada, devendo ser levada em conta a correlação entre as

variáveis. Desta forma, a combinação de todas as variáveis de qualidade define a qualidade de um produto. Assim, o desempenho dos processos é mais fortemente influenciado pelo conjunto das variáveis de processo do que por cada variável de processo individualmente (MacGregor, 1995).

A natureza multivariada dos dados de processo (medidas das variáveis de qualidade ou de processo) ocasiona alguns problemas que dificultam o monitoramento no CEP.

O primeiro deles diz respeito à dimensão do conjunto de dados de processo, que normalmente é muito grande. Além disso, os dados apresentam-se fortemente correlacionados. Poucos eventos comuns definem o desempenho do processo ao longo do tempo. Assim, medições de diferentes variáveis expressam, de maneira distinta, os efeitos do mesmo conjunto de eventos sobre o processo. Quando um evento especial ocorre no processo, altera-se não apenas a magnitude de cada variável, mas também a relação entre elas (Kourty & MacGregor, 1996).

Um outro problema ocorre quando a variação natural de uma variável monitorada ao longo do tempo é descrita por um perfil. Nestes casos, a média da variável de interesse, em condições normais de processo, muda ao longo do tempo. Neste contexto, não apenas a relação entre todas as variáveis ao longo do tempo é importante, mas também o histórico inteiro de suas trajetórias.

Nos processos em bateladas, o histórico das trajetórias das variáveis fornece todas as informações relativas ao desempenho do processo em cada batelada. A partir desses dados, é possível construir um modelo empírico que caracterize a operação satisfatória das bateladas. As maiores dificuldades estão em como trabalhar com a grande quantidade de medidas das variáveis, suas variações no tempo e sua estrutura fortemente correlacionada, além da natureza não linear das operações em bateladas (Nomikos & MacGregor, 1995).

Além disso, em processos em bateladas nem todos os dados sobre as trajetórias das variáveis estão disponíveis antes do término da batelada. Para o monitoramento do processo através de cartas de controle multivariadas, são necessários todos os dados acerca de todas as variáveis em todos os tempos.

Para solucionar os problemas provenientes da natureza multivariada dos processos e da falta de dados completos da batelada antes de seu término, utilizam-se métodos de projeção multivariada no tratamento estatístico dos dados de processo. Estes métodos simplificam a análise dos dados e, conseqüentemente, os problemas no monitoramento do processo.

Os métodos de projeção multivariada têm como objetivo reduzir a dimensionalidade de um conjunto de dados. A informação contida no conjunto de dados originais é projetada em um espaço dimensionalmente menor, chamado espaço de variáveis latentes, definido por vetores latentes. Os vetores latentes são combinações lineares dos dados provenientes das variáveis originais e definem o novo conjunto de dados a ser monitorado no CEP.

Entre os métodos de projeção multivariada mais utilizados estão a Análise de Componentes Principais (ACP; Jackson, 1991), a Análise de Correlação Canônica (ACC; Johnson & Wichern, 1992) e os Mínimos Quadrados Parciais (MQP; Martens & Naes, 1989). No presente trabalho, serão abordados os métodos ACP e MQP. O método ACP, em particular, constitui-se no método escolhido para tratamento dos dados na proposta metodológica desenvolvida e no estudo de caso apresentados nesta dissertação.

2.1.1 Análise de Componentes (ACP)

A Análise de Componentes Principais (ACP) é uma técnica estatística utilizada para resumir informações em conjuntos de dados multivariados. Considere um conjunto de dados composto de realizações de p variáveis aleatórias. Tal conjunto de dados possui p componentes principais (CP), sendo cada CP formado por uma combinação linear distinta das p variáveis aleatórias originais. Essas combinações são determinadas através de manipulações algébricas na matriz de covariâncias das p variáveis, conforme apresentado mais adiante. Cada CP captura uma *direção de variabilidade* do conjunto de dados originais. As direções capturadas por cada CP são ortogonais entre si, isto equivale a dizer que os CP são variáveis aleatórias independentes.

O objetivo da determinação de CP pode ser melhor entendido ao considerar-se que, via de regra, grande parte da variabilidade do conjunto de dados pode ser descrito por menos que a totalidade dos CP. Assim, pode-se reduzir a dimensionalidade do conjunto de dados (isto é, o número de CP que o descrevem) sem perda significativa de informações. Além disso, substitui-se um grande número de variáveis aleatórias *dependentes* por um número menor de variáveis aleatórias *independentes* (os CP). A Análise de CP tem sido utilizada com sucesso em diversas áreas do conhecimento; alguns exemplos de aplicação são apresentados em Hair, Jr. *et al.* (1995).

Conforme mencionado acima, CP são combinações lineares de variáveis aleatórias, identificadas por X_i , $i = 1, \dots, p$, com realizações designadas por $(x_{i1}, x_{i2}, \dots, x_{in})$. Essas combinações são obtidas a partir da matriz de covariâncias associada às p variáveis aleatórias.

A determinação algébrica dos CP é apresentada na sequência. Na exposição que se segue, matrizes são identificadas por letras maiúsculas em negrito, por exemplo, \mathbf{A} , e seus transpostos por \mathbf{A}^t ; vetores são identificados por letras minúsculas em negrito, por exemplo, \mathbf{a} , e seus transpostos por \mathbf{a}^t .

Seja Σ a matriz de covariâncias, de dimensão $(p \times p)$, associada à matriz de variáveis aleatórias $\mathbf{X} = [X_1, X_2, \dots, X_p]$. A dimensão de \mathbf{X} é $(n \times p)$, ou seja, dispõe-se de n observações de cada variável aleatória. O vetor \mathbf{x}^t denota uma linha qualquer de \mathbf{X} . Os p autovalores de Σ são designados por λ_i , $i = 1, \dots, p$, e os p autovetores designados por \mathbf{v}_i , $i = 1, \dots, p$, com elementos dados por (v_{i1}, \dots, v_{ip}) . Os autovalores e autovetores de uma matriz quadrada não-singular qualquer são determinados através de sua equação característica; ver Strang (1988), p. 246. Associado a cada autovalor λ_i , existe um autovetor \mathbf{v}_i . Assim, os pares $(\lambda_1, \mathbf{v}_1)$, $(\lambda_2, \mathbf{v}_2)$, ..., $(\lambda_p, \mathbf{v}_p)$ correspondem aos autovalores e autovetores de Σ , com autovalores arranjados tal que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

O $i^{\text{ésimo}}$ CP pode ser obtido pela expressão (Seber, 1984):

$$\begin{aligned} X_1 v_{i1} + X_2 v_{i2} + \dots + X_p v_{ip} &= \mathbf{x}^t \mathbf{v}_i \\ t_i = \mathbf{v}_i \mathbf{x}^t &= v_{i1} X_1 + v_{i2} X_2 + \dots + v_{ip} X_p. \end{aligned} \quad (2.1)$$

Na equação (2.1), os elementos do autovetor \mathbf{v}_i funcionam como pesos de importância de X_1, \dots, X_p na composição do $i^{\text{ésimo}}$ CP, e são denominados *cargas* do CP. Em CP obtidos a partir de variáveis aleatórias padronizadas, a magnitude da carga associada a uma variável descreve a sua importância relativa na composição dos CP. É bastante comum, como apresentado no exemplo mais adiante nesta seção, que alguns CP sejam majoritariamente descritos por uma única variável aleatória. O valor observado de t_i para um determinado vetor (x_1, \dots, x_p) de realizações de (X_1, \dots, X_p) é denominado *score* do CP. Cada CP apresenta n scores, correspondendo ao número total de realizações disponíveis de cada variável aleatória.

Cada CP descreve uma das direções de variabilidade do conjunto de dados. Em outras palavras, cada CP descreve uma porção da variância total apresentada pelo conjunto das p variáveis aleatórias. Sabe-se que a diagonal principal de Σ apresenta os valores de variância associados a cada variável aleatória. Logo, a variância total no conjunto de dados é dada pela soma dos elementos da diagonal principal de Σ ; este valor é igual à soma dos autovalores de Σ (já que o traço de uma matriz quadrada é igual à soma de seus autovalores; ver Rencher (1995) p. 39). Assim, pode-se utilizar os autovalores de Σ para determinar quanto da

variabilidade total presente no conjunto de dados está sendo descrita por cada CP. Mais especificamente, a proporção da variância total descrita pelo $i^{\text{ésimo}}$ CP é dada por:

$$\text{Proporção da variância descrita pelo } i^{\text{ésimo}} \text{ CP} = \frac{l_i}{l_1 + \dots + l_p}. \quad (2.2)$$

Através da equação (2.2), pode-se definir o número de CP a serem utilizados na representação do conjunto de dados originais. Essa definição é essencialmente qualitativa. Por exemplo, um analista pode considerar satisfatória uma representação dada por um grupo de CP que descreva 90% da variabilidade original. Algumas regras auxiliares na definição do número de CP que melhor representam o conjunto de dados em estudo são apresentadas na próxima seção deste capítulo.

O coeficiente de correlação entre o $i^{\text{ésimo}}$ componente principal (t_i) e a variável X_p é dado na equação (2.3). Tal coeficiente mede a importância da variável X_p na composição do escore t_i do CP em questão.

$$r_{t_i X_p} = \frac{v_{ip} \sqrt{l_i}}{S_{pp}} \quad (2.3)$$

onde σ_{pp} é o elemento da diagonal principal da matriz Σ .

Os conceitos apresentados acima são agora ilustrados através de um exemplo (Johnson & Wichern, 1992). Considere três variáveis aleatórias X_1, X_2, X_3 , com matriz de covariâncias Σ dada por:

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Os autovalores e autovetores de Σ são:

$$\lambda_1 = 5,83 \quad \mathbf{v}_1^t = [0,383, -0,924, 0],$$

$$\lambda_2 = 2,00 \quad \mathbf{v}_2^t = [0, 0, 1],$$

$$\lambda_3 = 0,17 \quad \mathbf{v}_3^t = [0,924, 0,383, 0].$$

Os CP são obtidos diretamente dos autovetores, sendo dados abaixo:

$$t_1 = \mathbf{xv}_1^t = 0,383X_1 - 0,924X_2$$

$$t_2 = \mathbf{xv}_2^t = 1,000X_3$$

$$t_3 = \mathbf{xv}'_3 = 0,924X_1 + 0,383X_2$$

Observe que o 2º CP captura uma direção de variância representada exclusivamente pela variável aleatória X_3 .

A proporção da variância total explicada pelos CP, obtida utilizando a equação (2.2) é:

$$CP_1 = 72,87\% \quad CP_2 = 25,00\% \quad CP_3 = 2,12\%.$$

É fácil verificar que aproximadamente 98% da variabilidade total presente nos dados está representada pelos dois primeiros CP. Assim, pode-se utilizar t_1 e t_2 no lugar das variáveis originais, sem grandes perdas de informação. Utilizando os t_1 e t_2 , reduz-se a dimensionalidade do problema de três variáveis (variáveis aleatórias originais) para duas variáveis (CP).

2.1.1.1 Seleção do Número de CP

Em muitas aplicações, deve-se decidir qual o número de CP necessários para representar de maneira eficiente o conjunto de dados originais. Neste trabalho, em particular, deseja-se definir o número de CP necessários para construir um modelo de ACP que descreva adequadamente o desempenho de bateladas em operação normal. Para tanto, muitos critérios podem ser utilizados, baseando-se em testes de significância ou procedimentos gráficos (Jackson, 1991).

Rencher (1995) apresenta quatro critérios para a definição do número adequado de CP.

O primeiro critério baseia-se em selecionar o número de CP suficiente para representar uma porcentagem específica da variância total, por exemplo, 80% da variância total do conjunto de dados. Este critério é essencialmente qualitativo.

O segundo critério baseia-se em excluir os CP com autovalores menores que a média dos autovalores, dada por:

$$\text{Média dos Autovalores} = \sum_{i=1}^p \frac{\lambda_i}{p} \quad (2.4)$$

Para a matriz de correlações, esta média é igual a 1. Como descrito na seção anterior, a variância total no conjunto de dados é dada pela soma dos elementos da diagonal principal de Σ ; este valor é igual à soma dos autovalores de Σ . Assim, a média dos autovalores é também a média da variância das variáveis individuais. Dessa forma, este critério retém os CP que representam variância maior que a média das variâncias das variáveis.

O terceiro critério proposto utiliza o *scree graph*, um gráfico onde o valor de λ_i é plotado no eixo vertical, para todos os i s ($i = 1, \dots, p$), no eixo horizontal. Este gráfico representa uma tentativa visual de encontrar uma separação natural entre autovalores grandes e pequenos. A Figura 2.1 apresenta um exemplo de *scree graph* (Rencher, 1995). Nota-se no gráfico que os últimos autovalores são bem menores que os dois primeiros e formam aproximadamente uma linha reta no gráfico. Existe um ponto de separação no gráfico, distinguindo os dois primeiros autovalores dos quatro restantes (este ponto de separação é chamado de “cotovelo” do *scree graph*; Jackson, 1991). A recomendação em Rencher (1995) é selecionar os CP correspondentes àqueles autovalores situados antes do ponto de inflexão no gráfico; Cattell & Jaspers *apud* Jackson (1991), por sua vez sugerem incluir também o CP correspondente ao autovalor no ponto de inflexão no grupo de CP selecionados.

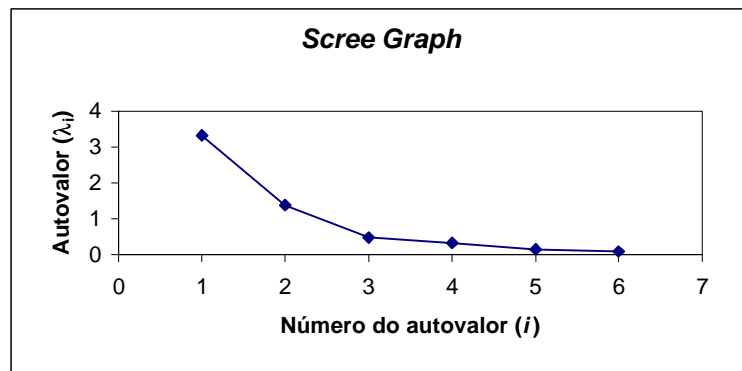


Figura 2.1 – Exemplo de *Scree graph* para autovalores (Rencher, 1995)

A partir da análise do *Scree graph* na Figura 2.1, seriam selecionados (i) os CP correspondentes aos dois primeiros autovalores (Rencher, 1995) ou (ii) os CP correspondentes aos três primeiros autovalores, incluindo o ponto de quebra (Cattell & Jaspers *apud* Jackson, 1991).

O *Scree graph* para o exemplo (Johnson & Wichern, 1992) apresentado na seção 2.1.1 deste capítulo é mostrado na Figura 2.2.

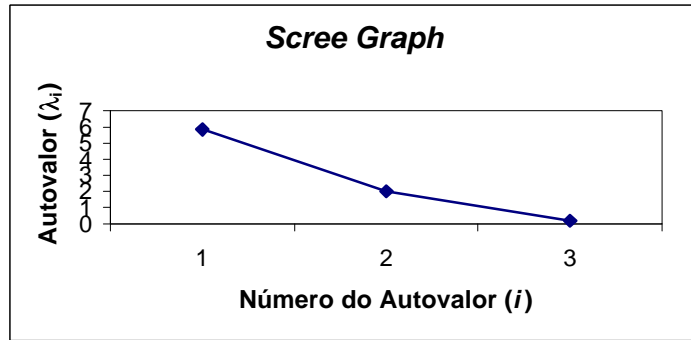


Figura 2.2 – *Scree graph* para o exemplo dado na seção 2.1.1

Na prática, o ponto de inflexão pode não ser tão distinto como na Figura 2.1, ou ainda pode existir mais de um ponto de inflexão no gráfico. Nestes casos, o critério que utiliza o *scree graph* para a seleção do número adequado de CP a serem retidos pode não levar a resultados conclusivos.

O quarto critério proposto por Rencher (1995) para seleção do número adequado de CP para representar um conjunto de dados, baseia-se em testes de significância. Estes testes supõem dados multivariados normalmente distribuídos, o que não é necessário para a estimação dos CP na ACP.

Para testar a significância dos maiores autovalores e, por conseguinte reter os CP a eles correspondentes, testa-se a hipótese de que os últimos k autovalores obtidos de Σ são pequenos e aproximadamente iguais. Este teste chama-se teste de Bartlett (Jackson, 1991), com hipótese nula dada na equação 2.5 (os autovalores devem ser ordenados em ordem crescente, de λ_1 a λ_p).

$$H_{0k}: \lambda_{p-k+1} = \lambda_{p-k+2} = \dots = \lambda_p \quad (2.5)$$

A suposição deste teste é que os primeiros CP capturam todas as dimensões essenciais do conjunto de dados, enquanto que os últimos CP refletem a variação aleatória nos dados. Se H_0 for verdadeira os últimos k autovalores amostrais tendem a apresentar um padrão semelhante ao mostrado na linha reta do *scree graph* da Figura 2.1.

Para testar a hipótese H_{0k} , calcula-se a média dos últimos k autovalores de Σ , dada na equação (2.6), e realiza-se o teste estatístico dado na equação (2.7).

$$\bar{\Gamma} = \sum_{i=p-k+1}^p \frac{\lambda_i}{k} \quad (2.6)$$

$$u = \left(n - \frac{2p+11}{6} \right) \left(k \ln \bar{l} - \sum_{i=p-k+1}^p \ln l_i \right) \quad (2.7)$$

Pode-se demonstrar que a quantidade na equação (2.7) aproxima-se de uma distribuição do χ^2 . Rejeita-se H_0 se $u \geq \chi^2_{\alpha, v}$, onde α é nível de significância e v é dado na equação (2.8).

$$v = \frac{1}{2}(k-1)(k+2) \quad (2.8)$$

O procedimento de teste inicia testando a hipótese $H_{02}: \lambda_{p-1} = \lambda_p$. Se a hipótese H_{02} for aceita, testa-se $H_{03}: \lambda_{p-2} = \lambda_{p-1}$, e assim, sucessivamente, até que H_{0k} seja rejeitada para algum valor de k .

A maior desvantagem do critério acima é que, na prática, ele tende a reter um número maior de CP do que o necessário.

Alguns outros métodos para seleção do número adequado de CP propostos em Jackson (1991) e Nomikos & MacGregor (1995) são apresentados a seguir.

Um método rápido para seleção do número de CP é a regra "*broken-stick*" (Jolliffe *apud* Nomikos & MacGregor, 1995). Este método baseia-se no fato de que, se um segmento de reta de comprimento unitário for aleatoriamente dividido em z segmentos, o comprimento esperado do $r^{\text{ésimo}}$ maior segmento será dado por:

$$G = 100 \frac{1}{z} \sum_{i=r}^z 1/i \quad (2.9)$$

De acordo com o teste *broken-stick*, sempre que a porcentagem da variância explicada por cada CP for maior que o valor calculado para G , deve-se reter os respectivos CP. Este método deve ser utilizado somente quando as variáveis aleatórias no conjunto de dados forem padronizadas. Apesar de bastante impreciso, o teste do *broken-stick* ainda é um dos métodos mais rápidos para julgar se um CP adiciona alguma informação estrutural sobre a variância dos dados ou explica apenas o ruído aleatório.

Um outro método para determinação do número ideal de CP é o da validação cruzada, proposto por Wold; Eastment & Krzanowski *apud* Jackson (1991). Esta abordagem é recomendada quando o objetivo do estudo é construir um modelo ACP para fins de previsão de valores futuros. O método mostra como o poder de previsão de um modelo ACP aumenta com a adição de mais um CP.

A validação cruzada consiste em dividir aleatoriamente a base de dados amostrais em g grupos de n/g observações cada. A ACP é realizada da seguinte maneira. Remove-se um grupo do conjunto de dados, geram-se os CP para os dados remanescentes e, a partir do modelo resultante, tenta-se estimar as observações do grupo excluído. Em seguida, calcula-se a diferença entre os valores preditos e observados para as variáveis aleatórias e gera-se a estatística *Press* (*PRediction Error Sum of Squares* – Soma do quadrado dos erros de predição). Retorna-se o grupo à base de dados e repete-se o procedimento com um segundo grupo. São realizadas as previsões dos CP para cada grupo excluído e também são obtidos os valores da soma do quadrado dos erros de predições destes grupos. A média do somatório dos quadrados dos erros de todas as predições origina a estatística *Press*. O procedimento inteiro é repetido construindo-se modelos com dois CP, modelos com três CP, e assim por diante. Para cada modelo calcula-se a estatística *Press* (Jackson, 1991). Uma maneira de escolher a dimensão do modelo é escolher aquele com menor valor para a estatística *Press* (Nomikos & MacGregor, 1995). Esta abordagem é recomendada quando o objetivo do estudo é a construção de um modelo contra o qual futuros conjuntos de dados serão avaliados.

Wold *apud* Nomikos & MacGregor (1995) e Krzanowski *apud* Nomikos & MacGregor (1995) propuseram dois critérios para escolha do número ótimo de CP utilizando a estatística *Press*. Wold verificou a razão R dada na equação (2.10), onde RSS_r é o somatório dos resíduos quadrados após a inclusão do $r^{\text{ésimo}}$ CP baseado no modelo ACP, construído usando-se a base de dados inteira e $Press_r$ é a estatística *Press* para o modelo com r CP.

$$R = \frac{Press_r}{RSS_{r-1}} \quad (2.10)$$

Este critério compara o poder de predição de um modelo baseado em r CP com a soma dos quadrados das diferenças entre os dados observados e o preditos utilizando-se $r-1$ CP. Um valor de R maior que a unidade indica que o $r^{\text{ésimo}}$ CP não melhorou o poder de predição do modelo, sendo melhor utilizar apenas $r-1$ CP.

Krzanowski *apud* Nomikos & MacGregor (1995) sugeriu testar-se a razão W , dada na equação (2.11), abaixo. Esta abordagem aplica-se ao contexto de utilização da ACP em processos produtivos realizados em batelada.

$$W = \frac{\left(\frac{(\text{Press}_{r-1} - \text{Press}_r)}{D_m} \right)}{\left(\frac{\text{Press}_r}{D_r} \right)} \quad (2.11)$$

onde,

$$D_m = I + JK - 2r \quad (2.12)$$

$$D_r = JK(I-1) - \sum_{i=1}^r (I + JK - 2i) \quad (2.13)$$

nas equações (2.12) e (2.13) I indica o número de bateladas, J o número de variáveis monitoradas, K a variável tempo. Os números D_m e D_r indicam os graus de liberdade necessários para excluir o $r^{\text{ésimo}}$ CP e os graus de liberdade remanescentes após a exclusão do $r^{\text{ésimo}}$ CP, respectivamente. Esta estatística fornece a razão entre a melhoria no poder de predição pela adição do $r^{\text{ésimo}}$ CP e o valor predito do mesmo CP. Se W for maior que a unidade, vale a pena inserir o $r^{\text{ésimo}}$ CP no modelo.

A utilização do procedimento de validação cruzada na determinação do número de CP necessários na elaboração do modelo de ACP apresenta algumas desvantagens. Não existe teste estatístico definitivo para o procedimento da validação cruzada. O principal problema está em não se saber, a priori, com quantos graus de liberdade se inicia e nem quantos graus de liberdade são perdidos com cada CP durante a análise (Nomikos & MacGregor, 1995). O número real de graus de liberdade de um processo é dado pelo número de variáveis que se alteram independentemente.

A dimensão do modelo ACP é definida pelo número de CP utilizados nele. O número de CP utilizados em um modelo ACP é também denominado número de fatores do modelo, por exemplo, um modelo ACP que utiliza 5 CP pode ser referido como um modelo de ACP com 5 fatores. Quando um processo possui forte não-linearidade, o número de fatores num modelo ACP determinado através da validação cruzada é geralmente maior que o número de graus de liberdade do processo. Neste caso, deve-se utilizar uma modelagem não-linear para o processo (Chen & McAvoy, 1996).

2.1.1.2 Interpretação Geométrica dos CP

Geometricamente, os CP podem ser representados por um sistema de eixos ortogonais com origem posicionada na média do conjunto de dados. O primeiro CP segue a direção de maior variabilidade dos dados. O segundo CP segue a segunda direção de maior variabilidade, ortogonal ao primeiro CP. Os demais CP são posicionados no espaço p -dimensional seguindo a mesma lógica. Dessa forma, a ACP define um novo sistema de eixos coordenados para a projeção das variáveis, resultante da translação do sistema coordenado original da origem dos dados para a média do conjunto de dados, e da rotação dos eixos coordenados na direção de máxima variância (Johnson & Wichern, 1992).

Por exemplo, suponha duas variáveis aleatórias X_1, X_2 , seguindo uma distribuição Normal bivariada $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com vetor de médias dado por $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. A função de densidade de $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é dada pela equação da elipse centrada em $\boldsymbol{\mu}$, conforme a equação (2.14).

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2 \quad (2.14)$$

Na equação (2.14), \mathbf{x} denota o vetor que contém as realizações das variáveis aleatórias X_1 e X_2 . Os eixos da elipse em (2.14) são dados por $\pm c(l_i)^{1/2} \mathbf{v}_i$, com $i = 1, 2$. Os CP $t_1 = \mathbf{x}' \mathbf{v}_1$ e $t_2 = \mathbf{x}' \mathbf{v}_2$ estão posicionados nas direções dos eixos da elipse de densidade constante, conforme ilustrado na Figura 2.3.

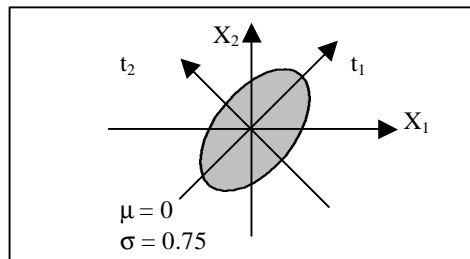


Figura 2.3 - Elipse de densidade constante e CP t_1 e t_2 .

A elipse de densidade constante define o novo espaço de projeção das variáveis latentes, que são os CP. A transformação dos dados em componentes principais é uma projeção dessas observações nos principais eixos da elipse. Esta elipse define o novo espaço de variáveis, idealmente com dimensão $k < p$.

Quando o contorno dado pela equação (2.14) é aproximadamente circular, isto é, quando os autovalores de $\boldsymbol{\Sigma}$ são iguais ($l_1 = l_2 = \dots = l_p$), a variação dos dados é homogênea

em todas as direções. Neste caso não é possível representar adequadamente os dados em dimensões menores que a dimensão original p (Johnson & Wichern, 1992).

2.1.1.3 Análise dos Resíduos dos CP

A ACP transforma o conjunto de dados originais \mathbf{X} ($n \times p$), de dimensão p , em k componentes principais, tal que, idealmente, $k < p$. Dessa forma, o conjunto de dados originais é projetado em um espaço ortogonal de dimensão k (Kourti & MacGregor, 1996).

$$\mathbf{X} = \sum_{i=1}^k t_i \mathbf{v}_i^t + \sum_{i=k+1}^p t_i \mathbf{v}_i^t = \hat{\mathbf{X}} + \mathbf{E} \quad (2.15)$$

Na equação (2.15) $\hat{\mathbf{X}}$ indica os valores estimados de \mathbf{X} (matriz dos dados originais) a partir dos k componentes principais retidos no modelo e \mathbf{E} é a matriz de resíduos.

O somatório do quadrado dos resíduos (Q) dos CP é dado pela equação (2.16) e representa o somatório do quadrado da distância de $(\mathbf{x} - \hat{\mathbf{x}})$. Q mede a contribuição dos componentes principais que não foram utilizados no modelo se fossem a ele adicionados (Jackson, 1991).

$$Q = (\mathbf{x} - \hat{\mathbf{x}})' (\mathbf{x} - \hat{\mathbf{x}}) \quad (2.16)$$

Tipicamente, apenas os primeiros poucos CP são considerados suficientes para resumir os dados, conforme apresentado na seção 2.1.1.1 deste capítulo. Entretanto, os últimos poucos componentes também podem fornecer informações úteis. Os últimos CP capturam pequenas variâncias. Se a variância de um CP é zero ou aproximadamente zero, o componente representa uma relação linear entre as variáveis, que é essencialmente constante. (Rencher, 1995).

No controle estatístico de processos, juntamente com a carta de controle multivariada, devem ser monitorados o gráfico dos escores dos componentes principais e o gráfico dos resíduos, devido a sua habilidade para detectar erros nas medições. Para os processos em bateladas, objeto da pesquisa nesta dissertação, esses dois gráficos têm interpretações específicas. O gráfico dos escores representa a projeção do histórico de cada batelada no plano reduzido definido pelos CP. O gráfico dos resíduos representa o quadrado da distância de cada batelada perpendicular ao plano de projeção definido pelos CP. Caso alguma variação anormal nos dados não seja capturada na carta dos escores, ela pode ser capturada na carta dos resíduos.

2.1.2 Mínimos Quadrados Parciais (MQP)

Mínimos quadrados parciais (MQP) ou Projeção de Estruturas Latentes (PEL) é um método alternativo de projeção multivariada de dados. Este método é aplicado quando tem-se duas matrizes de dados: a matriz dos dados de processo \mathbf{X} e a matriz dos dados de qualidade \mathbf{Y} , obtida medindo-se o desempenho de itens que emergem do processo. O MQP reduz simultaneamente a dimensão de \mathbf{X} e \mathbf{Y} , não apenas encontrando as maiores direções de variabilidade nos dados de processo, mas especificamente as maiores direções de variabilidade em \mathbf{X} que relatam as variações nos dados de qualidade \mathbf{Y} .

No MQP, a covariância amostral é dada pela matriz $\mathbf{X}^t\mathbf{Y}\mathbf{Y}^t\mathbf{X}$. A primeira variável latente $t_1 = \mathbf{w}_1^t\mathbf{x}$ é a combinação linear das variáveis X que maximiza a covariância entre \mathbf{X} e \mathbf{Y} . O primeiro vetor de cargas \mathbf{w}_1 é o primeiro autovetor da matriz $\mathbf{X}^t\mathbf{Y}\mathbf{Y}^t\mathbf{X}$. O vetor escore referente à primeira variável latente é obtido, para cada uma das m observações preliminares \mathbf{x}'_k que compõem a matriz \mathbf{X} , da seguinte forma: $\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1$. Após o cálculo do vetor escore \mathbf{t}_1 para o primeiro componente, o novo vetor de cargas correspondente, \mathbf{p}_1 , é obtido através da regressão das colunas de \mathbf{X} em \mathbf{t}_1 , como mostrado na equação (2.17).

$$\mathbf{p}_1 = \frac{\mathbf{X}^t\mathbf{t}_1}{\mathbf{t}_1^t\mathbf{t}_1} \quad (2.17)$$

A segunda variável latente, ortogonal à primeira, é calculada a partir da nova matriz de covariâncias $\mathbf{X}_2^t\mathbf{Y}_2\mathbf{Y}_2^t\mathbf{X}_2$, onde \mathbf{X}_2 e \mathbf{Y}_2 são calculados de acordo com as equações (2.18) e (2.19), respectivamente.

$$\mathbf{X}_2 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^t \quad (2.18)$$

$$\mathbf{Y}_2 = \mathbf{Y} - \mathbf{t}_1\mathbf{q}_1^t \quad (2.19)$$

onde \mathbf{q}_1 é obtido por regressão das colunas de \mathbf{Y} em \mathbf{t}_1 , conforme a equação (2.20).

$$\mathbf{q}_1 = \frac{\mathbf{t}_1^t\mathbf{Y}}{\mathbf{t}_1^t\mathbf{t}_1} \quad (2.20)$$

A segunda variável latente é dada por $\mathbf{t}_2 = \mathbf{w}_2^t\mathbf{x}$, onde \mathbf{w}_2 é o primeiro autovetor da matriz $\mathbf{X}_2^t\mathbf{Y}_2\mathbf{Y}_2^t\mathbf{X}_2$. Como na ACP, os vetores de escores ($\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p$) bem como os vetores de cargas ($\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$) são ortogonais entre si (Kourti & MacGregor, 1996).

Para grandes conjuntos de dados autocorrelacionados, Kourti & MacGregor (1996) sugerem calcular as variáveis latentes do método MQP através do algoritmo NIPALS, proposto por Geladi & Kowalski (1986a).

A diferença do método MQP com relação a ACP é que as variáveis latentes são calculadas utilizando-se ambas as matrizes de dados, \mathbf{X} e \mathbf{Y} . A carta de controle multivariada baseada em MQP, monitora as variações das variáveis de processo que são mais relevantes para as variáveis de qualidade do produto.

Em alguns processos, pode-se decompor as variáveis de processo \mathbf{X} em blocos $\mathbf{X}_1, \mathbf{X}_2, \dots$, onde, em cada bloco que representa uma parte (uma seção ou unidade de produção) do processo, existe um grupo peculiar de eventos subjacentes ocorrendo quando o processo está sob controle.

Nestes processos, um método de projeção multivariada de dados adequado é o MQP *Multi Block*. A utilização do MQP *Multi Block* não equivale a realizar a MQP, separadamente, em cada bloco. Neste método de projeção multivariada de dados, os blocos são analisados conjuntamente e ponderados de forma a maximizar sua covariância com o conjunto de variáveis de qualidade \mathbf{Y} . Os princípios deste método e o algoritmo para sua utilização podem ser encontrados em Wold *apud* Kourti & MacGregor (1996) e Wangen & Kowalski (1988). Aplicações deste método de projeção no monitoramento de processos químicos são apresentadas em Kourti & MacGregor (1996) e MacGregor *et al.* (1994).

2.2. Fundamentos das Cartas de Controle Estatístico de Processos

Todos os processos produtivos apresentam uma variabilidade natural. Quando esta variabilidade for originada devido a causas comuns, o processo é considerado estável, sob controle estatístico.

Além das variações oriundas das causas comuns, processos produtivos podem sofrer o efeito de variações originadas por causas especiais. Estas variações são relativamente grandes quando comparadas às anteriores e sua existência costuma resultar em desempenho inaceitável do processo. Causas especiais devem ser identificadas e corrigidas para que o processo permaneça dentro de um padrão esperado de desempenho.

O objetivo do controle estatístico de processo é monitorar o desempenho de processos ao longo do tempo, com vistas a detectar eventos incomuns que influenciem nas propriedades determinantes da qualidade do produto final (Montgomery, 1996). Uma vez encontradas as causas especiais responsáveis pelo evento incomum, melhorias no processo e na qualidade do produto podem ser obtidas.

As cartas de controle de *Shewhart* (CC) constituem uma técnica para monitoramento da variabilidade de processos. A base da teoria das CC está na diferenciação das causas de variação na qualidade, distinguindo causas comuns de causas especiais.

As CC podem monitorar CQs como o número de produtos conformes ou não conformes (número de defeituosos), ou o número de não conformidades (número de defeitos) em uma unidade do produto. Estes tipos de cartas são chamadas de CC para atributos.

As CC para atributos são a carta p , que monitora a fração de produtos não conformes, a carta np , que monitora o número de produtos não conformes, a carta c , que monitora o número de não conformidades por unidade do produto e a carta u , que monitora o número médio de não conformidades por unidade do produto.

Outras CC monitoram características de qualidade que podem ser medidas e expressas em uma escala contínua de valores, como pressão e temperatura, por exemplo. Estas cartas, conhecidas como CC para variáveis, são detalhadas a seguir.

2.2.1 Principais Cartas de Controle para Variáveis

As CC para variáveis monitoram a localização, através da média amostral (cartas \bar{X}), e a variabilidade, através da amplitude ou do desvio-padrão amostral (cartas R e S , respectivamente) do processo em estudo. Outros tipos de CC utilizadas para monitorar a média do processo são as cartas para a soma acumulada (cartas *CUSUM* – *Cumulative Sum Control Chart*) e as cartas para a média móvel ponderada exponencialmente (cartas *EWMA* – *Exponentially Weighted Moving Average Control Chart*).

As CC são gráficos apresentando os valores de medição da variável de interesse no eixo vertical e os pontos no tempo nos quais as medições são efetuadas no eixo horizontal.

As CC de médias apresentam uma linha central que representa a média da variável de interesse quando o processo está sob controle (livre de causas especiais), e duas outras linhas que representam os limites de controle do processo (LC; ver Figura 2.4). Os limites de controle são definidos de forma a compreender a maior parte dos valores da variável, estando o processo sob controle. A cada medição da variável de interesse, compara-se o resultado obtido com os limites de controle. Quando houver pontos situados fora da região definida pelos LC, ou os pontos na carta apresentarem um comportamento sistemático ou não aleatório atípico, o processo é dito fora de controle. Um comportamento sistemático pode ser uma série de pontos consecutivos de um mesmo lado da carta, acima ou abaixo da linha central, mesmo que nenhum ponto caia fora dos limites de controle. A Figura 2.4, obtida utilizando o

software PROCEP,2000, apresenta em vermelho pontos fora da região definida pelos LC e pontos descrevendo um comportamento sistemático não-aleatório.

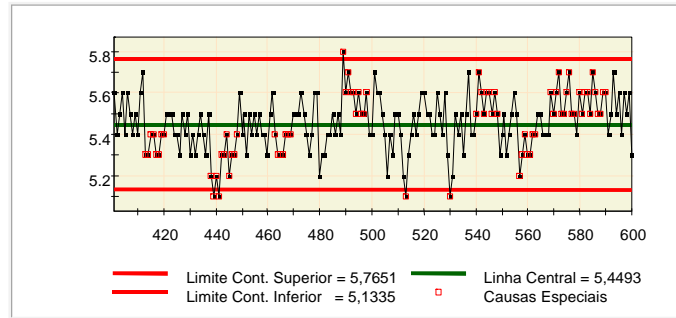


Figura 2.4 – Exemplo de carta de controle \bar{X} .

Os LC e a linha central em CC são definidos a partir de um modelo geral, onde a CQ de interesse é representada pela estatística amostral w , com média m_w e desvio-padrão s_w . O limite de controle superior (LC), a linha central (LCT) e o limite de controle inferior (LCI) são dados por (Montgomery,1996):

$$LC = \mu_w + L\sigma_w; \quad LCT = \mu_w; \quad LCI = \mu_w - L\sigma_w, \quad (2.21)$$

sendo L a distância, em desvios-padrão, entre os limites de controle e a linha central. O valor de L mais comumente utilizado é $L = 3$, o que pode ser justificado pelos bons resultados obtidos na prática.

Existem dois tipos de erros possíveis de ocorrer no monitoramento de processos através de CC. O erro tipo I é caracterizado pela ocorrência de um alarme falso. O erro tipo II é caracterizado pela não detecção de uma situação real de processo fora de controle. A escolha dos LC influencia diretamente as probabilidades de ocorrência dos erros tipo I e tipo II. Quanto mais distantes da linha central estiverem os limites, menor será a probabilidade de erro tipo I, aumentando, todavia, a probabilidade de erro tipo II. Em contrapartida, aproximar os limites de controle da linha central provoca uma situação inversa.

Um outro aspecto indispensável ao planejamento de uma CC é a determinação do tamanho da amostra e da frequência de amostragem. Para tanto, devem ser considerados o custo de amostragem, as perdas pela produção de produtos de baixa qualidade e o custo da localização de causas especiais. Além disso, devem ser levados em consideração aspectos estatísticos como a minimização de alarmes falsos, isto é, a sinalização da ocorrência de uma causa especial quando não houve alteração no processo e, ainda, a maximização da capacidade de detecção de mudanças reais no processo.

Assim, o tamanho da amostra e a frequência de amostragem devem ser definidos tendo em vista (i) o tamanho da mudança no processo a ser detectada pela CC e (ii) a rapidez com que se deseja detectar esta mudança. Dessa forma, pequenas mudanças no processo podem ser rapidamente detectadas utilizando-se grandes amostras tomadas freqüentemente, embora esta estratégia possa ser economicamente inviável (Michel & Fogliatto, 2000).

Uma outra maneira de definir o tamanho da amostra e a frequência de amostragem é através do comprimento médio de corrida (ARL - *Average Run Length*) da carta de controle. O ARL é o número médio de pontos que devem ser plotados antes que um ponto indique uma causa especial na carta de controle. O ARL pode ser usado também para avaliar o desempenho da carta de controle. Para uma carta de controle de *Shewhart*, o ARL pode ser calculado como (Montgomery, 1996):

$$ARL = \frac{1}{P} , \quad (2.22)$$

onde P é a probabilidade de um ponto exceder os limites de controle.

Os limites de controle utilizados na carta de *Shewhart* de controle são dados na equação (2.21), onde $L = Z_{\alpha/2} = \pm 3$. Assim, $\alpha = 0,0027$, ou seja, existe uma probabilidade de 0,27% da carta de controle acusar incorretamente uma mudança na média do processo. 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, esperamos retirar, em média, 370 amostras sucessivas antes que a carta gere um alarme falso.

Quando a média do processo desviar-se do valor nominal em um desvio padrão ($m = m_0 + s$), $ARL = 1/(1-\beta) = 1/0,0228 \approx 44$ amostras, onde $(1-\beta)$ é a probabilidade de que a carta detecte um desvio de 1s na média do processo. Isto significa que serão necessárias 44 amostras sucessivas para que a carta sinalize uma mudança na média do processo, de m_0 para m .

Uma prática adequada, bastante utilizada na indústria, é a escolha entre pequenas amostras tomadas freqüentemente ou grandes amostras tomadas em intervalos de tempo maiores (Michel & Fogliatto, 2000).

2.3. Cartas de Controle Multivariadas

As Cartas de Controle Multivariadas (CCM) são utilizadas no monitoramento conjunto de duas ou mais variáveis de processo ou produto. Nessas cartas, pontos p -dimensionais (ou

seja, relativos a p variáveis aleatórias ou estatísticas de interesse delas derivadas) são representados uni-dimensionalmente e plotados em gráficos similares às cartas de *Shewhart*. Desta forma, simplifica-se a tarefa de controle simultâneo de variáveis.

As CCM são particularmente recomendadas em situações onde exista correlação significativa entre as variáveis a serem monitoradas. A significância das correlações pode ser determinada aplicando-se o teste de Fisher sobre correlações amostrais; ver Freund & Simon (1997), p. 538. No caso de correlações significativas, o monitoramento das p variáveis de interesse utilizando p cartas univariadas resulta em grande número de alarmes falsos, o que é indesejável em qualquer esquema de monitoramento. Na verdade, mesmo variáveis independentes monitoradas individualmente apresentam grande incidência de alarmes falsos, o que torna seu monitoramento simultâneo, usando CCM, recomendado (Jackson, 1991).

As cartas *Qui-quadrado* e de *Hotelling* (T^2), descritas na seqüência, são cartas de controle do tipo *Shewhart*. Estas cartas utilizam informações obtidas apenas na amostragem mais recente do processo, sendo pouco sensíveis para pequenas e moderadas mudanças no vetor de médias.

As cartas para a soma acumulada (cartas *CUSUM*) e as cartas para a média móvel ponderada exponencialmente (cartas *EWMA*) foram desenvolvidas para fornecer maior sensibilidade a pequenas alterações na média do processo; estas cartas também podem ser adaptadas para contemplar o controle de qualidade multivariado (Montgomery, 1996), conforme apresentado nas seções 2.3.2 e 2.3.3.

2.3.1 Cartas de Hotelling ou cartas T^2

As CCM foram originalmente introduzidas por Hotelling (1947) para o monitoramento de médias amostrais; estas cartas são descritas a seguir.

Suponha variáveis de interesse seguindo uma distribuição Normal p -variada, com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Tomam-se amostras de tamanho n para cada uma das p variáveis de interesse (a serem monitoradas). Calcula-se a média amostral de cada variável e escreve-se o resultado num vetor de médias $\bar{\mathbf{x}}' = [\bar{X}_1, \dots, \bar{X}_p]$. A estatística a ser monitorada na CCM é:

$$\begin{aligned} (i) \chi_0^2 &= n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \quad \text{ou} \\ (ii) T^2 &= n(\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}}). \end{aligned} \tag{2.23}$$

A estatística (i) será utilizada sempre que os parâmetros populacionais μ e Σ forem conhecidos. A estatística (ii) será utilizada quando os parâmetros populacionais μ e Σ não forem conhecidos, sendo estimados por $\bar{\bar{x}}$ e \mathbf{S} . O estimador de $\bar{\bar{x}}$ é dado por (Montgomery,1996)

$$\bar{\bar{x}}^t = [\bar{\bar{X}}_1, \dots, \bar{\bar{X}}_p] \text{ e } \bar{\bar{X}}_i = \frac{1}{mn} \sum_{k=1}^m \sum_{j=1}^n x_{jki}, \quad i = 1, \dots, p, \quad (2.24)$$

onde m denota o número total de amostras de tamanho n utilizadas na estimação. O estimador da matriz $\mathbf{S} = [\bar{S}_{gh}]$ é:

$$\bar{S}_{gh} = \frac{1}{m} \frac{1}{(n-1)} \sum_{k=1}^m \sum_{j=1}^n (x_{jqk} - \bar{x}_{qk})(x_{jhk} - \bar{x}_{hk}) \quad (2.25)$$

As estatísticas (i) e (ii) na equação (2.23) representam a distância quadrada padronizada entre o vetor de médias amostrais e o vetor de médias do processo. Esta distância foi denominada por Morrison *apud* Pignatiello & Runger (1990) como *Distância Mahalanobis* e por Johnson & Wichern (1992) como *Distância Estatística*.

Na maioria das vezes, os parâmetros populacionais não são conhecidos, sendo estimados utilizando as equações (2.24) e (2.25) e dados coletados do processo em estudo, a CCM, nestes casos, é construída utilizando a estatística (ii) da equação (2.23).

A elaboração da carta de controle divide-se em duas fases distintas, cada uma delas tendo limites de controles específicos (Alt, 1984). A Fase I, consiste em utilizar a carta para verificar se o processo estava sob controle estatístico quando m amostras preliminares foram coletadas e estimar os parâmetros do processo. As estimativas dos parâmetros populacionais caracterizarão a distribuição de referência, contra a qual observações futuras do processo serão comparadas. Desta forma, os dados utilizados na determinação das estimativas devem ser coletados quando o processo apresentar operação estável e desejável. Na Fase II, os parâmetros estimados na fase anterior são utilizados na carta para monitorar amostras futuras. O limite de controle superior (LCS) da CCM para cada fase é dado abaixo (Ryan, 1989).

Fase I

$$LCS = \frac{p(m-1)(n-1)}{mn - m - p + 1} F_{\alpha, p, mn-m-p+1} \quad (2.26)$$

Fase II

$$LCS = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha, p, mn-m-p+1} \quad (2.27)$$

Nas equações acima, $F_{\alpha, p, mn-m-p+1}$ denota o valor do $\alpha^{\text{ésimo}}$ percentil da distribuição F com p e $mn-m-p+1$ graus de liberdade e $(1-\alpha)$ denota o nível de significância do teste de hipóteses representado pela CCM. O limite de controle inferior da carta é zero em ambas as fases, por definição.

Para ilustrar a operacionalização das CCM, considere duas variáveis de interesse, medidas de um mesmo processo (Montgomery, 1996). Coletam-se amostras de tamanho $n = 10$. Os limites de controle foram calculados a partir de $m=20$ amostras obtidas do processo em condições desejáveis de operação. A estatística amostral utilizada para monitorar o desempenho das variáveis é:

$$T^2 = \left(\frac{n}{\bar{S}_1^2 \bar{S}_2^2 - \bar{S}_{12}^2} \right) \left[\bar{S}_2 (\bar{x}_1 - \bar{X}_1) + \bar{S}_1 (\bar{x}_2 - \bar{X}_2) - 2\bar{S}_{12} ((\bar{x}_1 - \bar{X}_1)(\bar{x}_2 - \bar{X}_2)) \right], \quad (2.28)$$

ou seja, a expressão na equação (2.23) adaptada para o caso especial onde $p = 2$. Os dados utilizados, as estimativas dos parâmetros populacionais μ e Σ e o valor de T^2 para cada amostra estão apresentadas no Quadro 2.1. A carta de controle resultante, com LC calculado pela equação (2.27), está apresentada na Figura 2.5 ($\alpha = 0.001$). Como esperado, os pontos na carta caracterizam um processo estável.

Tabela 2.1 Dados do exemplo de CCM

Amostra	Médias Amostrais		Variâncias e Covariâncias			Hotelling
	\bar{x}_{1k}	\bar{x}_{2k}	\bar{S}_{1k}^2	\bar{S}_{2k}^2	\bar{S}_{12k}	
1	115.25	1.04	1.25	0.87	0.80	2.16
2	115.91	1.06	1.26	0.85	0.81	2.14
3	115.05	1.09	1.30	0.90	0.82	6.77
4	116.21	1.05	1.02	0.85	0.81	8.29
5	115.90	1.07	1.16	0.73	0.80	1.89
6	115.55	1.06	1.01	0.80	0.76	0.03
7	114.98	1.05	1.25	0.78	0.75	7.54
8	115.25	1.10	1.40	0.83	0.80	3.01
8	116.15	1.09	1.19	0.87	0.83	5.92
10	115.92	1.05	1.17	0.86	0.95	2.41
11	115.75	0.99	1.45	0.79	0.78	1.13
12	114.90	1.06	1.24	0.82	0.81	9.96
13	116.01	1.05	1.26	0.55	0.72	3.86
14	115.83	1.07	1.17	0.76	0.75	1.11
15	115.29	1.11	1.23	0.89	0.82	2.56
16	115.63	1.04	1.24	0.91	0.83	0.70
17	115.47	1.03	1.20	0.95	0.70	0.19
18	115.58	1.05	1.18	0.83	0.79	0.00
19	115.72	1.06	1.31	0.89	0.76	0.35
20	115.40	1.04	1.29	0.85	0.68	0.62

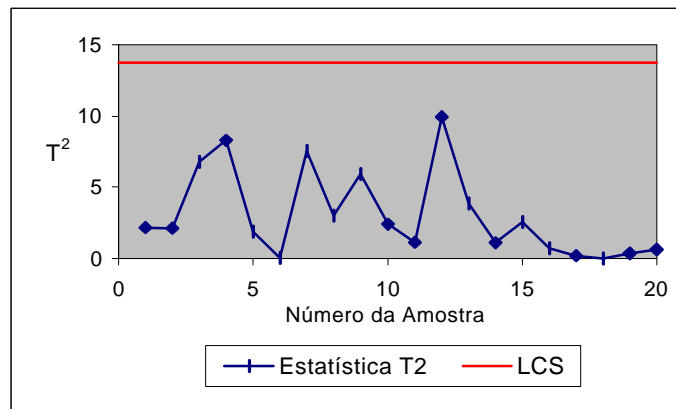


Figura 2.5– Exemplo de CCM de Hotelling.

Quando os parâmetros populacionais μ e Σ forem estimados a partir de um grande número de amostras preliminares, o $LCS = \chi^2_{\alpha,p}$ (aproximação pela distribuição *Qui-quadrado* com significância α e p graus de liberdade) em ambas as fases (Montgomery,1996).

Para a carta de médias, no caso univariado, utilizando amostras preliminares com $m \geq 20$ ou 25, a distinção entre os limites nas fases I e II é desnecessária. Entretanto, para as cartas de controle multivariadas a seleção dos LC deve ser cuidadosa. Lowry & Montgomery (1995) apresentaram tabelas indicando o número mínimo recomendado de amostras preliminares, m , para tamanhos de amostras de $n = 3, 5$ e 10 e para $p = 2, 3, 4, 5, 10$ e 20 variáveis, para que os LC da fase II sejam bem aproximados através do limite *Qui-quadrado*. Os valores recomendados para m foram sempre maiores que 50. A medida que o número de variáveis, p , aumenta, maior o valor m necessário para uma boa aproximação dos LC da fase II através do limite *Qui-quadrado*.

Em muitos processos, o monitoramento é realizado com amostras de tamanho unitário ($n=1$). Isto pode ocorrer em processos com baixa taxa de produção, onde o intervalo de tempo entre duas amostras consecutivas produzidas é grande, sendo inconveniente coletar amostras de tamanho maior que o unitário (Montgomery, 1996). Nas cartas construídas a partir de observações individuais, utilizando-se m amostras preliminares na fase I para o monitoramento de p variáveis, o LC exato durante a fase II é definido como ilustrado na equação (2.29); Ryan (1989).

$$LCS = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p}, \quad (2.29)$$

onde $F_{\alpha, p, m-p}$ significa o percentual da distribuição F com p e $m-p$ graus de liberdade e $(1-\alpha)$ denota o nível de significância do teste de hipóteses representado pela CCM. O limite de controle inferior é zero, por definição.

Quando utilizam-se grandes quantidades de amostras preliminares na fase I, ($m > 100$) o LC da fase II pode ser aproximado pela equação abaixo (Jackson, 1985):

$$LCS = \frac{p(m-1)}{m-p} F_{\alpha, p, m-p} \quad (2.30)$$

O LC calculado através da equação (2.30) ou aproximado através da distribuição χ^2 quando utilizam-se amostras de tamanho unitário, deve ser utilizado com cautela. A comparação do LC exato na fase II (dado pela equação 2.29), para amostras de $n=1$, com o LC aproximado pela distribuição *Qui-Quadrado*, para $p = 2, 3, 4, 5, 10$ e 20 variáveis foi apresentada em Lowry & Montgomery (1995). A medida que p (número de variáveis monitoradas) aumenta, maior é o valor de m (amostras preliminares utilizadas na fase I)

necessário para manter o erro relativo entre o LC exato e a aproximação c^2 . Para $p=3$, por exemplo, o número mínimo de m , para manter o erro relativo em 0,1 é 150.

Na fase I, quando são selecionadas m amostras preliminares de tamanho unitário representando o processo sob controle, o LC para a estatística T^2 utiliza a distribuição *Beta*, conforme a equação 2.31(Tracy *et al.*,1992).

$$LCS = \frac{(m-1)^2}{m} B_{a, \frac{p}{2}, \frac{(m-p-1)}{2}} \quad (2.31)$$

Cálculos para o LC, em ambas as fases de construção das cartas de controle multivariadas para observações individuais ($n=1$), foram apresentados por Sullivan & Woodall (1996). Os autores apresentaram diferentes estimadores da matriz de covariâncias Σ e analisaram sua utilização nas cartas de controle quanto a sensibilidade às mudanças no vetor de médias do processo. Dois tipos de alterações no vetor de médias do processo foram investigadas: mudanças bruscas (*step shift*) ou mudanças graduais (*ramp shift*). Observações adjacentes tendem a ser obtidas do mesmo vetor de médias do processo. Dessa forma, se houver uma mudança brusca 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 (Sullivan & Woodall, 1996).

Um estimador para Σ calculado a partir da diferença entre pares de observações sucessivas foi proposto por Holmes & Mergen (1993). Seja v_i (com $i = 1, \dots, m-1$) a diferença entre os pares de observações sucessivas, dado pela equação (2.32), e \mathbf{V} a matriz formada pelas diferenças v_i ,

$$v_i = x_{i+1} - x_i \quad (2.32)$$

Assim, o estimador proposto para Σ é dado pela equação (2.33).

$$\bar{\mathbf{S}} = \frac{1}{2} \frac{\mathbf{V}^t \mathbf{V}}{(m-1)} \quad (2.33)$$

$\bar{\mathbf{S}}$ é um estimador mais robusto para Σ do que o estimador $\bar{\mathbf{S}}_{qh}$ dado na equação (2.25). A suposição é de que observações sucessivas deveriam ter aproximadamente o mesmo vetor de médias. Dessa forma, a carta de controle T^2 , utilizando o estimador de Σ dado na

equação (2.33), é mais efetiva para detectar mudanças no vetor de médias do processo (Sullivan & Woodall, 1996).

A determinação do tamanho da amostra e da frequência de amostragem é um aspecto indispensável ao planejamento da CC multivariada. Esses parâmetros devem ser selecionados de modo a tornar a CC mais sensível à presença de causas especiais. Para tanto, agrupam-se os dados multivariados do processo em subgrupos racionais. Esses subgrupos devem ser selecionados, de acordo com as características do processo produtivo, de forma a maximizar as diferenças entre os subgrupos e minimizar as diferenças dentro dos subgrupos, sempre que causas especiais estiverem presentes (Sullivan & Woodal, 1996). Conforme as características do processo produtivo, existem diferentes estratégias para obtenção desses subgrupos racionais, para maiores detalhes, ver Montgomery (1996).

Em determinadas situações, de acordo com o processo produtivo, os dados são estruturados apenas como observações individuais. Não é possível, nestes casos, a seleção de subgrupos homogêneos de tamanho grande. Alguns métodos são recomendados por autores que tratam o caso de observações multivariadas individuais; ver, por exemplo, Jackson (1985), Tracy *et al.* (1992), Lowry & Montgomery (1995) e Wierda (1994).

A comparação entre as cartas multivariadas de Hotelling e as cartas controle de Causa-efeito (*Cause-Selecting Control Charts*) propostas por Zhang (1984) foi apresentada por Wade & Woodall (1993). As cartas de controle de Causa-efeito utilizam um modelo de regressão para estabelecer a relação entre as medidas de processo e de qualidade. Segundo os autores, estas cartas apresentam algumas vantagens com relação as cartas multivariadas de Hotelling. Uma das vantagens é que as cartas de Causa-efeito indicam mais facilmente quando o processo está fora de controle. As cartas de Hotelling indicam quando o processo está fora de controle, mas não indicam qual etapa específica do processo está fora de controle estatístico; isto pode ser visualizado na carta de Causa-efeito. Outra vantagem das cartas de Causa-efeito é que elas não requerem uma relação linear entre as medidas das variáveis do processo. O uso das cartas de Hotelling, em contrapartida, pressupõe uma distribuição multivariada normal, o que implica em uma relação linear entre medidas das variáveis do processo.

2.3.2 Carta de controle multivariada para soma acumulada (MCUSUM)

As cartas de controle do tipo *Shewhart* baseiam-se apenas nas informações mais recentes do processo, sendo insensíveis a pequenas e moderadas alterações no vetor de médias. As cartas para soma acumulada (cartas *CUSUM*) e as cartas para a média móvel ponderada exponencialmente (cartas *EWMA*) são alternativas em situações onde a detecção de pequenas mudanças nos parâmetros do processo é importante.

Para o caso univariado, a carta *CUSUM* captura diretamente a informação da seqüência dos dados amostrais através do somatório cumulativo dos desvios dos valores amostrais em relação ao valor alvo μ_0 (média do processo). Seja \bar{x}_j a média da $j^{\text{ésima}}$ amostra de tamanho $n \geq 1$; a estatística da carta *CUSUM* é dada pela equação (2.34).

$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0) \quad (2.34)$$

onde C_i representa a soma das diferenças entre a média de cada amostra e a média do processo até a $i^{\text{ésima}}$ amostra. Quando o processo permanece sob controle estatístico, C_i constitui-se em um passeio aleatório em torno do zero. Entretanto, quando a média amostral se altera para algum valor $\mu >$ (ou $<$) μ_0 , diferenças consecutivas positivas (ou negativas) serão acumuladas causando uma inflação positiva (ou negativa) no valor C_i .

A carta de controle *CUSUM* tabular bilateral considera os desvios acumulados das médias amostrais acima do valor alvo μ_0 , através da estatística C_i^+ , e abaixo do valor alvo μ_0 , através da estatística C_i^- , dadas nas equações (2.35) e (2.36). Se os valores C_i^+ e C_i^- excederem o intervalo de decisão H , o processo é considerado 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 $LC = H$ e $LCI = -H$.

$$C_i^+ = \max\{0, x_i - (\mu_0 + K) + C_{i-1}^+\} \quad (2.35)$$

$$C_i^- = \max\{0, (\mu_0 - K) - x_i + C_{i-1}^-\} \quad (2.36)$$

Nas equações (2.35) e (2.36), K é chamado de valor de referência (ou tolerância), sendo calculado como a metade da diferença entre a média do processo fora de controle que se quer detectar, μ e o valor alvo da média do processo, μ_0 . Se a alteração na média do processo for expressa em unidades de desvio padrão como $\mu = \mu_0 + ds$, onde s é o desvio

padrão e d é dado na equação (2.37), K pode ser calculado como a metade dessa alteração, de acordo com a equação (2.38) a seguir (Montgomery, 1996).

$$d = \frac{|\bar{m}_1 - m_0|}{s} \quad (2.37)$$

$$K = \frac{d}{2}s = \frac{|\bar{m}_1 - m_0|}{2} \quad (2.38)$$

A escolha dos parâmetros H e K para a carta *CUSUM* tabular baseia-se no tamanho médio de corrida (ARL, ver equação 2.22). Define-se $H = hs$ e $K = ks$ (normalmente, $k = \delta/2$), sendo s o desvio padrão da variável usada na construção da carta *CUSUM*.

A escolha de k fundamenta-se no tamanho da alteração de média do processo que se deseja detectar. ARL_0 e ARL são os parâmetros usados para medir o desempenho de uma carta *CUSUM*, a partir de uma determinada escolha de h e k . Uma vez escolhido o valor de k , para um dado ARL_0 , devemos escolher h de forma a maximizar a chance de detectar um deslocamento de tamanho δ na média μ_0 do processo. Uma vez dada a probabilidade de alarme falso que estamos dispostos a tolerar e o tamanho do deslocamento da média do processo que estamos interessados em detectar, escolhe-se o valor de h que minimize o número médio de amostras sucessivas necessárias (ARL) para detectar o deslocamento na média do processo. Utilizando $h = 4$ ou $h = 5$ e $k = \delta/2$ geralmente obtém-se uma carta *CUSUM* com boas propriedades ARL para uma alteração em torno de $1s$ na média do processo (Montgomery, 1996).

Os valores para os ARLs em cartas de controle *CUSUM* são obtidos utilizando uma cadeia de Markov na obtenção das probabilidades de alteração de um estado sob controle para um estado fora de controle no processo (Brook & Evans *apud* Montgomery, 1996). O cálculo aproximado dos ARLs, dados h e k , para uma carta *CUSUM* tabular unilateral, C_i^+ ou C_i^- , é dado pela equação (2.39) (Montgomery, 1996).

$$ARL = \frac{[\exp(-2\Delta b) + 2\Delta b - 1]}{2\Delta^2}, \text{ se } D \neq 0, \\ = b^2, \text{ se } D = 0, \quad (2.39)$$

Na equação (2.39), $D = \delta - k$ para o caso da soma acumulada positiva C_i^+ e $D = -\delta - k$ para o caso da soma acumulada negativa e C_i^- , $b = h + 1$ para ambos os casos. Quando $\delta =$

0, a expressão (2.39) resulta em ARL_0 . Quando $\delta \neq 0$, obtém-se o valor de ARL referente ao desvio de tamanho δ em relação a μ_0 que deseja-se detectar.

Para a carta *CUSUM* tabular bilateral, o valor de ARL é calculado através da equação (2.40).

$$\frac{1}{ARL} = \frac{1}{ARL^+} + \frac{1}{ARL^-} \quad (2.40)$$

onde ARL^+ refere-se à soma acumulada positiva (C_i^+) e ARL^- refere-se à soma acumulada negativa (C_i^-), obtidos respectivamente através das equações (2.35) e (2.36).

Quando $k = 1/2$, $h = 5$ e $D = 0$ obtém-se $ARL_0^+ = ARL_0^- = 938,2$, para a carta unilateral, utilizando a equação (2.39). Com os mesmos valores de k , h e D na equação (2.40), obtém-se $ARL_0 \approx 469$ para a carta bilateral. Adotando $k = 1/2$ e $h = 5$, obtém-se, nas equações (2.35) e (2.36), respectivamente, $D = 0,5$ em relação a C_i^+ e $D = -1,5$ em relação a C_i^- . Usando as equações (2.39) e (2.40), $ARL \approx 10$. Este resultado mostra que a carta *CUSUM* é mais sensível para detectar mudanças na média da ordem de 1S se comparada a carta de Shewhart, cujo ARL é de aproximadamente 44 amostras.

As cartas multivariadas podem ser separadas em duas categorias distintas: os esquemas de monitoramento *direcionalmente invariantes* e os esquemas de monitoramento *direcionalmente específicos* (Pignatiello & Runger, 1990). As principais cartas em cada categoria são apresentadas a seguir.

No controle multivariado, o tamanho do desvio de interesse é representado através da distância de *Mahalanobis*, dada na equação (2.23). A distância de *Mahalanobis* representa a distância, em p dimensões, entre o vetor de médias μ do processo fora de controle e o vetor de médias μ_0 do processo sob controle (é a distância, em termos do número de desvios padrões entre μ e μ_0). Este procedimento não captura a direção específica de deslocamento do vetor μ_0 . As cartas de controle multivariadas com desempenho ARL dado como função da distância de *Mahalanobis* são denominadas cartas de controle com *direcionalidade invariante* (Pignatiello & Runger, 1990). A maioria das cartas univariadas bilaterais de controle, entre elas as cartas para médias de *Shewhart* e a *CUSUM* tabular, possuem direcionalidade invariante.

No entanto, múltiplas cartas univariadas podem ser utilizadas no monitoramento de processos multivariados. Nestes casos, tais esquemas de monitoramento devem apontar os

desvios da média do processo em alguma direção particular, como por exemplo, ao longo dos eixos das CQs monitoradas ou na direção dos componentes principais resultantes delas. Dessa forma, quando estamos exercendo o controle multivariado, num processo com p variáveis de interesse, utilizando p cartas univariadas, não estamos utilizando um procedimento de controle com direcionalidade invariante, pois o desempenho ARL depende da direção específica do deslocamento do vetor μ_0 (Pignatiello & Runger, 1990). Estes procedimentos de controle são chamados *de direcionalidade específica*.

No controle estatístico multivariado, os procedimentos baseados na filosofia *CUSUM* classificam-se em dois grupos distintos: (i) procedimentos de controle que utilizam múltiplas cartas *CUSUM* univariadas (denominado múltiplo *CUSUM* univariado, abreviados por *MCU*), desconsiderando assim a correlação entre as variáveis e (ii) procedimentos de controle que utilizam uma carta *CUSUM* multivariada (abreviados por *MCUSUM*), isto é, utilizam a matriz Σ de covariâncias das variáveis.

Através do procedimento de controle multivariado *MCU* e direcionalmente específico, tem-se o monitoramento de cada uma das p características de qualidade individualmente, através de cartas *CUSUM*, proposto por Woodall & Ncube (1985). A matriz de covariâncias Σ 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 nas equações (2.35) e (2.36), com os limites de controle dados por H . Assim, a média da $j^{\text{ésima}}$ variável é monitorada obtendo-se os escores C_j^+ e C_j^- a cada nova amostra da variável j (com $j=1, \dots, p$). A $j^{\text{ésima}}$ carta *CUSUM* sinaliza que a média da $j^{\text{ésima}}$ variável sofreu um deslocamento quando $C_j^+ > H_j$ ou $C_j^- < -H_j$ para valores específicos de K_j e H_j (parâmetros para o monitoramento da variável j). O *MCU* indicará que o processo está fora de controle quando pelo menos uma das p cartas univariadas detectar algum desvio em relação a média da respectiva variável.

As cartas *MCU* devem ser usadas num processo onde existe o interesse em detectar desvios numa direção específica em relação ao vetor μ_0 (Pignatiello & Runger, 1990), já que possuem direcionalidade específica. Estes desvios podem ser caracterizados através de deslocamentos das variáveis do processo sobre seus eixos (desvios específicos em relação as médias das variáveis) ou na direção dos eixos dos *componentes principais* (variáveis independentes que são combinações lineares das variáveis originais; ver seção 2.1 deste capítulo).

Quando há interesse em detectar um desvio da média em uma determinada direção, uma carta *CUSUM* univariada estruturada nesta direção fornecerá os melhores resultados em termos de desempenho ARL (Healy, 1987). Entretanto, se existir interesse em detectar deslocamentos em várias direções em relação a $\boldsymbol{\mu}_0$, cartas *CUSUM* univariadas podem ser pouco sensíveis (Pignatiello & Runger, 1990). Alguns procedimentos de controle direcionalmente específicos são sugeridos na literatura (ver Pignatiello & Runger, 1990, Crosier, 1988 e Lowry & Montgomery, 1995).

O procedimento *MCUSUM* foi inicialmente proposto por Crosier (1988), consistindo de duas cartas de controle multivariadas *CUSUM*. A carta com melhor desempenho ARL, denominada *CCUSUM*, dada na equação (2.41), se constitui em uma extensão multivariada da carta univariada *CUSUM*.

$$\mathbf{c}_l = \left[(\mathbf{s}_{l-1} + \mathbf{x}_l - \hat{\boldsymbol{\mu}}_0) \boldsymbol{\Sigma}^{-1} (\mathbf{s}_{l-1} + \mathbf{x}_l - \hat{\boldsymbol{\mu}}_0) \right]^{1/2} \quad (2.41)$$

Na equação (2.41), s_{l-1} representa o vetor das diferenças consecutivas superiores a K (valor de referência, dado na equação (2.38)), acumuladas até a amostra $l-1$, \mathbf{x}_l representa o vetor de observações referentes à amostra l e \mathbf{c}_l é a distância, em P dimensões, entre as diferenças sucessivas acumuladas superiores a K , e o vetor das médias $\boldsymbol{\mu}_0$ do processo sob controle. O vetor \mathbf{s}_l é dado na equação (2.42), onde $\mathbf{s}_0 = \mathbf{0}$ e $K > 0$ (Crosier, 1988).

$$\begin{aligned} s_l &= \mathbf{0} \quad \text{se } \mathbf{c}_l \leq K \\ s_l &= (\mathbf{s}_{l-1} + \mathbf{x}_l - \boldsymbol{\mu}_0) / (1 - K / \mathbf{c}_l) \quad \text{se } \mathbf{c}_l > K, \end{aligned} \quad (2.42)$$

A estatística de controle a ser plotada na carta *CCUSUM* é dada na equação (2.43).

$$\tilde{\mathbf{a}}_l = \left\{ \mathbf{s}_l^t \boldsymbol{\Sigma}^{-1} \mathbf{s}_l \right\}^{1/2} \quad (2.43)$$

O processo será considerado fora de controle sempre que $\gamma_l > H$, para k e h fixos. Sugere-se como valor de k , a exemplo do caso univariado, a metade da distância padronizada entre $\boldsymbol{\mu}$ e $\boldsymbol{\mu}_0$ ($\boldsymbol{\mu}$ representa o vetor de médias do processo fora de controle); ou seja, $k = \delta / 2$.

No monitoramento de p variáveis simultaneamente, se o vetor de médias do processo deslocar-se para $\boldsymbol{\mu}$, a carta *CCUSUM* fornece indicações em relação à possível direção deste deslocamento. Além disso, este esquema fornece uma detecção mais rápida de pequenas alterações no vetor de médias que o procedimento multivariado de Hotelling (1947). Para um dado ARL_0 , as cartas *MCUSUM* apresentam um desempenho ARL superior em relação à carta de Hotelling (ver Woodal & Ncube, 1985 e Crosier, 1988).

Dois outros procedimentos *MCUSUM*, foram propostos por Pignatiello & Runger (1990): a carta *MC1* e a carta *MC2*. A diferença entre estes dois procedimentos está centrada na forma como a acumulação (somatório) é realizada. No procedimento *MC1*, realiza-se inicialmente o somatório dos vetores de médias e após isso calcula-se o quadrado desse somatório. No procedimento *MC2*, cada vetor de médias é elevado ao quadrado e depois o somatório é realizado.

O procedimento com melhor desempenho ARL, *MC1*, é baseado no vetor de somas \mathbf{c}_t , dado na equação (2.44).

$$\mathbf{c}_t = \sum_{l=t-n_t+1}^t (\mathbf{x}_l - \hat{\mathbf{i}}_0) \quad (2.44)$$

$$n_t = n_{t-1} + 1 \text{ se } MC1_{t-1} > 0,$$

$$\text{ou } n_t = 1, \text{ caso contrário} \quad (2.45)$$

Na equação (2.45) n_t representa o número de subgrupos na carta *CUSUM*. O vetor $\frac{1}{n_t} \mathbf{c}_t$, dado na equação (2.46), representa a diferença entre média amostral acumulada e o valor alvo da média do processo. No tempo t a média do processo multivariado pode ser estimada na equação (2.47).

$$\frac{1}{n_t} \mathbf{c}_t = \left(\frac{1}{n_t} \sum_{l=t-n_t+1}^t \mathbf{x}_l \right) - \hat{\mathbf{i}}_0 \quad (2.46)$$

$$\frac{1}{n_t} \mathbf{c}_t + \mathbf{m}_0 \quad (2.47)$$

A equação (2.48) apresenta o vetor \mathbf{c}_t normalizado e a equação (2.49) apresenta a estatística *MC1* da carta multivariada.

$$\|\mathbf{c}_t\| = \sqrt{(\mathbf{c}_t^t \Sigma^{-1} \mathbf{c}_t)} \quad (2.48)$$

$$MC1 = \max\{\|\mathbf{c}_t\| - kn_t, 0\} \quad (2.49)$$

Na equação (2.49) $n_t = n_{t-1} + 1$ se $MC_{t-1} > 0$ e 1, caso contrário, e $K > 0$. O processo será considerado fora de controle se $MC_t > H$, para K e H fixos.

O procedimento *MC2* considera o quadrado da distância de cada vetor de média amostral do vetor de média \mathbf{m}_b , d_t^2 , dada na equação (2.50), e então realiza o somatório destas distâncias quadradas.

$$d_t^2 = (x_t - \mathbf{m}_b)' \Sigma^{-1} (x_t - \mathbf{m}_b) \quad (2.50)$$

A carta *MC2* é dada na equação (2.51), onde K é o valor de referência (ou tolerância).

$$MC2_t = \max\{0, MC2_{t-1} + d_t^2 - K\} \quad (2.51)$$

O desempenho ARL das cartas *MC1*, *MC2*, carta c^2 e cartas múltiplo *CUSUM* univariado (*MCU*), propostas por Woodall & Ncube (1985), foi estudado por Pignatiello & Runger (1990). Foi analisado o desempenho das cartas para dois casos de alterações na média do processo: o primeiro caso do tipo $\boldsymbol{\mu} = (d, 0, \dots, 0)^t$ e o segundo caso do tipo $\boldsymbol{\mu} = (d, d, \dots, d)^t$, para $p = 2, 3$ e 10 variáveis. As alterações da média foram investigadas para as distâncias de $0,5; 1,0; 1,5; 2,0; 2,5$ e $3,0$ do valor alvo $\boldsymbol{\mu}_0 = (0, \dots, 0)^t$ para ambos os casos, para os três diferentes valores de p utilizados. As cartas que apresentaram os melhores valores de ARL para detectar pequenas mudanças no vetor de médias do processo normal multivariado considerado foram as cartas *MC1* e *MCU*.

As cartas *CCUSUM* e *MC1* apresentadas acima utilizam a matriz de covariâncias Σ do processo em suas estatísticas de controle. Desta forma, estes procedimentos apresentam *direcionalidade invariante*, ou seja, o desempenho ARL dessas cartas é determinado apenas em função da distância entre o vetor $\boldsymbol{\mu}_0$ do processo sob controle e o vetor $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ do processo fora de controle, e não em função de um deslocamento específico na média de uma determinada variável do processo.

2.3.3 Carta de controle multivariada para média móvel exponencialmente ponderada (*MEWMA*)

Duas aplicações para carta *EWMA* são identificadas na literatura (MacGregor & Harris, 1990). A primeira aplicação consiste da sua utilização no controle estatístico de processo, como uma ferramenta para detectar quando uma causa especial atua no sistema. A segunda aplicação diz respeito a utilização da estatística *EWMA* na previsão de observações futuras a partir do processo de média móvel integrada de primeira ordem. Nesta segunda aplicação, a estatística pode ser usada como parte de um algoritmo para ajustar o processo, desta forma reduzindo o erro quadrado médio em torno do valor alvo.

Assim como a carta de controle *CUSUM*, a carta *EWMA* pode ser utilizada como alternativa à carta de controle de Shewhart quando há interesse em detectar pequenas alterações na média do processo. A carta de controle para média móvel exponencialmente ponderada foi introduzida por Roberts (1959). Lucas & Saccucci (1990) apresentaram uma boa discussão sobre sua utilização.

A carta de controle *EWMA* para o caso univariado ($p=1$) e para $n = 1$ é dada pela estatística da equação (2.52) (Lowry *et al.*, 1992).

$$z_l = rx_l + (1-r)z_{l-1} \quad (2.52)$$

Na equação (2.52) x_l representa a observação referente à $l^{\text{ésima}}$ observação amostral unitária, r é a constante de ponderação ($r \in (0,1]$) e z_{l-1} é o escore obtido na amostra $l-1$. Considerando que $z_0 = \bar{m}_0$, onde \bar{m}_0 é a média do processo, e que as suposições de normalidade das observações x_l ($x_l \sim N(\bar{m}_0, S)$) e de independência entre amostras são satisfeitas, o desvio padrão de z_l é dado na equação (2.53) (Lowry *et al.*, 1992).

$$\sigma_{z_l} = \sqrt{\sigma^2 \left(\frac{r}{2-r}\right) [1-(1-r)]^{2l}} \quad (2.53)$$

Os limites de controle para a carta *EWMA*, quando $\bar{m}_0 = 0$, são dados na equação (2.54), onde B representa o número de desvios padrões tolerados em relação a z_l .

$$\begin{aligned} \text{L.S.C.} &= +BS_{z_l} \\ \text{L.C.} &= 0 \\ \text{L.I.C.} &= -BS_{z_l} \end{aligned} \quad (2.54)$$

Os critérios para a escolha dos parâmetros r e B para a carta *EWMA* univariada foram discutidos detalhadamente por Lucas & Saccucci (1990). A escolha dos limites de controle baseia-se na forma assintótica do desvio padrão de z_i , dada na equação (2.55).

$$S_{z_l} = \sqrt{\frac{r}{2-r}} S \quad (2.55)$$

O critério para a escolha dos valores de r e B é semelhante ao critério de escolha de k e h para a carta *CUSUM* tabular. 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 B que

minimizam ARL (número médio de amostras consecutivas necessárias para que a carta detecte o desvio especificado). Para a detecção de pequenos desvios na média do processo, $B = 3$ e pequenos valores de r devem ser utilizados (Montgomery, 1996).

Uma versão multivariada para carta de controle *EWMA* (abreviada como *MEWMA*) foi proposta em Lowry *et al.* (1992). Para o caso multivariado ($p > 1$), a estatística da equação (2.52) é expandida conforme a equação (2.56) (Lowry *et al.*, 1992).

$$z_l = \mathbf{R}\mathbf{x}_l + (\mathbf{I} - \mathbf{R})\mathbf{z}_{l-1} \quad (2.56)$$

Na equação (2.56) \mathbf{x}_l é o vetor de observações amostrais p -dimensional referente a $l^{\text{ésima}}$ amostra unitária ($n = 1$), \mathbf{R} é a matriz diagonal (r_1, r_2, \dots, r_p) , onde r é a constante de ponderação $\{r_p \in (0,1], \}$ e \mathbf{z}_{l-1} é o vetor p -dimensional dos escores referentes a amostra $l-1$, supões-se $\mathbf{z}_0 = 0$. Quando não houver razão para atribuir diferentes pesos para as variáveis monitoradas na carta, então $r_1 = r_2 = \dots = r_p = r$. Neste caso, monitora-se um processo multivariado utilizando o procedimento *MEWMA* através da estatística de controle dada na equação (2.57).

$$T_l^2 = \mathbf{z}_l^t \Sigma_{\mathbf{z}_l}^{-1} \mathbf{z}_l \quad (2.57)$$

A matriz de covariâncias $\Sigma_{\mathbf{z}_l}$ é dada na equação (2.58) (ver Lowry *et al.*, 1992).

$$\Sigma_{\mathbf{z}_l} = \left\{ \frac{r[1 - (1-r)^{2l}]}{(2-r)} \right\} \Sigma \quad (2.58)$$

O processo será considerado em controle se $T_l^2 < h$ na equação (2.57). O valor de h é escolhido a partir do desempenho ARL desejado para a carta. Quando as p características de qualidade recebem o mesmo peso r , o desempenho ARL da carta *MEWMA* depende apenas do parâmetro de não-centralidade, dado pela distância de *Mahalanobis*. Dessa forma, este procedimento de controle também apresenta a característica de *direcionalidade invariante*, podendo-se comparar o desempenho dessa carta em relação às cartas de Hotelling, *CCUSUM* e *MC1*. A determinação do comprimento médio de corrida sob controle de uma carta de controle *MEWMA* através da utilização de uma equação integral e tabelas com valores apropriados para o LC, em função do número de características de qualidade, p foram apresentados em Rigdon *apud* Lowry & Montgomery (1995).

A carta *MEWMA* possui um desempenho ARL superior em relação as demais cartas (Hotelling, *CCUSUM* e *MC1*), quando implantada em um processo inicialmente fora de controle ($\mu \neq \mu_0$). Neste caso, a carta *MEWMA* apresenta uma rápida resposta, o que é uma característica inerente a sua estatística de controle, analogamente a estatística da carta univariada *EWMA*. Da mesma forma, quando o processo está inicialmente sobre controle, a carta *MEWMA* é pelo menos tão boa quanto as demais cartas multivariadas na detecção de deslocamentos no vetor de médias μ_0 (Lowry *et al.*, 1992).

Em contrapartida, as cartas *EWMA* univariadas reagem lentamente a mudanças na média, se comparadas a outras cartas de controle univariadas (Yashchin, 1987). Podendo resultar em uma defasagem no tempo de detecção da alteração na média. O atraso na detecção de mudanças na média pode ocorrer com ambas as cartas multivariadas, *MEWMA* e *MCUSUM*, isto ocorre devido ao seu elevado grau de inércia. Para contornar este problema, de maneira análoga ao caso univariado, recomenda-se o uso da carta de Hotelling em conjunto com essas cartas, estabelecendo-se um compromisso entre baixa inércia (carta de Hotelling) e rápida detecção de pequenos deslocamentos no vetor de médias (*MCUSUM* e *MEWMA*) (Lowry *et al.*, 1992).

2.3.4 Carta de Controle para Monitoramento da Variabilidade de Processos

O monitoramento de processos com relação a sua variabilidade é realizado com cartas de controle específicas para monitorar a sua matriz de covariâncias. A matriz de covariâncias ($p \times p$) descreve a variabilidade de um processo multivariado de ordem p . As variâncias das p variáveis envolvidas situam-se na diagonal principal e as covariâncias entre as variáveis são dadas pelos elementos fora da diagonal principal da matriz.

Existem três procedimentos mais utilizados no monitoramento da variabilidade em processos multivariados. O primeiro procedimento consiste em a partir de sucessivas amostras, testar se a estrutura de covariâncias do processo está bem representada por uma matriz de covariâncias Σ_0 (Alt, 1984). Este procedimento baseia-se na carta de controle univariada ($p = 1$) para variabilidade, a carta de controle S^2 (ver Montgomery, 1996). A equação (2.59) apresenta a estatística de controle, calculada para a $i^{\text{ésima}}$ amostra. O limite de controle superior desta estatística é dado na equação (2.60), o limite inferior é zero por definição.

$$\mathbf{w}_i = -pn + pn \ln(n) - n \ln\left(\frac{|\mathbf{A}_i|}{|\Sigma_0|}\right) + \text{tr}(\Sigma_0^{-1} \mathbf{A}_i) \quad (2.59)$$

$$LCS = c^2_{a, p(\frac{p+1}{2})} \quad (2.60)$$

Na equação (2.59) $\mathbf{A}_i = (n-1) \mathbf{S}_i$, onde \mathbf{S}_i é a matriz de covariâncias amostrais obtidas a partir da amostra i , $|\mathbf{A}_i|$ é o determinante de \mathbf{A}_i e tr representa o traço da matriz resultante do produto $\hat{\mathbf{a}}_0^{-1} \mathbf{A}_i$.

O LCS, na equação (2.60), é dado pela distribuição Qui-quadrado com $p(p+1)/2$ graus de liberdade. Quando o valor calculado de \mathbf{w}_i na equação (2.59) for superior ao LCS, existirão indícios de que a variabilidade do processo pode ter se alterado, significando uma situação de processo fora de controle ($\Sigma \neq \Sigma_0$).

O segundo procedimento para monitoramento da variabilidade é baseado em uma medida denominada *variância generalizada* (Alt, 1995). A *variância generalizada* de um processo, estimada através de uma amostra i , é representada por $|\mathbf{S}_i|$, ou seja, pelo determinante da matriz de covariâncias amostrais (ver Montgomery & Wadsworth, 1972).

Além destes dois procedimentos apresentados acima, existe um terceiro procedimento, também baseado na variância generalizada, apresentado por Montgomery (1996). Ele consiste na construção de uma carta de controle supondo-se que grande parte da distribuição de probabilidade de $|\mathbf{S}|$ está contida no intervalo $E(|\mathbf{S}|) \pm 3\sqrt{V(|\mathbf{S}|)}$. O valor $E(|\mathbf{S}|)$, que representa a média (valor esperado) de $|\mathbf{S}|$ e $\sqrt{V(|\mathbf{S}|)}$ que representa o desvio padrão de $|\mathbf{S}|$, são dados nas equações (2.61) e (2.62), respectivamente; os parâmetros b_1 e b_2 , naquelas equações são definidos nas equações (2.63) e (2.64).

$$E(|\mathbf{S}|) = b_1 |\Sigma| \quad (2.61)$$

$$V(|\mathbf{S}|) = b_2 |\Sigma|^2 \quad (2.62)$$

$$b_1 = \frac{1}{(n-1)^p} \prod_{i=1}^p (n-i) \quad (2.63)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{i=1}^p (n-1) \left[\prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j) \right] \quad (2.64)$$

Através da análise de amostras preliminares, a matriz Σ pode ser estimada através da matriz de covariâncias amostrais S . Alguns estimadores de Σ foram apresentados na seção 2.3.1 deste capítulo. Montgomery (1996) recomenda a utilização de $|S|/b_1$ como estimador não-tendencioso de $|\Sigma|$.

Os limites de controle para a carta de controle proposta por Montgomery (1996) são dados na equação (2.65).

$$\begin{aligned} \text{LCS} &= \frac{|S|}{b_1}(b_1 + 3b_2^{1/2}) \\ \text{LCT} &= |S| \\ \text{LCI} &= \frac{|S|}{b_1}(b_1 - 3b_2^{1/2}) \end{aligned} \quad (2.65)$$

A utilização da *variância generalizada* pode resultar em uma caracterização insuficiente da estrutura de covariância do processo. Isso deve-se ao fato de que ela considera o determinante da matriz de covariâncias amostrais, $|S|$, não capturando a estrutura de correlação entre as variáveis. Embora diferentes matrizes possam resultar em um mesmo valor de variância generalizada, isso não significa que todas as variáveis tenham a mesma estrutura de correlação nas matrizes, podendo ser correlacionadas positivamente em uma e negativamente em outra. Como alternativa para solucionar este problema, pode-se utilizar utilização de p cartas de controle univariadas para monitorar a variabilidade do processo, simultaneamente ao uso da carta de controle para $|S|$; Montgomery (1996).

2.3.5 Controle Estatístico de Processos para Dados de Processo Auto-Correlacionados.

Para que seja possível a utilização das cartas de controle, apresentadas nas seções anteriores deste capítulo, duas suposições devem ser validadas, partindo-se da situação de processo em controle estatístico. A primeira suposição é a de independência estatística dos dados gerados no processo. A segunda suposição é a de que os dados de processo seguem uma distribuição normal com média μ e desvio padrão σ . Quando ambas as suposições são satisfeitas, para a situação de controle estatístico, os dados de processo ($x_t, t=1,2,\dots$) seguem o modelo de *Shewhart*, dado na equação (2.66).

$$x_t = \bar{\mu} + \hat{\sigma} a_t \quad (2.66)$$

Na equação (2.66), μ é a média do processo e e_t é uma seqüência de variáveis independentes e aleatoriamente distribuídas.

Nas situações onde a suposição de normalidade é violada em graus fraco ou moderado, as cartas de controle vistas anteriormente ainda oferecem um desempenho razoável. Por outro lado, quando a suposição de independência entre as observações do processo não é satisfeita, as cartas apresentadas anteriormente não oferecem um desempenho satisfatório.

A presença de autocorrelação entre as observações tem profundos efeitos nas cartas de controle desenvolvidas usando a suposição de independência entre as observações. Um destes efeitos é o aumento da frequência com que sinais falsos são gerados. Mesmo correlações fracas produzem distúrbios nas cartas de controle levando a conclusões erradas sobre o estado de controle do processo. Na presença de um sinal fora de controle, torna-se difícil distinguir se o mesmo é devido a atuação de uma causa especial ou se é um alarme falso induzido pela estrutura de autocorrelação dos dados de processo.

A suposição de independência das observações não é satisfeita em alguns processos industriais. Em processos químicos, monitoramentos automatizados ou procedimentos de inspeção *on line*, onde medidas consecutivas de características de produto ou processo são tomadas em intervalos curtos de tempo, os dados podem resultar fortemente correlacionados.

A abordagem de diversos autores para trabalhar com dados autocorrelacionados baseia-se no ajuste de um modelo apropriado de séries temporais às observações e, então, a aplicação de cartas de controle de Shewhart sobre os resíduos do modelo ajustado (Montgomery & Mastrangelo, 1991). A classe de modelos de séries temporais mais utilizada é a média móvel integrada autoregressiva (*ARIMA – Autoregressive Integrated Moving Average*) (ver Box *et al.*, 1994). Um modelo autoregressivo de primeira ordem, pertencente a esta classe, é dado para a observação x_t da característica de processo X no instante t , na equação (2.67) (Montgomery, 1996). Outros modelos de ordem mais elevada pertencentes a esta classe são apresentados em Box *et al.* (1994).

$$x_t = \mu + f x_{t-1} + e_t \quad (2.67)$$

Na equação (2.67), μ e f ($-1 < f < 1$) são constantes desconhecidas do modelo, e e_t é normalmente e independentemente distribuído, com média zero e desvio padrão S .

As observações x_t dadas na equação (2.67) têm média e desvio padrão dados nas equações (2.68) e (2.69), respectivamente.

$$\bar{x} = \frac{X}{1-f} \quad (2.68)$$

$$s = \frac{S}{(1-f)^{1/2}} \quad (2.69)$$

Os resíduos deste modelo, e_t , são dados na equação (2.70), sendo distribuídos normalmente e independentes, com média zero e variância constante.

$$e_t = x_t - \hat{x}_t \quad (2.70)$$

Cartas de controle convencionais podem ser aplicadas no monitoramento dos resíduos na equação (2.70). Pontos fora de controle ou padrões incomuns na carta podem indicar que o parâmetro f se alterou, implicando que x_t está fora de controle (Montgomery, 1996).

O ajuste de modelos de séries temporais para dados multivariados também é possível, apesar de não muito favorecido em aplicações práticas (Mason *et al.*, 1997). A utilização das cartas de controle *EMWA* em certas situações onde os dados de processo são autocorrelacionados é recomendada em Montgomery & Mastrangelo (1991).

A estrutura de autocorrelação das variáveis de processo pode ser identificada utilizando-se técnicas estatísticas de projeção multivariada de dados (Montgomery *et al.*, 1993). Uma das técnicas utilizadas é a análise de componentes principais (ACP; ver seção 2.1.1) 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 das variáveis de processo. Quando dispõem-se de dados do processo, X , e dados de qualidade, Y , recomenda-se o uso do procedimento mínimos quadrados parciais (MQP; ver seção 2.1.2), para capturar a estrutura de autocorrelação dos dados, em lugar da ACP (Kresta *et al.*, 1991). Para identificar um sinal fora de controle obtido via ACP ou MQP, pode-se utilizar cartas de controle univariadas (uma para cada variável de processo envolvida).

2.4. Carta de Controle Multivariada do tipo Hotelling T^2 baseada em Métodos de Projeção Multivariada de Dados.

As cartas de controle multivariadas tradicionais apresentadas anteriormente são muitas vezes insuficientes e inadequadas no monitoramento de processos industriais devido aos diversos fatores apresentados na seção 2.1 desta dissertação.

Uma alternativa para o controle estatístico destes processos é a utilização das CCM de Hotelling T^2 baseadas em métodos de projeção multivariada de dados. Estas cartas utilizam os métodos de projeção multivariada de dados apresentados na seção 2.1, em conjunto com a CCM de Hotelling, apresentada na seção 2.3.1.

Os métodos de projeção multivariada de dados (ACP ou MQP, ver seções 2.1.1 e 2.1.2) são utilizados como técnica para a redução da dimensionalidade dos dados históricos obtidos do processo em situação de controle estatístico. Os principais objetivos da utilização destas técnicas são: (i) capturar a estrutura de correlação dos dados de processo, (ii) representar os dados satisfatoriamente através um número menor de novas variáveis independentes e (iii) fornecer um modelo de referência para o processo em situação de controle estatístico e quando produziu produto de boa qualidade.

O primeiro passo desta estratégia de controle de processos é a construção do modelo de referência dos dados de processo, contra o qual o monitoramento futuro será comparado. Dessa forma, o modelo de referência é uma representação dos dados originais do processo em um espaço de projeção dimensionalmente mais baixo de novas variáveis independentes.

A carta de controle multivariada de Hotelling T^2 é construída utilizando como dados de entrada as variáveis definidas pelo modelo de referência resultante do método de projeção utilizado. O controle estatístico é realizado monitorando-se em conjunto com a carta de controle multivariada de Hotelling T^2 os gráficos dos escores e dos resíduos das novas variáveis resultantes do modelo de projeção multivariada. Utilizando então, o modelo de referência de ACP ou MQP construído e a variação dos escores e dos resíduos calculados a partir dele, definem-se os espaços de projeções dos escores e dos resíduos das novas variáveis independentes (ver seção 2.1.1 e 2.1.2). O monitoramento do desempenho futuro do processo é realizado plotando-se a projeção dos dados no espaço de escores e na carta de controle dos resíduos.

O foco desta dissertação concentra-se no desenvolvimento de CCM do tipo Hotelling T^2 baseada em componentes principais (doravante denominada CCP) apresentada na seção 2.4.1. deste capítulo. A elaboração da carta de controle multivariada baseada em componentes principais (CCP) fundamenta-se na utilização da Análise de Componentes Principais (seção 2.1.1) sobre os dados de processo e na aplicação da CCM de Hotelling T^2 (seção 2.3.1).

As CCP foram inicialmente desenvolvidas para monitoramento de processos contínuos, onde diversas variáveis de processo/produto são monitoradas *on-line*; ver Kresta *et al.* (1991), Skagerberg *et al.* (1992), Miller *et al.* (1993) e MacGregor *et al.* (1994). Mais

recentemente, a aplicação das CCP no monitoramento de processos em batelada foi sugerida por Nomikos & MacGregor (1994, 1995). O desenvolvimento aqui apresentado aplica-se a processos em batelada, estando baseado nos trabalhos de MacGregor (1995), Nomikos & MacGregor, (1994, 1995) e Kourti & MacGregor (1996).

Embora os procedimentos de controle multivariado de processos utilizando métodos de projeção multivariadas de dados sejam desenvolvimentos teóricos recentes, alguns desenvolvimentos e aplicações são encontrados na literatura. Na exposição a seguir, serão apresentados brevemente os principais estudos encontrados sobre o tema.

A transformação das variáveis originais de interesse em novas variáveis ortogonais, utilizando Análise de Componentes Principais (ACP) foi proposta inicialmente em Jackson (1985). Em seguida, Geladi & Kowalski (1986a) desenvolveram algoritmos para aplicação destas técnicas. Trabalhos posteriores sobre a utilização das técnicas de projeção multivariada de dados para ajuste das variáveis de processo foram realizados por Lowry *et al.* (1992) e Montgomery *et al.* (1993), os quais demonstraram como obter novas variáveis independentes através dos CP originados a partir das variáveis de interesse no processo, quando o processo está em estado de controle estatístico, para a elaboração do modelo de referência do processo. A utilização de Mínimos Quadrados Parciais (MQP) para obter as novas variáveis independentes ao invés de obtê-las a partir dos CP foi recomendada em Kresta *et al.*(1991).

A utilização de CCM em processos contínuos foi apresentada em Kresta *et al.* (1991), MacGregor *et al.*(1994) e Kourti & MacGregor (1996). Estas cartas foram elaboradas utilizando a estatística de Hotelling e como variáveis de entrada nas cartas, as projeções multivariadas dos dados originais de processo, obtidas através da ACP ou MQP. Analogamente, Nomikos & MacGregor (1994) e Kourti *et al.* (1995) apresentaram essa mesma abordagem de controle estatístico multivariado aplicada em processos em bateladas.

Considerações sobre abordagem estatística multivariada utilizada em controle estatístico de processos foram realizadas por MacGregor (1995). Os principais problemas decorrentes da abordagem tradicional univariada dos processos industriais frente a natureza multivariada de suas variáveis de interesse foram seus objetos de estudo. O autor também apresentou considerações sobre a utilização de métodos de projeção estatística multivariada, como ACP, MQP e Análise de Correlação Canônica (ACC) para tratamento das variáveis originais de processo. A aplicação das CCM usando a estatística de Hotelling e ACP para um processo industrial em bateladas foi ilustrada em um estudo de caso na produção de borracha butadieno-estireno (SBR).

A abordagem proposta por MacGregor (1995) foi ampliada por Nomikos & MacGregor (1995). Os autores avaliaram o problema do uso de dados com trajetórias variáveis no tempo medidos em muitas variáveis de processos ao longo da duração finita de um processo em batelada. Para tanto, os autores utilizaram ACP para comprimir a informação contida nas trajetórias dos dados dentro de um espaço dimensionalmente baixo, descrevendo a operação das bateladas anteriores e desenvolveram uma CCM para o monitoramento *on-line* do progresso de novas bateladas. Os LC para CCM foram desenvolvidos utilizando informações obtidas a partir da distribuição de referência histórica dos dados, obtidos de bateladas anteriores sobre as quais não houve incidência de causas especiais.

Métodos para o monitoramento multivariado de processos e produtos, baseados nas estatísticas *Qui-Quadrado* e de Hotelling utilizando ACP e MQP para reduzir a dimensionalidade das variáveis de interesse no processo foram apresentados em Kourti & MacGregor (1996). Os autores também sugeriram métodos para detecção das variáveis de processo que contribuem para o sinal fora de controle na CCM. A ilustração dos métodos propostos foi realizada em um estudo de caso aplicado em um processo de fabricação de polietileno de baixa densidade.

Algumas aplicações práticas de metodologias de controle multivariado de processos utilizando técnicas de projeções multivariadas de dados podem ser encontradas na literatura. A aplicação de CCM utilizando dados baseados em modelos estatísticos multivariados (ACP e MQP) foi ilustrada por Chen & McAvoy (1996) em estudos de caso realizados em uma coluna de destilação binária e para o processo industrial da Tennessee Eastman (ver Downs & Vogel (1993)). Martin *et al.* (1996) apresentou aplicações da CCM baseada em ACP e análise de componentes principais não linear (ACPNL) realizadas em dois estudos de caso utilizando processos industriais.

Um método alternativo para definir a região normal de controle no controle multivariado de processos baseado na ACP robusta, através de projeção de busca (PB - *Projection Pursuit*) foi apresentado em Chen *et al.* (1996). O método foi elaborado para minimizar a excessiva sensibilidade da aplicação de ACP na presença de dados expúrios no monitoramento. A técnica de projeção de busca pesquisa um subespaço de baixa dimensão, de forma que a configuração dos dados neste subespaço reflita a estrutura e as falhas dos dados originais de uma maneira otimizada.

Uma metodologia para otimização da distribuição de peso molecular de polímeros lineares produzidos em reatores em batelada e semi-batelada foi elaborada por Clarke-Pringle

& MacGregor (1998). A abordagem proposta combina um método para otimização do processo batelada a batelada com o controle estatístico multivariado baseado em ACP. A metodologia desenvolvida baseou-se em conhecimentos fundamentais de polimerização para simplificar o problema a ser otimizado e fornecer liberdade na manipulação da seleção de variáveis. O controle multivariado de processos foi utilizado para definir quando a otimização de uma nova batelada era requerida. A metodologia desenvolvida foi aplicada em uma simulação em reator semi-batelada para produção de poliestireno.

2.4.1 Carta de Controle Multivariada Baseada em Componentes Principais para Processos em Batelada

Processos em batelada são utilizados para fabricação de produtos alimentícios, bioquímicos, farmacêuticos e químicos, entre outros. Caracterizam-se por serem processos de duração finita no tempo. Neste tipo de processo as matérias-primas são alimentadas, sofrem transformações durante um certo período de tempo, correspondente à duração da batelada. Ao término da batelada, o produto final é descarregado.

Existem três categorias diferentes de variáveis nos processos 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 obtidas por amostragem *on-line* do desempenho do processo. As variáveis de qualidade do produto final Y são obtidas através de análises laboratoriais no produto acabado, após o término da batelada.

Os processos em batelada, em geral, não são estacionários. Isto significa que a média ou valor alvo das variáveis de processo não é constante. Normalmente as variáveis de processo descrevem um perfil de variação ao longo do tempo de duração da batelada. Este perfil de variação no tempo caracteriza a trajetória da variável durante a batelada. Desta forma, a variação normal de uma variável de processo no tempo de duração da batelada é descrita através de sua trajetória padrão.

As variações entre bateladas podem ser consequência de uma combinação inadequada de matérias primas utilizada no início da batelada, impurezas presentes no processo e desvios das variáveis de processo em relação as suas trajetórias padrão (Nomikos e MacGregor, 1995). O controle deste tipo de processo visa manter uma variabilidade mínima entre as bateladas, gerando um produto final dentro dos padrões de qualidade desejados. Na realidade o objetivo principal é que, em cada batelada, a trajetória padrão de cada variável de processo não seja significativamente alterada (ver Marsh & Tucker, 1991, Nomikos & MacGregor,

1995, e Martin & Morris, 1996). Quando houver incidência de alguma causa especial de variação sobre o processo o resultado será uma variabilidade excessiva entre bateladas. Indicando que ocorreram alterações na trajetória padrão de uma ou mais variáveis do processo. Como consequência o produto final estará fora das especificações de qualidade.

A filosofia do controle estatístico de processos é desenvolver um modelo empírico que caracterize o processo e realizar o monitoramento futuro através da comparação com este modelo. O modelo é construído utilizando a distribuição de referência dos dados obtidos quando o processo operou em estado de controle estatístico e resultou em produto de qualidade satisfatória. Num processo em bateladas, obter a distribuição de referência das variáveis de processo implica em modelar sua trajetória padrão, que é a trajetória realizada durante bateladas que resultaram em produtos aceitáveis.

Em cada batelada, amostras consecutivas são obtidas das variáveis de processo. Ao final da batelada, dispõe-se de várias medidas 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 correlação entre variáveis de processo, mas a estrutura de autocorrelação de cada variável dentro das bateladas. Para tanto utiliza-se a ACP como técnica estatística de tratamento dos dados de processo. A variação das trajetórias das variáveis de processo entre bateladas é caracterizada em um espaço de projeção de variáveis latentes definido pelo modelo ACP.

O desempenho de novas bateladas é então comparado com a distribuição de referência testando-se a seguinte hipótese H_0 : As medidas das trajetórias das variáveis de processo no tempo corrente em uma nova batelada estão consistentes com a batelada em operação normal, definida pela distribuição de referência (Nomikos & MacGregor, 1995).

Os dados de processo provenientes das bateladas podem ser organizados de três maneiras diferentes. Considere uma batelada com $j = 1, 2, \dots, J$ variáveis de processo sendo medidas em $k = 1, 2, \dots, K$ intervalos de tempo durante a batelada. Dados similares são obtidos em muitas outras $i = 1, 2, \dots, I$ bateladas do processo. Todos os dados podem ser resumidos na matriz tridimensional $\underline{\mathbf{X}}$ ($I \times J \times K$), ilustrada na Figura 2.6 (Nomikos & MacGregor, 1995). Na Figura 2.6, as diferentes bateladas são organizadas no eixo vertical, as medidas das variáveis no eixo horizontal e sua evolução no tempo ocupa a terceira dimensão. Cada fatia horizontal de $\underline{\mathbf{X}}$ é uma matriz de dados ($J \times K$) representando as trajetórias no tempo das variáveis de uma única batelada (i). As fatias verticais de $\underline{\mathbf{X}}$ são matrizes de dados ($I \times J$)

representando os valores de todas as variáveis para todas as bateladas em um intervalo de tempo específico (k).

O método de projeção multivariada utilizado no tratamento de dados de processos em batelada organizados como na Figura 2.6 é a Análise de Componentes Principais *Multiway* (ACPM), proposto em Geladi *et al.*(1986). Este método é consistente com a ACP, apresentando os mesmos objetivos e benefícios. O ACPM é aplicado sobre o desdobramento da matriz tridimensional $\underline{\mathbf{X}}$, fatia por fatia, de três maneiras diferentes. O desdobramento resulta em uma grande matriz \mathbf{X} bidimensional (que pode ser organizada de duas maneiras) onde então é aplicada a ACP. Cada uma destas seis maneiras de desdobramento de $\underline{\mathbf{X}}$ corresponde à representação de um tipo diferente de variabilidade presente nos dados.

Para analisar o monitoramento do processo em batelada, Nomikos & MacGregor (1995) sugerem desdobrar $\underline{\mathbf{X}}$ colocando suas fatias verticais ($I \times J$) lado a lado, iniciando com a fatia correspondente ao primeiro intervalo de tempo. O resultado é a matriz \mathbf{X} com dimensões ($I \times JK$) na Figura 2.6. Este desdobramento permite analisar a variabilidade entre as bateladas resumindo as informações dos dados com relação as variáveis e suas variações no tempo.

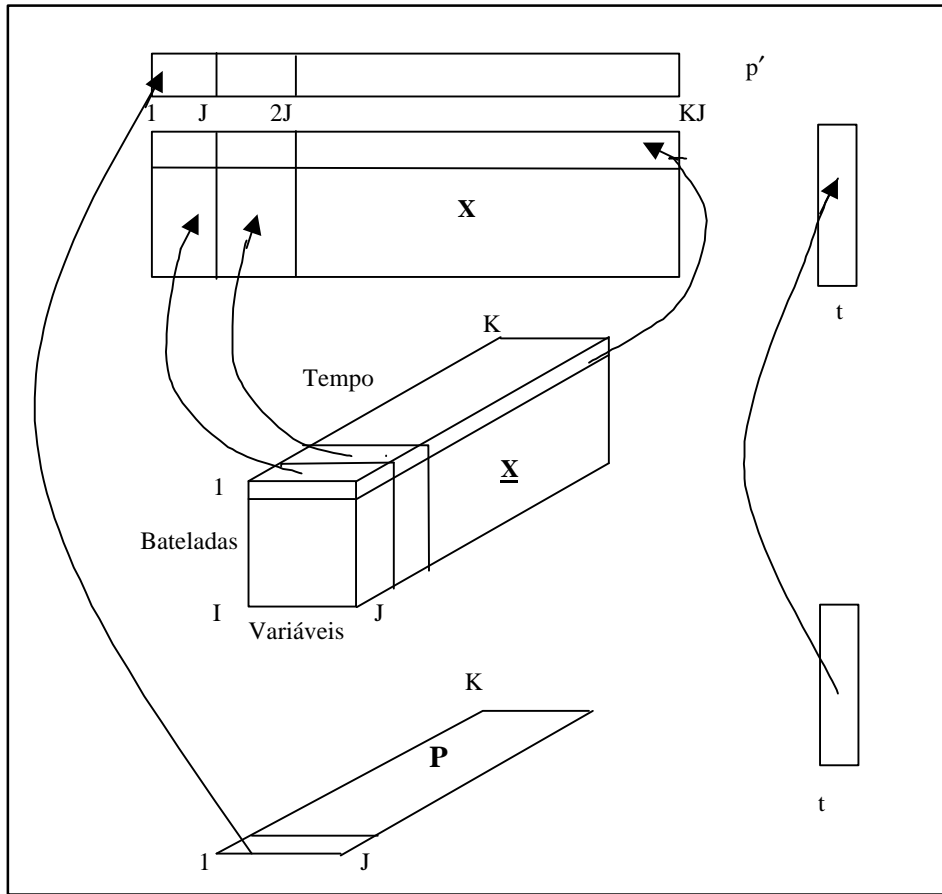


Figura 2.6 - Arranjo dos dados da batelada; três maneiras de decomposição da matriz \underline{X} através da ACPM (Nomikos & MacGregor, 1995)

Antes de aplicar a ACPM na matriz \underline{X} , os dados devem ser normalizados. Isto é efetuado subtraindo a média das colunas da matriz \underline{X} e dividindo cada coluna de variáveis pelo seu desvio padrão. Ao subtrair-se a média de cada coluna na matriz, subtrai-se a trajetória média de cada variável e remove-se as principais não-linearidades e componentes dinâmicos dos dados. A divisão das colunas da matriz pelo seu desvio padrão elimina as diferenças de unidades entre as variáveis e pondera igualmente todas as variáveis em cada intervalo de tempo. A ACPM aplicada a estes dados normalizados da matriz \underline{X} fornece um estudo para a variação das trajetórias no tempo de todas as variáveis em todas as bateladas com relação as suas trajetórias médias.

A ACPM decompõe os dados de \underline{X} (ou de \underline{X}) em uma série de componentes principais. Estes componentes principais permitem a representação dos dados como R produtos de vetores de escores (\underline{t}) e matrizes de cargas (\underline{P} ou \underline{p}), acrescidos de uma matriz de

resíduos ($\underline{\mathbf{E}}$ ou \mathbf{E}). As equações (2.71) e (2.72) apresentam a decomposição dos dados de $\underline{\mathbf{X}}$ e de \mathbf{X} em componentes principais.

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{t}_r \otimes \mathbf{P}_r + \underline{\mathbf{E}} \quad (2.71)$$

$$\mathbf{X} = \sum_{r=1}^R \mathbf{t}_r \mathbf{p}'_r + \mathbf{E} \quad (2.72)$$

A Figura 2.7 (Nomikos & MacGregor, 1994) demonstra como a ACPM explica a variação das medidas das variáveis em torno de suas trajetórias médias calculadas a partir do modelo de referência. O $i^{\text{ésimo}}$ elemento do vetor de escores \mathbf{t} corresponde a $i^{\text{ésima}}$ batelada e resume a variação total desta batelada durante seu tempo de duração com relação as outras bateladas do modelo de referência. A matriz de cargas \mathbf{P} resume a variação no tempo das medidas das variáveis sobre suas trajetórias médias. Os elementos da matriz \mathbf{P} são pesos que quando aplicados a cada variável em cada intervalo de tempo dentro da batelada resultam no escore \mathbf{t} para aquela batelada (Nomikos & MacGregor, 1994).

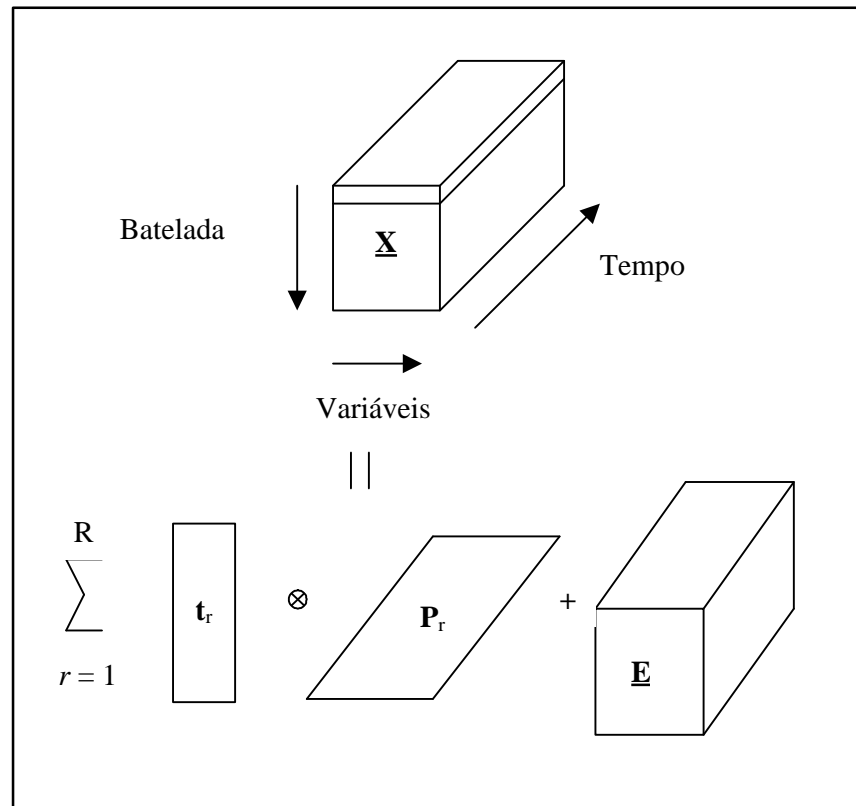


Figura 2.7 – Decomposição da matriz $\underline{\mathbf{X}}$ através da ACPM (Nomikos & MacGregor, 1994)

O algoritmo NIPALS (*Nonlinear interactive partial least squares*), recomendado para o cálculo seqüencial dos componentes principais é apresentado em Geladi & Kowalski (1986).

O modelo de distribuição de referência das bateladas em operação normal é construído a partir dos dados históricos do processo. A ACPM é utilizada para selecionar as bateladas sujeitas apenas às causas comuns de variação do processo. Os gráficos de escores e de resíduos são utilizados para diferenciar as bateladas boas das ruins nos dados históricos. As bateladas ruins, ou que geraram produtos fora dos padrões de qualidade desejados, devem ser retiradas da base de dados históricos. A ACPM aplicada à base de dados históricos do processo contendo apenas bateladas boas constitui o modelo contendo a distribuição de referência dos dados. O número R de componentes principais necessários no modelo de referência pode ser determinado utilizando-se uma das regras apresentadas na seção 2.1.1.1.

Os escores t são combinações lineares das variáveis aleatórias da matriz de dados \mathbf{X} , que foram normalizados. Os escores t dos R CP selecionados no modelo de referência são multinormalmente distribuídos, com média zero e matriz de covariâncias $\mathbf{S}(R \times R)$. A matriz \mathbf{S} é uma matriz diagonal devido a ortogonalidade dos escores t . A estatística de Hotelling é dada na equação (2.73) (Nomikos & MacGregor, 1995).

$$T^2 = \mathbf{t}'_R \mathbf{S}^{-1} \mathbf{t}_R \frac{I}{(I-1)^2} \approx B_{\frac{R}{2}, \frac{(I-R-1)}{2}} \quad (2.73)$$

Na equação (2.73) o vetor \mathbf{t}_R contém os escores de uma dada batelada a partir dos R CP retidos no modelo e I é o número total de bateladas. A estatística T^2 corresponde à distância de *Mahalanobis* no espaço reduzido (definido pelos R CP do modelo de ACPM) entre a posição de uma batelada (definida pelos escores t) e a origem no espaço reduzido, que designa o ponto com a mínima variação no desempenho do processo.

Os valores críticos da variável beta, B , para um nível de significância α , podem ser encontrados a partir dos valores críticos da distribuição F , utilizando-se a relação dada na equação (2.74).

$$B_{\frac{R}{2}, \frac{(I-R-1)}{2}} = \frac{\left(\frac{R}{I-R-1}\right) F_{R, I-R-1, \alpha}}{1 + \left(\frac{R}{I-R-1}\right) F_{R, I-R-1, \alpha}} \quad (2.74)$$

A evolução de uma nova batelada em tempo real é monitorada no espaço reduzido definido pelos CP do modelo ACPM. A matriz de cargas \mathbf{P} contém toda informação estrutural

de como as medidas das variáveis poderiam desviar-se de seus valores médios em cada intervalo de tempo. A predição dos escores t para cada um dos R CP e dos resíduos \mathbf{E} para uma nova batelada $\mathbf{X}_{nova}(K \times J)$, desdobrada em $\mathbf{X}_{nova}^t(1 \times JK)$ são dadas na equação (2.75) e (2.76), respectivamente (Nomikos & MacGregor, 1995).

$$t_r = \mathbf{X}_{nova}^t \mathbf{p}_r \quad (2.75)$$

$$\mathbf{E} = \mathbf{X}_{nova} - \sum_{r=1}^R t_r \mathbf{p}_r \quad (2.76)$$

Nas equações (2.75) e (2.76) a matriz a matriz \mathbf{X}_{nova} contém as medidas das variáveis, após serem normalizadas e t_r é o escore correspondente ao $r^{ésimo}$ CP.

O somatório do quadrado dos resíduos em cada batelada i , para o intervalo de tempo k , é dado na equação (2.77).

$$Q_i = \sum_{k=1}^K \sum_{j=1}^J \mathbf{E}(k, j) = \sum_{j,k=1}^{JK} (x_{j,k} - \hat{x}_{j,k})^2 \quad (2.77)$$

Os limites de controle para a estatística Q_i , são obtidos, a partir da matriz \mathbf{X} de dados preliminares do processo, utilizando resultados aproximados da distribuição das formas quadráticas (Jackson & Mudholkar, 1979).

Fazendo:

$$q_1 = \sum_{l=r+1}^R l_l, \quad q_2 = \sum_{l=r+1}^R l_l^2, \quad q_3 = \sum_{l=r+1}^R l_l^3 \quad \text{e} \quad h_0 = 1 - \frac{2q_1 q_3}{3q_2^2}; \quad (2.78)$$

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

$$Z = \frac{q_1[(Q_i/q_1)^{h_0} - 1 - q_2 h_0 (h_0 - 1)/q_1^2]}{\sqrt{2q_2 h_0^2}} \quad (2.79)$$

segue uma distribuição Normal com média zero e desvio padrão unitário [$Z \sim N(0,1)$]. Os limites de controle da estatística Q_i são dados na equação (2.80).

$$\text{LSC} = q_1 \left[\frac{Z_\alpha \sqrt{2q_2 h_0^2}}{q_1} + 1 + \frac{q_2 h_0 (h_0 - 1)}{q_1^2} \right]^{1/h_0} \quad \text{e}$$

$$\text{LIC} = 0, \quad (2.80)$$

onde $|Z_a|$ representa o desvio positivo máximo tolerado na distribuição Normal Z , dado a 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.

Existem duas maneiras de verificar se uma nova batelada apresenta desvios do modelo ACPM elaborado para o processo. A primeira delas é quando os valores dos escores $t_{r,k}$ estão fora dos limites definidos pela região de controle no espaço reduzido. A outra maneira é quando os resíduos $\mathbf{E}(k,j)$ são grandes. No primeiro caso, o modelo elaborado ainda é válido, mas a magnitude da variação durante a nova batelada é grande. No segundo caso, o modelo não pode ser mais considerado válido, devido à ocorrência de novos eventos, não contemplados na base de dados de referência; dessa forma, os dados da nova batelada não estão projetados adequadamente no espaço reduzido e os resíduos capturam alguma variabilidade não suficientemente descrita no modelo de referência (Nomikos & MacGregor, 1994). Se os escores situam-se próximos da origem e os resíduos são pequenos, o desempenho da nova batelada é similar ao desempenho das bateladas da distribuição de referência, significando que o processo permanece em controle estatístico.

No monitoramento *on-line*, a melhor maneira de rastrear se algo diferente ocorreu em um instante particular de tempo é através dos resíduos. Para tanto, o mais adequado é a utilização do quadrado dos erros de predição (QEP), calculado de acordo com a equação (2.81) (Nomikos & MacGregor, 1994).

$$QEP_k = \sum_{j=1}^J \mathbf{E}(k, j)^2 \quad (2.81)$$

A justificativa para a utilização do QEP no monitoramento dos resíduos baseia-se no fato de que ele refere-se ao somatório do quadrado dos erros diretamente relacionados com as últimas medidas do processo, no tempo k . O Q , além estar relacionado com as medidas no tempo k , refere-se também ao erro associado à previsão das futuras observações na matriz \mathbf{X}_{nova} . Desta forma, se o interesse é verificar em algum intervalo específico de tempo alguma causa especial, o QEP é a melhor alternativa (Nomikos & MacGregor, 1994).

Para uma nova batelada, a matriz de dados \mathbf{X}_{nova} não estará completa antes do término da batelada. Em cada intervalo de tempo durante a operação da batelada, a matriz \mathbf{X}_{nova} possui apenas as medidas de processo até o intervalo de tempo considerado. O restante das medidas da matriz referentes aos intervalos de tempo até o término da batelada são desconhecidas. Quando a batelada está no $k^{ésimo}$ intervalo de tempo, a matriz \mathbf{X}_{nova} contém apenas as primeiras

k linhas completas, faltando os dados correspondentes aos demais intervalos até o término da batelada. Para aplicar a ACPM, todavia a matriz \mathbf{X}_{nova} de dados deve estar completa.

Nomikos & MacGregor (1995) apresentam três abordagens para o preenchimento dos dados que faltam na matriz \mathbf{X}_{nova} entre o tempo corrente k e o término da batelada. O objetivo é preencher os dados futuros na matriz de forma que os valores preditos para os escores t a cada intervalo de tempo sejam tão próximos quanto possível dos valores verdadeiros ao final da batelada. Estes métodos envolvem a utilização de todos os vetores de cargas \mathbf{p}_r ($JK \times 1$) da distribuição de referência obtida do modelo ACPM e o preenchimento dos dados futuros na matriz \mathbf{X}_{nova} de diferentes maneiras.

A primeira abordagem para o preenchimento dos dados em \mathbf{X}_{nova} considera que os dados futuros estão de acordo com as trajetórias médias calculadas no modelo de referência elaborado. Isto implica em admitir que a batelada deverá operar normalmente até seu término. Desta forma, deve-se preencher com zeros os dados que faltam na matriz \mathbf{X}_{nova} . Esta abordagem origina uma boa representação da operação em batelada no gráfico dos escores t e rápida detecção de anormalidades no gráfico *QEP*. A desvantagem é que o gráfico dos escores apresenta uma inércia na detecção de anormalidades na operação, especialmente no início da batelada.

A segunda abordagem considera que os desvios futuros nas trajetórias das variáveis com relação a suas trajetórias médias durante o tempo restante da batelada permanecem de acordo com os valores correntes no intervalo de tempo k . Desta forma, supõe-se que os mesmos erros persistam no processo até o término da batelada. Nesta abordagem, o gráfico dos escores t detecta mais rapidamente anormalidades, mas o gráfico *QEP* não é tão sensível a alterações quanto na abordagem anterior.

A terceira abordagem considera as observações futuras como valores que faltam no modelo ACPM elaborado. A partir desta consideração, os CP da base de dados de referência podem ser usados na predição dos valores que faltam. Isto é feito considerando-se os valores que faltam consistentes com os valores já obtidos no intervalo de tempo k . Com as estrutura de correlações das variáveis no modelo de referência definido pela matriz de cargas \mathbf{P} , as observações conhecidas $\mathbf{X}_{nova,k}$ ($1 \times JK$) podem ser projetadas no espaço reduzido. A partir disso, os escores t podem ser calculados em cada intervalo de tempo pela equação (2.82).

$$\mathbf{t}_{R,k} = (\mathbf{P}_k^t \mathbf{P}_k)^{-1} \mathbf{P}_k^t \mathbf{X}_{nova,k} \quad (2.82)$$

Na equação (2.82), $\mathbf{P}_k(JK \times R)$ é uma matriz contendo nas colunas todos os elementos dos de cargas \mathbf{p}_r até o intervalo de tempo k para todos os CP. Esta abordagem é superior às anteriores em casos onde 10% da batelada já é conhecida (Nomikos & MacGregor, 1995).

A escolha de qual abordagem utilizar depende das características do processo a ser monitorado. Se o processo em batelada não apresenta distúrbios persistentes ou variáveis com descontinuidades em suas trajetórias, as melhores abordagens são a primeira ou a terceira. Mas se o processo, todavia apresentar distúrbios persistentes, a segunda abordagem é a mais indicada (Nomikos & MacGregor, 1995).

A realização do monitoramento *on line* é efetuada através da CCM baseada em CP, das cartas de Shewhart para os R escores dos CP retidos no modelo ACPM elaborado e da carta para os resíduos QEP.

Para tanto, sabe-se que os escores são combinações lineares das variáveis e, por força do teorema do limite central, podem ser considerados distribuídos Normalmente. Os desvios da normalidade resultam das abordagens utilizadas para preencher os dados ausentes em $\mathbf{X}_{nova,k}$. Considerando os escores normalmente distribuídos e independentes a equação (2.83) apresenta os limites de controle, em um nível de significância α , da carta de Shewhart para escores.

$$\pm t_{n-1, \frac{\alpha}{2}} S_{ref} \left(1 + \frac{1}{n}\right)^{1/2} \quad (2.83)$$

Na equação (2.83), n é o número de observações, S_{ref} é o desvio padrão estimado do escore t no intervalo de tempo k e $t_{n-1, \alpha/2}$ é o valor crítico de uma variável de *Student* com $n-1$ graus de liberdade e nível de significância $\alpha/2$.

No monitoramento *on line*, estatística de Hotelling para um novo vetor de escores independentes \mathbf{t} é dada na equação (2.84). Os comprimentos dos eixos da elipsóide de confiança na direção do $r^{ésimo}$ CP são dados na equação (2.85) (Nomikos & MacGregor, 1995).

$$T^2 = \mathbf{t}_{R,k}^t \mathbf{S}^{-1} \mathbf{t}_{R,k} \frac{I(I-R)}{R(I^2-1)} \approx F_{R, I-R} \quad (2.84)$$

$$\pm \sqrt{S(r, r) F_{2, I-2, \alpha} \frac{2(I^2-1)}{I(I-2)}} \quad (2.85)$$

Os erros de predição $\mathbf{E}(k,j)$ são também distribuídos normalmente, embora não sejam independentes. O QEP é uma forma quadrática que pode ser bem aproximada pela distribuição $g\chi_h^2$, onde g é uma constante e h são os graus de liberdade da distribuição χ^2 (Box, 1954). Os valores de g e h podem ser estimados a partir dos autovalores da matriz Σ ou através do somatório dos momentos da distribuição $g\chi_h^2$ com os momentos obtidos da distribuição de referência em cada intervalo de tempo k . Isto é feito igualando-se a média ($m = gh$) e a variância ($S^2 = 2g^2h$) da distribuição $g\chi_h^2$ à média (m) e à variância (v) do QEP amostral em cada intervalo de tempo k .

Os limites de controle para o QEP , em um nível de significância α para o intervalo de tempo k são dados na equação (2.86), onde $c_{\frac{2m^2}{v}, \alpha}^2$ denota o valor crítico de uma variável Qui-quadrado com $2m^2/v$ graus de liberdade e nível de significância α (Nomikos & MacGregor, 1995).

$$QEP_{\alpha} = \left(\frac{v}{2m} \right) c_{\frac{2m^2}{v}, \alpha}^2 \quad (2.86)$$

Informações sobre alterações na distribuição dos resíduos ao longo da batelada podem ser obtidas segundo análise de cartas para as estatísticas g e h . Baixos valores nos graus de liberdade h indicam que a distribuição é dominada por uma grande variabilidade. Isto significa que apenas poucas medidas das variáveis permanecem em torno de suas trajetórias médias. Valores elevados de h ocorrem em períodos mais estáveis. Nestes períodos, os desvios das variáveis em relação as suas trajetórias médias estão contribuindo no QEP . A constante g é um fator de escala que possibilita efetuar a soma dos momentos.

Considerando apenas o intervalo de tempo corrente, o erro do tipo I (caracterizado pela ocorrência de um alarme falso, ver Seção 2.2.1) é dado pelo valor α utilizado no teste estatístico da carta de controle. Para o monitoramento de uma batelada inteira, este procedimento não é válido. Neste caso, o erro do tipo I está associado com o tempo total de duração da batelada e não apenas com o valor de α para o teste estatístico instantâneo em um intervalo de tempo. Além disso, os valores dos escores t e dos QEP em intervalos sucessivos de tempo não são independentes. O erro total do tipo I para as cartas dos escores e para as cartas QEP pode ser estimado como sendo o número de valores que no teste estatístico situam-se fora dos limites de controle, nos dados da distribuição de referência, dividido pelo número total de observações (IK) (Nomikos & MacGregor, 1995).

Finalmente, algumas vantagens da estratégia de monitoramento descrita acima são apresentadas na seqüência:

- As CCP permitem o controle simultâneo de um grande número de variáveis de processo, oferecendo-se como ferramenta eficiente no monitoramento de processos automatizados.
- As CCP permitem o controle de variáveis que descrevem um perfil de variação ao longo do tempo, essas variáveis são de ocorrência bastante comum na indústria.
- A utilização das CCP não pressupõe variáveis Normalmente distribuídas; na prática, as variáveis podem apresentar qualquer distribuição de probabilidade.
- As CCP podem ser aplicadas quando sucessivos pontos amostrais das variáveis de processo apresentarem dependência.

2.4.2 Identificação das variáveis fora de controle em cartas multivariadas

Para que o controle estatístico multivariado de qualidade seja satisfatório é necessário: (i) especificar a probabilidade de alarme falso a ser tolerada e (ii) ser capaz de sinalizar rapidamente quando o processo de fato está fora de controle estatístico.

Existem três propriedades importantes para definir um bom procedimento de controle multivariado de processos: (i) o procedimento deve ser capaz de responder a questão: "O processo está sob controle?"; (ii) a probabilidade de erro do tipo I deve estar estabelecida com precisão; (iii) o procedimento deve levar em conta as correlações entre as variáveis (dada na matriz Σ) na determinação do valor crítico da estatística utilizada (Jackson, 1985).

Uma vez que uma falha ou evento especial tenha ocorrido em uma das cartas (carta de Hotelling, equação (2.84), e carta *QEP*, equação (2.81) ou nas cartas dos escores, equação (2.75)) é importante diagnosticar e encontrar as suas possíveis causas. Isto envolve duas etapas: a primeira etapa consiste em encontrar qual(s) variável(s) contribuiu para o sinal fora de controle e a segunda etapa consiste em identificar o que ocorreu no processo para ocasionar o(s) desvio(s) nesta(s) variável(s). Para tanto, vários métodos foram propostos por diversos autores, podendo ser divididos em dois grupos: os procedimentos de diagnóstico via cartas univariadas e os procedimentos de diagnóstico via decomposição da estatística T^2 de Hotelling.

A estratégia para a decomposição de T^2 denominada *step-down* foi apresentada em Wierda (1994b). Este procedimento baseia-se numa ordenação das médias das p variáveis de interesse. De acordo com a ordenação, o procedimento particiona o vetor de médias em q ($< p$) subvetores e define q sub-hipóteses; cada sub-hipótese testa a possibilidade de um determinado subvetor de médias não ter sofrido alterações. Uma carta de controle é associada a cada uma das q sub-hipóteses.

Outras abordagens para detecção e diagnóstico no controle multivariado de processos foram propostas. Métodos baseados em ajustes por regressão foram sugeridos por Hawkins (1993), Wade & Woodall (1993) e Zhang (1985). Estes métodos utilizam os gráficos de resíduos das variáveis monitoradas e tentam remover os efeitos das variáveis fora de controle por regressão.

Métodos de decomposição da estatística de Hotelling T^2 foram sugeridos por Mason, Tracy & Young (1995). Os autores demonstraram que os procedimentos sugeridos por Murphy (1987) e procedimento *step-down* de Wierda *apud* Sullivan & Woodal (1996) são subconjuntos de seu método proposto.

Quando um grande número de variáveis devem ser monitoradas, os procedimentos de diagnóstico via decomposição da estatística de Hotelling podem não ser viáveis na prática. Uma alternativa são os procedimentos de diagnóstico via cartas univariadas.

Uma abordagem de diagnóstico utilizando a carta de Hotelling T^2 e o gráfico de escores normalizados dos CP para identificar qual CP gerou o sinal fora de controle na carta de controle foi proposto por Jackson (1980). Quando apenas poucas variáveis estão envolvidas, as cartas univariadas podem detectar os desvios com eficiência. Além disso, quando os escores possuem um significado físico, então um grupo de variáveis podem afetar apenas um CP específico. A partir da interpretação física deste CP no processo monitorado, o problema pode ser diagnosticado. Este procedimento é menos eficiente quando os escores não possuem significado físico no processo em estudo. Uma vez que os CP são combinações lineares de todas as variáveis de processo, a interpretação dos resultados passa a ser uma tarefa mais complexa (Kourti & MacGregor, 1996).

Jackson (1991) modificou o procedimento sugerido em Jackson (1980) utilizando limites de controle de *Bonferroni* nas cartas univariadas dos escores para o diagnóstico. Uma vez que os escores são independentes, a utilização destes limites de controle soluciona problemas de erro do tipo I.

A utilização de p cartas univariadas de *Shewhart* simultaneamente ao uso da carta multivariada de Hotelling no controle do processo foi proposta por Alt (1984). Neste procedimento, a carta multivariada é utilizada para detectar mudanças no processo. Quando é gerado um sinal fora de controle na carta multivariada, verifica-se, através das cartas univariadas, qual grupo de variáveis pode ter sido responsável pelo sinal. Os limites de controle de *Bonferroni* devem ser utilizados nas cartas univariadas para minimizar o erro tipo I. Estes limites são obtidos substituindo $\alpha/2$ por $\alpha/2p$ em cada uma das p cartas univariadas. Assim, amplia-se os limites de controle nestas cartas para obter uma probabilidade total de erro tipo I (alarme falso) de no máximo α .

Uma discussão sobre a utilização dos limites controle de *Bonferroni* nas cartas univariadas para o diagnóstico das variáveis responsáveis pelo sinal fora de controle na carta multivariada foi apresentado em Hayer & Tsui (1994). Os autores apresentaram um procedimento baseado na construção de intervalos de confiança simultâneos para as médias das variáveis. Para tanto, utilizaram procedimentos paramétricos e não paramétricos e avaliaram os pontos críticos (valores do teste estatístico da carta de controle) através de tabelas de frequências, integração numérica e simulação.

O procedimento de diagnóstico através da construção de um gráfico de barras para os erros normalizados das variáveis e um gráfico de barras para os escores normalizados dos CP foi sugerido em Kourti & MacGregor (1996). Nestes gráficos, as variáveis que possuem elevados valores para os erros normalizados ou para os escores normalizados são detectadas. Uma variação deste procedimento é o monitoramento conjunto da carta multivariada de Hotelling e do gráfico de barras para as variáveis normalizadas, o que pode ser uma maneira eficiente de inspeção visual (Fuchs & Benjamini, 1994).

Suponha que no monitoramento das CCP num dado período k , um ponto amostral é plotado além dos limites de controle na CCM de Hotelling. Neste caso, um ou mais dos CP apresentou desvio anormal de seus valores de média, devendo ser identificados. O objetivo do monitoramento é determinar qual variável de processo desviou-se de sua trajetória esperada e atuar corretivamente sobre ela. Uma vez identificado o CP responsável pelo sinal na CCM de Hotelling, é necessário identificar qual (ou quais) das variáveis de processo contribuiu para o escore anormal apresentado pelo CP.

A abordagem mais usual para determinação da contribuição de variáveis de processo para o comportamento anormal de escores em CP utiliza gráficos de contribuição (Miller *et al* (1993), MacGregor *et al.* (1994), Kourti & McGregor (1996)).

Os gráficos de contribuição indicam o quanto cada variável está envolvida na composição de um determinado escore. São gráficos de barra onde são plotados os valores observados das variáveis de processo no período onde a CCM de Hotelling sinalizou uma causa especial. Como as variáveis foram padronizadas, os gráficos apresentam uma linha central em zero. Variáveis com valores positivos ou negativos muito diferentes de zero (por exemplo, valores maiores que $|2|$) são candidatos à investigação detalhada. Um bom procedimento de diagnóstico é o monitoramento dos gráficos de contribuição juntamente com os gráficos dos escores. Este procedimento permite a fácil visualização de quais as variáveis que contribuíram para o sinal fora de controle na carta multivariada, caso ele venha a ocorrer.

A contribuição de cada variável normalizada z_j para o cálculo do escore do CP a é dada na equação (2.87). Sugere-se investigar as variáveis com contribuição elevada e com mesmo sinal do escore (Kourti & MacGregor, 1996).

$$cont_{a,j} = p_{a,j}(z_j - m_j) \quad (2.87)$$

Freqüentemente, mais de um escore pode apresentar valor elevado no gráfico dos escores. Neste caso, recomenda-se o cálculo da contribuição média total por variável para todos os escores com valores elevados. Para cada um dos escores normalizados com valores elevados (acima de 2,5), deve ser calculada a contribuição das variáveis com mesmo sinal do escore.

Sugere-se duas etapas para o cálculo da contribuição total da variável z_j para todos os n ($n < R$) escores com valores elevados (Kourti & MacGregor, 1996).

1. Repetir o procedimento (i) e (ii) abaixo para todos os n escores com valores elevados ($n < R$):

- (i) Calcular a contribuição da variável z_j no escore normalizado (t_a/s_a^2) através da equação (2.88).

$$cont_{a,j} = \frac{t_a}{s_a^2} p_{a,j}(z_j - m_j) = \frac{t_a}{|a|} p_{a,j}(z_j - m_j) \quad (2.88)$$

- (ii) Se o sinal da $cont_{a,j}$ calculada na equação (2.88) for contrário ao sinal do escore, considerá-la igual a zero.

2. Calcular a contribuição total para a variável z_j através da equação (2.89).

$$CONT_j = \sum_{a=1}^n cont_{a,j} \quad (2.89)$$

O gráfico de contribuição pode não revelar explicitamente a causa do evento especial sinalizado na carta de controle. Ele identifica o grupo de variáveis de processo que não estão consistentes com as condições de operação normal da batelada. A atenção dos operadores e engenheiros deve estar direcionada para utilização de seus conhecimentos sobre o processo produtivo permitindo a dedução das causas que provocaram as alterações nas variáveis identificadas.

2.5. Interface entre Controle de Processos na Engenharia e Controle Estatístico de Processo.

O controle de processos freqüentemente utilizado em engenharia mecânica, elétrica e química está associado ao controle progressivo/retroativo (*feedforward-feedback control*) ou controle proporcional integral diferencial (*proportional-integral-differential control*, PID) de processos. Estas ferramentas de controle estão associados com teorias de sistemas lineares. O controle de processos utilizado em Engenharia Industrial está associado ao monitoramento através de cartas de controle.

Faz-se necessário distinguir-se entre controle de processos de engenharia (CPE) e controle estatístico de processos (CEP). Os objetivos de ambos os controles são os mesmos: atingir os níveis alvos de processo com pequena variabilidade. A diferença entre estes dois tipos de controle encontra-se na estratégia utilizada para atingir o objetivo comum (Crowder *et al.*, 1997).

Em muitos sistemas de engenharia existe a possibilidade de descrever seu desempenho sob controle, seu estado fora de controle, os eventos que contribuiriam para os desvios no processo e as ações corretivas a serem tomadas. Isto possibilita elaborar uma forma de intervenção que mantenha o sistema em estado de equilíbrio com pequena variabilidade. Para tanto, supõe-se que a média do sistema em controle sofra alteração contínua governada por um processo estocástico. O controle é atingido através de ações corretivas sistemáticas tomadas freqüentemente para cada observação ou em pontos determinados pela observação dos dados. Estas ações sistemáticas podem tipicamente ser descritas através de uma fórmula matemática e freqüentemente implantadas utilizando-se reguladores automáticos ou dispositivos similares. A qualidade do controle efetuado é mensurada através variância na saída do sistema. O CPE, com controladores do tipo PID, utilizam esta abordagem.

A aplicação tradicional do CEP supõe que a média e a variância do processo em condições normais são estáveis, embora alterações bruscas em ambas possam ocorrer em

momentos desconhecidos de tempo. Sob condições normais de operação, a variabilidade do processo é causada apenas por causas comuns, inerentes ao processo produtivo. O efeito destas causas comuns freqüentemente é muito difícil de ser eliminado do processo. O processo atinge níveis de controle inaceitáveis na presença de causas especiais de variabilidade. O objetivo deste controle é, então, identificar rapidamente a existência de causas especiais e eliminá-las do processo ou neutralizar seus efeitos através de ajustes apropriados. As ferramentas utilizadas no CEP são as cartas de controle associadas a sistemas de diagnóstico e análise de dados. Nesta abordagem, o controle de qualidade do processo é medido em termos de índices de capacidade, ARL ou curvas de aprendizagem.

As ferramentas que são apropriadas no controle de sistemas estocásticos bem definidos são sub-ótimas no controle de sistemas sujeitos a alterações imprevistas e bruscas. É importante observar que existem ferramentas de CPE envolvendo modelos que descrevem o desempenho de processos e ferramentas de CEP capazes de descrever alterações bruscas e imprevistas nos processos. Ambas metodologias podem ser utilizadas em conjunto, complementando-se. Muitos processos industriais onde existe a utilização conjunta de ambas as filosofias de controle tem sido relatados na literatura (ver Box & Kramer, 1992; Yashchin, 1993b; Tucker *et al.*, 1993).

CAPÍTULO 3

3. MÉTODO PROPOSTO

Neste capítulo propõe-se um método para elaboração da Carta de Controle Multivariada baseada em Componentes Principais para processos em batelada.

Processos industriais realizados em bateladas são constituídos de três etapas :

- (i) Alimentação das matérias-primas no reator industrial;
- (ii) Processamento das matérias-primas dentro do reator, onde estas sofrem uma série de transformações. Este processamento deve ser controlado através do monitoramento das trajetórias das variáveis de processo, tais como temperaturas e pressões;
- (iii) Descarga da batelada, etapa na qual o produto final é analisado em laboratório quanto às suas características de qualidade.

Nos processos em bateladas três tipos de variáveis são importantes no monitoramento da qualidade e controle de processo:

- Variáveis de qualidade provenientes de medições de qualidade feitas sobre as matérias-primas, denominadas de variáveis de qualidade Z ;
- Variáveis provenientes de mensurações *on-line* do desempenho do processo, em pontos próximos no tempo, denominadas variáveis de processo X ;

- Variáveis de qualidade final do produto que são mensuradas no produto acabado, após a conclusão da batelada, denominadas de variáveis de qualidade *Y*.

A aplicabilidade deste tipo de controle estatístico de processos depende de três condições:

- Disponibilidade de informações históricas sobre o comportamento do processo (variáveis de processo e de qualidade).
- Coleta automatizada dos dados de processo.
- Bateladas com tempo fixo de duração.

A disponibilidade de dados históricos sobre as variáveis de processo selecionadas é importante para etapa posterior de elaboração da distribuição de referência contra a qual o desempenho do processo será avaliado.

A coleta automatizada dos dados relativos às variáveis de processo monitoradas é importante para possibilitar a utilização *on line* da Análise de Componentes Principais *Multiway* (ACPM) nos dados selecionados.

O processo produtivo deve ser realizado em bateladas com tempo fixo de duração. Isto permite, através da elaboração da distribuição de referência das variáveis de processo, a avaliação das trajetórias médias destas variáveis durante as bateladas. Dessa forma, se o processo está sob controle, cada variável estará seguindo sua trajetória padrão. Uma batelada pode ser considerada conforme se as variáveis de processo apresentarem variações toleráveis em relação às suas trajetórias médias.

A principal diferença entre processos em bateladas e processos contínuos reside no fato de que processos em bateladas normalmente não são estacionários (Marsh & Tucker, 1991). Isto significa que a média ou valor alvo das variáveis de processo não é constante. As variáveis de processo descrevem um perfil de variação (trajetória da variável) ao longo do tempo de duração da batelada. Desta forma, a variação normal de uma variável de processo no tempo de duração da batelada é descrita através de sua trajetória padrão. Processos contínuos são estacionários porque as variáveis apresentam pequenas variações em torno de suas médias fixas quando apenas causas comuns de variabilidade estão atuando.

O principal objetivo do controle de processos em bateladas é assegurar a mínima variabilidade entre as bateladas, gerando um produto final dentro dos padrões de qualidade requeridos. Para tanto, necessita-se garantir que a trajetória média de cada variável de

processo não seja significativamente alterada. A presença de causa especial de variação no processo ocasiona variabilidade excessiva entre bateladas; como consequência, pode haver alteração da trajetória média de uma ou mais variáveis do processo. O resultado pode ser a obtenção de um produto final fora dos padrões de qualidade.

A variabilidade entre bateladas pode ser consequência de uma combinação inadequada de matérias-primas na etapa de alimentação da batelada, presença de impurezas no processamento e desvios das variáveis de processo em relação as suas trajetórias médias (Nomikos & MacGregor, 1995).

O método aqui proposto para o monitoramento *on line* de processos em batelada constitui-se das seguintes etapas:

1. Identificação das variáveis de interesse no processo em estudo.
2. Elaboração de uma distribuição de referência para os dados do processo. Isto é realizado a partir de informações históricas sobre dados (variáveis de processo e de qualidade) de bateladas e da aplicação da Análise de Componentes Principais *Multway* (ACPM) nos dados selecionados.
3. Elaboração da carta de controle multivariada baseada em componentes principais.
4. Procedimento de monitoramento *on line* de novas bateladas.
5. Procedimentos para diagnóstico das variáveis que ocasionaram causa especial, no caso da CCM ou da carta de controle dos resíduos, sinalizarem uma situação fora de controle.

Cada uma das etapas propostas será detalhada na seqüência.

3.1. Identificação das Variáveis de Interesse

Esta etapa compreende a identificação do objeto de estudo. Esta identificação se refere à seleção das variáveis de interesse para o monitoramento e controle do processo produtivo. Compreende a definição de quais as variáveis de processo e quais as variáveis de qualidade do produto são relevantes para o monitoramento. As variáveis de processo, como por exemplo, temperaturas, fluxos, pressão são medidas diretamente no processo produtivo. As variáveis de qualidades são obtidas mediante análises posteriores do produto final. Esta etapa requer conhecimento do processo produtivo quanto à definição das variáveis controláveis determinantes do desempenho do processo e da qualidade do produto final.

A abordagem proposta nesta dissertação utiliza as variáveis de processo, X , para o monitoramento e controle estatístico do processo produtivo. Isto se deve ao fato de que estas variáveis são coletadas diretamente do processo, na ordem de segundos ou minutos, o que permite uma detecção de eventos especiais atuantes no processo bem como sua eliminação de forma rápida e eficaz. Em contrapartida, as variáveis de qualidade do produto final são obtidas após algumas horas do término da batelada, mediante análises laboratoriais, o que não permite uma atuação sobre o processo produtivo. A utilização das variáveis de qualidade do produto final restringe-se a confirmação se o produto final encontra-se dentro dos limites de especificação adotados.

3.2. Elaboração de uma Distribuição de Referência para os Dados de Processo

O desenvolvimento de cartas de controle estatístico multivariadas para processos em batelada baseia-se em dados históricos de bateladas com bom desempenho e que originaram produtos de boa qualidade. A idéia chave é a construção de uma distribuição de referência para os dados de processo que capture a trajetória média das variáveis de processo selecionadas na etapa anterior. Dessa forma, a elaboração de uma distribuição de referência tem como objetivo refletir um padrão de desempenho das variáveis de processo durante a batelada, contra o qual os dados em cada novo instante de uma nova batelada possam ser comparados.

A análise de dados históricos de bateladas do processo em estudo visa selecionar bateladas que tenham ocorrido de forma semelhante. Isto significa que as variáveis monitoradas ao longo destas bateladas seguiram as trajetórias médias esperadas com alguma variabilidade tolerável. Além disso, este estudo permite a obtenção de um maior entendimento com relação às possíveis fontes de variação entre bateladas.

A distribuição de referência deve conter apenas as bateladas selecionadas nos dados históricos de processo que apresentaram bom desempenho, sem causas especiais de variação. A ACPM é utilizada para selecionar as bateladas sujeitas apenas às causas comuns de variação do processo.

As etapas para a elaboração da distribuição de referência são apresentadas a seguir.

3.2.1 Arranjo dos Dados das Bateladas

Os dados das bateladas devem ser dispostos de acordo com a matriz \mathbf{X} mostrada na Figura 2.6 (Nomikos & MacGregor, 1995).

A Figura 3.1 apresenta um detalhamento do arranjo dos dados provenientes das bateladas, proposto por Nomikos & MacGregor (1995).

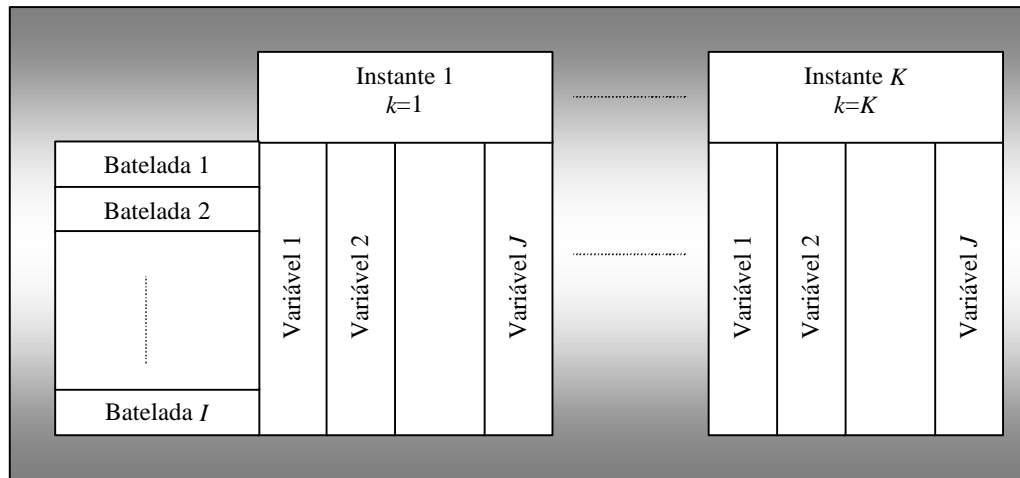


Figura 3.1 - Arranjo dos dados da batelada na matriz \mathbf{X} (detalhamento da Figura 2.6).

Considere uma batelada com $j = 1, 2, \dots, J$ variáveis de processo sendo medidas em $k = 1, 2, \dots, K$ intervalos de tempo durante a batelada. Dados similares são obtidos em muitas outras $i = 1, 2, \dots, I$ bateladas do processo. Todos os dados podem ser resumidos na forma $\underline{\mathbf{X}}$ ($I \times J \times K$) ilustrada na Figura 2.6 (Nomikos & MacGregor, 1995). Na forma $\underline{\mathbf{X}}$, as bateladas são organizadas no eixo vertical, as medidas das variáveis, no eixo horizontal e sua evolução no tempo ocupa a terceira dimensão.

No entanto, para possibilitar a aplicação da ACPM, Nomikos & MacGregor (1995) sugerem desdobrar $\underline{\mathbf{X}}$ colocando suas fatias verticais ($I \times J$) lado a lado, iniciando com a fatia correspondente ao primeiro intervalo de tempo. O resultado é a matriz \mathbf{X} com dimensões ($I \times JK$) na Figura 3.1. Este desdobramento permite analisar a variabilidade entre as bateladas resumindo as informações dos dados com relação as variáveis e suas variações no tempo.

3.2.2 Normalização dos Dados da Matriz \mathbf{X}

A normalização dos dados da matriz \mathbf{X} é efetuada subtraindo a média das colunas da matriz e dividindo cada coluna de variáveis pelo seu desvio padrão. Ao subtrair a média de cada coluna na matriz, subtrai-se a trajetória média de cada variável e remove-se as principais não-linearidades e componentes dinâmicos dos dados. A divisão das colunas da matriz pelo

seu desvio padrão elimina as diferenças de escala entre as variáveis e pondera igualmente todas as variáveis em cada intervalo de tempo.

3.2.3 Aplicação da ACPM nos Dados Normalizados da Matriz \mathbf{X}

O objetivo da aplicação da ACPM nos dados normalizados da matriz \mathbf{X} é a obtenção de um estudo para a variação das trajetórias no tempo de todas as variáveis em todas as bateladas com relação as suas trajetórias médias.

Para aplicar a ACPM na matriz \mathbf{X} , deve-se tratar cada uma das JK colunas desta matriz como sendo uma variável aleatória distinta. No arranjo dos dados em \mathbf{X} tem-se para cada linha da matriz K replicações das mesmas J variáveis. Isto resulta em JK variáveis fortemente correlacionadas.

Os R CPs extraídos da ACPM devem descrever satisfatoriamente a estrutura de correlação dos dados. A informação obtida com estas novas variáveis, os CPs, captura tanto a estrutura de correlação das J variáveis dentro de cada batelada como também a estrutura de autocorrelação de cada variável durante as bateladas. Desta forma, os CPs indicam as principais direções de variabilidade do processo, descrevendo satisfatoriamente as trajetórias médias das suas J variáveis.

A estrutura geral de correlação dos dados da matriz \mathbf{X} é estimada através da matriz de covariâncias amostrais \mathbf{S} , de dimensão $JK \times JK$. Os JK autovalores e respectivos autovetores podem ser calculados a partir da matriz \mathbf{S} , utilizando a equação (2.1).

A ACPM decompõe os dados de \mathbf{X} em uma série de componentes principais. Estes componentes principais são R produtos de vetores de escores (\mathbf{t}) e matrizes de cargas (\mathbf{p}) mais a matriz de resíduos (\mathbf{E}). A equação (2.72) apresenta decomposição dos dados de \mathbf{X} em componentes principais. O número R de componentes principais necessários no modelo de referência pode ser determinado utilizando-se uma das regras apresentadas na seção 2.1.1.1.

O escore t_r , referente ao $r^{\text{ésimo}}$ CP pode ser obtido através da equação (3.1).

$$t_r = \mathbf{X}\mathbf{p}_r^t \quad (3.1)$$

Na equação (3.1), \mathbf{p}_r representa o $r^{\text{ésimo}}$ autovetor obtido de \mathbf{S} e contém as cargas referentes ao $r^{\text{ésimo}}$ CP. Na seção 2.1.1. desta dissertação, foi mostrado que os CPs são variáveis independentes e, portanto, cada autovetor representa a projeção das JK variáveis originais na $r^{\text{ésima}}$ maior direção de variabilidade dos dados, ortogonal em relação às demais

($JK-1$) direções. O correspondente autovalor λ_r , que representa a variância do $r^{\text{ésimo}}$ CP, é o estimador da variabilidade dos dados originais.

A matriz \mathbf{E} , na equação (2.72), contém os resíduos obtidos do modelo que são a diferença entre o observado e o estimado através do modelo proposto. A equação (3.2) apresenta o cálculo da matriz de resíduos.

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} \quad (3.2)$$

Na equação (3.2), a matriz $\hat{\mathbf{X}}$ contém os dados originais acerca das I bateladas amostradas, constituintes do modelo de distribuição de referência elaborado, estimados em função dos R primeiros CPs, retidos segundo algum dos critérios apresentados na seção 2.1.1.1. A matriz \mathbf{X} contém os dados amostrados obtidos no processo.

O somatório do quadrado dos resíduos em cada batelada para o intervalo de tempo k , Q_i é dado na equação (2.77). Os limites de controle para o carta de resíduos são dados na equação (2.80).

3.2.4 Seleção de Bateladas no Conjunto de Dados Históricos

A análise da carta de resíduos, Q_i , e da carta de Hotelling deve ser efetuada para diferenciar as bateladas boas das ruins nos dados históricos. As bateladas ruins, ou que geraram produtos fora dos padrões de qualidade desejados, devem ser retiradas da base de dados históricos. A ACPM aplicada à base de dados históricos do processo contendo apenas bateladas boas constitui o modelo contendo a distribuição de referência dos dados.

Na carta de resíduos, são plotados os valores de Q , calculados segundo a equação (2.77), para cada batelada da distribuição de referência. Se alguma determinada batelada i , apresentar valor de Q_i acima dos limites de controle estabelecidos, dados na equação (2.80) (geralmente com limites de confiança de 95% ou 99%), significa que esta batelada apresentou resíduos elevados, devendo ser eliminada da distribuição de referência.

3.3. Elaboração da CCM de Hotelling Baseada em Componentes Principais

Além das cartas de escores e de resíduos deve ser analisada a estatística de Hotelling para as bateladas constituintes da distribuição de referência. Isto é efetuado através do cálculo da estatística T^2 , dado na equação (2.73). A carta contendo os valores da estatística T^2 para

cada batelada fornece a distância de *Mahalanobis* no espaço reduzido (definido pelos R CP do modelo de ACPM) entre a posição de uma batelada (definida pelos escores t) e a origem.

Os valores da estatística T^2 , referentes às bateladas preliminares utilizadas na distribuição de referência, devem ser comparados a limites de controle, para excluir da distribuição de referência as bateladas que tenham apresentado variação além do tolerado. Na fase I do controle estatístico (ver seção 2.3.1), quando são selecionadas m amostras preliminares de tamanho unitário representando o processo sob controle, o LCS para a estatística T^2 utiliza a distribuição *Beta*, conforme a equação 2.31(Tracy *et al.*,1992).

Os valores críticos da variável beta, B , em nível de significância α , podem ser encontrados a partir dos valores críticos da distribuição F utilizando-se a relação dada na equação (2.74).

3.4. Procedimento de Monitoramento *on line* de Novas Bateladas

No monitoramento *on line*, a evolução de uma nova batelada em tempo real é monitorada no espaço reduzido definido pelos CP do modelo ACPM elaborado. A matriz de cargas \mathbf{P} contém toda informação estrutural de como as medidas das variáveis poderiam desviar-se de seus valores médios em cada intervalo de tempo. A predição dos escores t para cada um dos R CP e dos resíduos \mathbf{E} para uma nova batelada $\mathbf{X}_{nova}(K \times J)$, desdobrada em $\mathbf{X}_{nova}^t(1 \times JK)$ são dadas na equação (2.75) e (2.76), respectivamente (Nomikos & MacGregor, 1995). Nas equações (2.75) e (2.76), a matriz \mathbf{X}_{nova} contém as medidas das variáveis, após normalização e t_r é o escore correspondente ao $r^{ésimo}$ CP.

Para uma nova batelada, o vetor de dados \mathbf{x}_{novo} não está completo até o término da batelada. Quando a batelada está no $k^{ésimo}$ intervalo de tempo, o vetor \mathbf{x}_{novo} contém apenas as primeiras k linhas completas, faltando os dados correspondentes aos demais intervalos até o término da batelada. Para aplicar a ACPM, todavia, a matriz de dados deve estar completa. A etapa seguinte da metodologia aborda este assunto.

3.4.1 Preenchimento dos Dados que Faltam no vetor \mathbf{x}_{novo} entre o Tempo Corrente k e o Término da Batelada.

O objetivo nesta etapa é preencher os dados futuros no vetor de dados da batelada monitorada de forma que os valores preditos para os escores t , a cada intervalo de tempo, sejam tão próximos quanto possível dos valores verdadeiros ao final da batelada. Esta etapa do método proposto envolve a utilização de todos os vetores de cargas \mathbf{p}_r ($JK \times 1$) da

distribuição de referência obtida do modelo ACPM e o preenchimento dos dados futuros no vetor $\mathbf{x}_{\text{nov}}^t$ de diferentes maneiras.

Para o preenchimento dos dados no vetor $\mathbf{x}_{\text{nov}}^t$ pode-se utilizar uma das três abordagens sugeridas por Nomikos & MacGregor (1995), apresentadas na seção 2.4.1. desta dissertação. A escolha de qual abordagem utilizar depende das características do processo a ser monitorado. Se o processo em batelada não apresenta distúrbios persistentes ou variáveis com descontinuidades em suas trajetórias as melhores abordagens são a primeira ou a terceira; se o processo apresenta distúrbios persistentes a segunda abordagem é a mais indicada (Nomikos & MacGregor, 1995).

3.4.2 Cálculos dos Escores e da Estatística de Hotelling para cada Instante K

No monitoramento *on line* de uma nova batelada, calculam-se os escores e a estatística de Hotelling e os resíduos para cada instante k . A predição dos escores t para cada um dos R CP e dos resíduos E para uma nova batelada $\mathbf{X}_{\text{nova}}(K \times J)$, desdobrada no vetor $\mathbf{x}_{\text{nov}}^t(1 \times JK)$ são dadas na equação (2.75) e (2.76), respectivamente.

A estatística de Hotelling para um novo vetor de escores independentes t é dada na equação (2.84). Os comprimentos dos eixos da elipsóide de confiança na direção do $r^{\text{ésimo}}$ CP são dados na equação (2.85).

Uma maneira alternativa de efetuar os cálculos acima é apresentada a seguir.

Para o caso multivariado, a estatística de Hotelling pode ser decomposta em função dos R CPs retidos no modelo ACPM elaborado. O cálculo da estatística de Hotelling pode ser efetuado de acordo com a equação (3.3).

$$T_{r,k}^2 = \sum_{r=1}^R \frac{t_{r,k}^2}{\sqrt{I_r}} = \sum_{r=1}^R y_{r,k}^2 = y_{1,k}^2 + \dots + y_{r,k}^2 \quad (3.3)$$

Na equação (3.3), $t_{r,k}$ é o escore devido à variável CP_r , para $r=1,2,\dots,R$, referente à $k^{\text{ésima}}$ amostra e $y_{r,k}$ representa a padronização do escore $t_{r,k}$ dada na equação 3.4.

$$y_{r,k} = \frac{t_{r,k}}{\sqrt{I_r}} \quad (3.4)$$

Os limites de controle para esta estatística de Hotelling são dados na equação (3.5).

$$\text{LSC} = \frac{R(I+1)(I-1)}{I(I-R)} F_{\alpha, R, I-R}$$

$$\text{LIC} = 0. \tag{3.5}$$

Os limites de controle são estabelecidos a partir da suposição de que as variáveis (os R CPs retidos) tenham distribuição Normal R -variada. Esta suposição baseia-se no fato de que o valor $T_{r,k}^2$ é obtido somando os devidos escores $t_{r,k}$ referentes aos R CPs utilizados. Utilizando o teorema do limite central e sabendo-se que cada CP representa uma combinação linear das J variáveis originais conclui-se que cada CP apresenta uma distribuição aproximadamente normal.

O valor $T_{r,k}^2$ para uma observação \mathbf{x}_k^t , referente a uma batelada futura, deve ser comparado aos limites de controle apresentados na expressão (3.5). Um valor elevado de $T_{r,k}^2$ indica que a batelada em questão apresentou uma variabilidade acima do tolerável. Isto evidencia a presença de causas especiais atuando no processo, isto é, algumas variáveis devem ter apresentado trajetórias com desvios em relação às suas trajetórias médias.

Pode-se construir uma carta de Shewhart para monitorar a média de cada componente padronizado y_r . A média de cada escore padronizado será monitorada através da variável $y_{r,k}$, que possui média zero e desvio padrão unitário. Esta variável segue uma distribuição de Student com parâmetro $m-1$ (onde m é o número de amostras). Os limites de controle para monitorar a média de cada componente y_r , na carta de Shewhart, são dados na equação (3.6).

$$\text{LSC} = t_{m-1, \alpha/2} ,$$

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

$$\text{LIC} = -t_{m-1, \alpha/2} . \tag{3.6}$$

onde $t_{m-1, \alpha/2}$ é o valor da estatística de Student com $m-1$ graus de liberdade e nível de significância α .

A cada nova amostra m obtida do processo, os respectivos escores $y_{r,k}$ calculados serão comparados aos limites de controle na equação (3.6).

3.4.3 Verificar se uma Nova Batelada Apresenta Desvios do Modelo ACPM Elaborado para o Processo

A verificação de algum desvio do modelo ACPM elaborado, no decorrer de uma nova batelada, é realizada através da análise das cartas de escores e das cartas dos resíduos. Quando os valores dos escores $t_{r,k}$ estão fora dos limites de controle a magnitude da variação durante a nova batelada é grande. Quando os resíduos $E(k,j)$ são grandes, o modelo não pode ser mais considerado válido, devido a ocorrência de um novo evento, que não está contemplado na base de dados de referência. Os dados da nova batelada não estão projetados adequadamente no espaço reduzido. Com isso, os resíduos capturam alguma variabilidade que não está suficientemente descrita no modelo de referência (Nomikos & MacGregor, 1994).

No monitoramento *on-line* a melhor maneira de rastrear se algo anormal ocorreu em um instante particular de tempo é através dos resíduos. O mais adequado é a utilização do quadrado dos erros de predição (QEP), calculado de acordo com a equação (2.81) (Nomikos & MacGregor, 1994). O QEP refere-se ao somatório do quadrado dos erros diretamente relacionados com as últimas medidas do processo, no tempo k (ver seção 2.4.1). Os limites de controle para o QEP em um nível de significância α para o intervalo de tempo k são dados na equação (2.86).

O monitoramento de um processo em bateladas é realizado através de amostras \mathbf{x}_k^t de novas bateladas. A partir destas amostras, aplica-se a ACPM e obtêm-se os escores $t_{r,k}$ (equação 3.1), os escores padronizados $y_{r,k}$ (equação 3.4), o valor para $T_{r,k}^2$ (equação 3.3) e o resíduo QEP_k (equação 2.81). Estes valores são comparados com os seus respectivos limites de controle.

Na CCM de Hotelling, com os valores da estatística $T_{r,k}^2$, analisa-se se a variabilidade de alguma das r variáveis em torno das suas médias está de acordo com a distribuição de referência elaborada para o processo. A presença de algum valor elevado de $T_{r,k}^2$ (acima do LCS) indica que alguma variável, em algum instante durante a batelada, apresentou desvio considerável em relação à sua trajetória média. Este fato evidencia a provável presença de causas especiais atuando no processo.

No monitoramento dos resíduos, através da estatística QEP_k , é possível analisar se a estrutura de correlação das JK variáveis dentro da batelada futura está de acordo com a estrutura de correlação capturada no modelo de referência. A presença de algum valor elevado

de QEP_k indica que o modelo de referência elaborado não descreveu satisfatoriamente esta observação. Este fato denota a provável presença de um evento atípico, não incluído no modelo de referência, podendo haver presença de causas especiais no processo.

3.5. Procedimento para Diagnóstico das Variáveis que Ocasionaram Causa Especial

Quando uma das cartas monitoradas (CCM de Hotelling, equação (2.84) ou equação (3.3) e CC dos resíduos QEP , equação (2.81) e nas CC dos escores, equação (2.75)) sinaliza uma causa especial é importante diagnosticar qual(s) variável(s) contribuiu para o sinal fora de controle e depois identificar o que ocorreu no processo para ocasionar o(s) desvio(s) nesta(s) variável(s).

O monitoramento dos gráficos de contribuição juntamente com as CC dos escores normalizados é recomendado. Este procedimento permite a fácil visualização de quais as variáveis que contribuíram para o sinal fora de controle na carta multivariada.

A abordagem para diagnosticar a variável(s) responsável pelo sinal fora de controle sugerida aqui está baseada em Jackson (1980) e foi proposta em Kourti & MacGregor (1996). Para tanto, utilizam-se os cartas de contribuição apresentados na seção 2.4.2 desta dissertação.

A abordagem é constituída por três etapas descritas a seguir.

3.5.1 Identificação de Sinal Fora de Controle na CCM de Hotelling e/ou na Carta QEP

Verificar na CCM de Hotelling se o valor da estatística T^2 para observação \mathbf{x}_k situa-se acima do limite de controle superior (ver equação (3.5)). Da mesma forma, verificar se a CC dos resíduos QEP, apresenta, em algum instante k , valor acima do limite de controle.

3.5.2 Análise das CC de Shewart dos Escores Normalizados dos CP

Verificar nas CC de Shewart para os escores normalizados (os escores normalizados são calculados segundo a equação (3.4)) para observação \mathbf{x}_k : encontrar os escores com valores elevados. Utilizar os limites de controle (equação 3.6) com nível de significância sugerido por *Bonferroni* nas cartas de Shewart para os escores normalizados (se o nível de significância utilizado na carta T^2 foi α , utilizar o nível de significância α/R na carta Shewart para os escores, onde R é o número de CP retidos no modelo).

3.5.3 Cálculo da Contribuição das Variáveis

Calcular a contribuição das variáveis para os escores com valores elevados (equações (2.88) e (2.89)). Investigar as variáveis com valores de contribuição elevados.

O mesmo procedimento deve ser utilizado quando a carta *QEP* (equação (2.81)) sinalizar algum ponto além dos LCs.

Com a utilização desta abordagem não é necessária a interpretação para o significado físico dos escores. Da mesma forma, a construção do gráfico dos escores normalizados e do gráfico de contribuição das variáveis não é necessária uma vez que as etapas 2 e 3 são facilmente programáveis em computadores.

O gráfico de contribuição identifica o grupo das variáveis de processo que não estão consistentes com as condições de operação normal da batelada. Os operadores e engenheiros devem utilizar seus conhecimentos sobre o processo produtivo para encontrar as causas que provocaram as alterações nas variáveis identificadas.

CAPÍTULO 4

4. ESTUDO DE CASO

Este capítulo é destinado à descrição do estudo de caso realizado para esta dissertação. O estudo de caso foi realizado em uma indústria do setor de bebidas, localizada na região metropolitana de Porto Alegre.

O processo definido para elaboração deste estudo de caso é a etapa de fermentação integrante do processo de fabricação de cerveja. A cerveja é fabricada através da fermentação alcoólica dos carboidratos presentes nos grãos de cereais, como a cevada (Lehninger, 1986).

A fabricação da cerveja inicia-se quando a cevada sofre um processo chamado maltagem. Neste processo inicial, as sementes do cereal germinam até a formação de enzimas apropriadas, necessárias à hidrólise dos polissacarídeos da parede celular das sementes, bem como do amido e de outros polissacarídeos da reserva alimentar existente no interior da célula. A germinação é, então, detida por aquecimento controlado, o que impede a semente de continuar a crescer. O produto resultante deste processo inicial é denominado malte, o qual contém enzimas α -amilase e maltase capazes de hidrolisar o amido em maltose, glicose e outros açúcares simples.

A etapa seguinte do processo de fabricação de cerveja, constitui-se na preparação do mosto. O mosto é o meio nutriente necessário à etapa subsequente, denominada fermentação, a ser realizada pelas células de levedura. Para a preparação do mosto, o malte é misturado com água e então macerado. Isto permite que as enzimas formadas durante a preparação do malte exerçam sua atividade sobre os polissacarídeos do cereal e produzam maltose, glicose e

outros açúcares simples, que são solúveis em meio aquoso. O material celular restante é filtrado e o mosto líquido é fervido com lúpulo para aromatiza-lo. Após isso, o mosto deve ser resfriado e aerado.

Para a realização da etapa de fermentação, células de levedura são adicionadas ao mosto. No mosto aeróbico a levedura se reproduz muito rapidamente, empregando a energia obtida pela metabolização de parte dos açúcares existentes no meio. Nesta fase não ocorre a formação de álcool, pois a levedura, tendo muito oxigênio à disposição, oxida o composto chamado de piruvato, proveniente do ciclo do ácido cítrico, em CO_2 e H_2O . O metabolismo aeróbico da levedura permite uma multiplicação muito rápida das células, o que pode ser controlado através de adição de oxigênio. Quando todo o oxigênio dissolvido existente no tanque de fermentação é consumido, as células de levedura, que são facultativas, passam a utilizar anaerobicamente o açúcar existente no mosto. A partir deste ponto a levedura fermenta os açúcares presentes no meio etanol e dióxido de carbono. O processo de fermentação é controlado, em parte pela concentração de etanol que se forma, pelo pH do meio e pela quantidade de açúcar remanescente. Após a interrupção da fermentação, as células são removidas e a cerveja bruta está pronta para ser submetida ao processamento final.

Nas etapas finais de fabricação de cerveja, o controle da espuma, ou “colarinho”, formado por proteínas dissolvidas, é realizado. Normalmente, este controle é feito através do emprego de enzimas proteolíticas provenientes do preparo do malte. Esta etapa é chamada de maturação.

As informações obtidas na indústria, com relação aos dados do processo e produto em estudo, permitiram a aplicação do método proposto para o controle multivariado de processos em bateladas proposto no capítulo 3 desta dissertação.

A seguir será apresentada a aplicação das etapas constituintes do método proposto no capítulo 3 desta dissertação aos dados do estudo de caso.

4.1. Identificação das Variáveis de Interesse

Foram identificadas pelos engenheiros de processos da empresa três variáveis de processos para aplicação da ACPM. As variáveis selecionadas satisfazem as condições de aplicabilidade do método proposto no capítulo 3. As variáveis são coletadas automaticamente no processo e existem dados históricos disponíveis acerca destas variáveis para bateladas já realizadas.

Dispõe-se de dados que descrevem as trajetórias das variáveis selecionadas nesta operação do processo: *Variável A* (X_1) e *Variável B* (X_2), e *Variável C* (X_3) durante 15 bateladas completas. Em cada batelada, foram obtidas medições simultâneas das variáveis, em intervalos fixos de 6 horas, durante o tempo de duração das bateladas (90hs), perfazendo um total de 15 medições consecutivas das variáveis X_1 e X_3 . Para a aplicação da ACPM, consideram-se as medições da variável X_2 a partir do segundo instante de tempo $k=2$, totalizando 14 replicações para esta variável dentro da batelada. Dessa forma, tem-se então 44 variáveis fortemente correlacionadas.

As Figuras 4.1, 4.2 e 4.3 apresentam, respectivamente, as trajetórias das variáveis X_1 e X_2 e X_3 nas 15 bateladas preliminares disponíveis. Na aplicação da ACPM, deve ser considerado em cada instante cada uma das três variáveis de processo envolvidas como sendo uma variável aleatória distinta. A partir da aplicação da ACPM nestas variáveis, captura-se, a partir das trajetórias de cada variável, a estrutura de correlação e autocorrelação dessas variáveis ao longo das bateladas preliminares bem sucedidas, obtendo assim a trajetória média de cada variável.

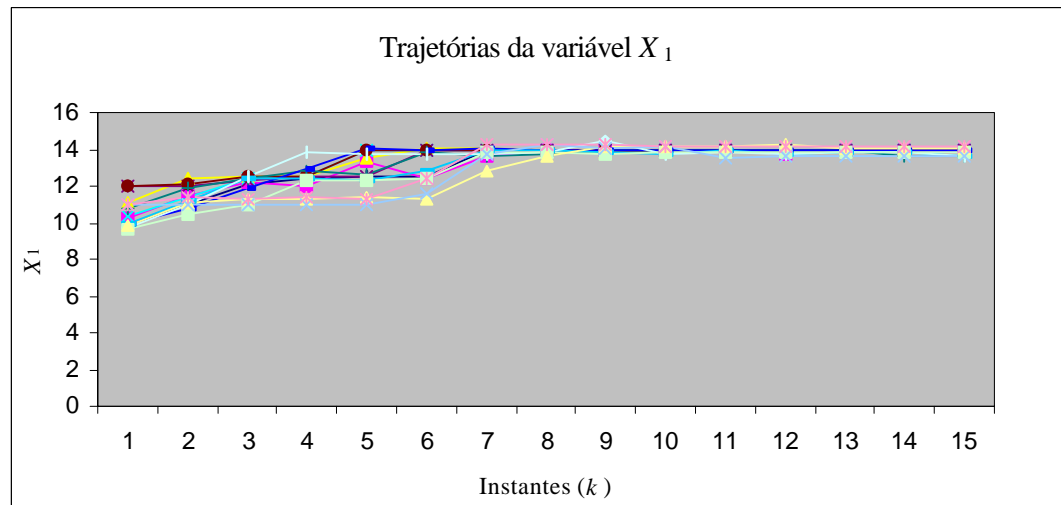


Figura 4.1 - Trajetórias da variável X_1 observadas em 15 instantes, nas 15 bateladas preliminares.

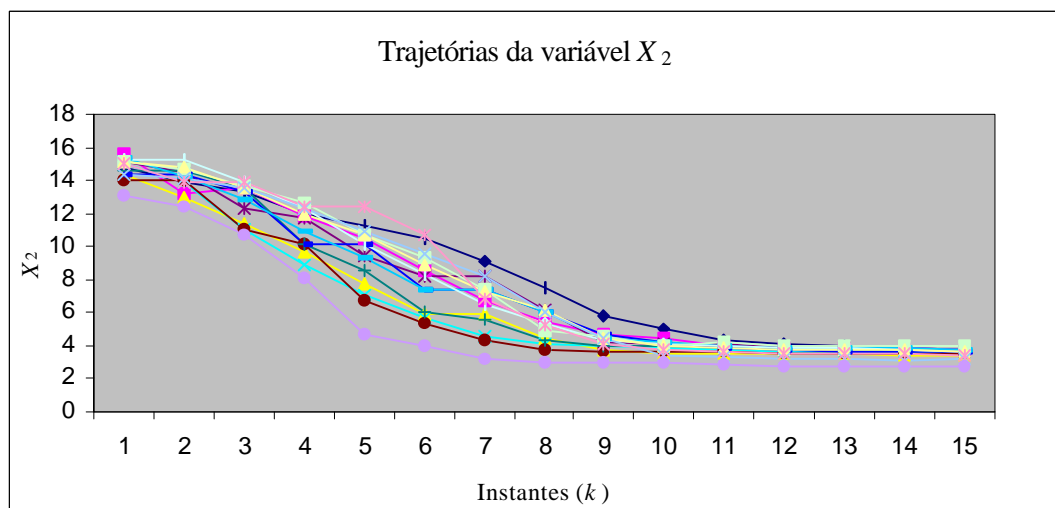


Figura 4.2 - Trajetórias da variável X_2 observadas em 15 instantes, nas 15 bateladas preliminares.

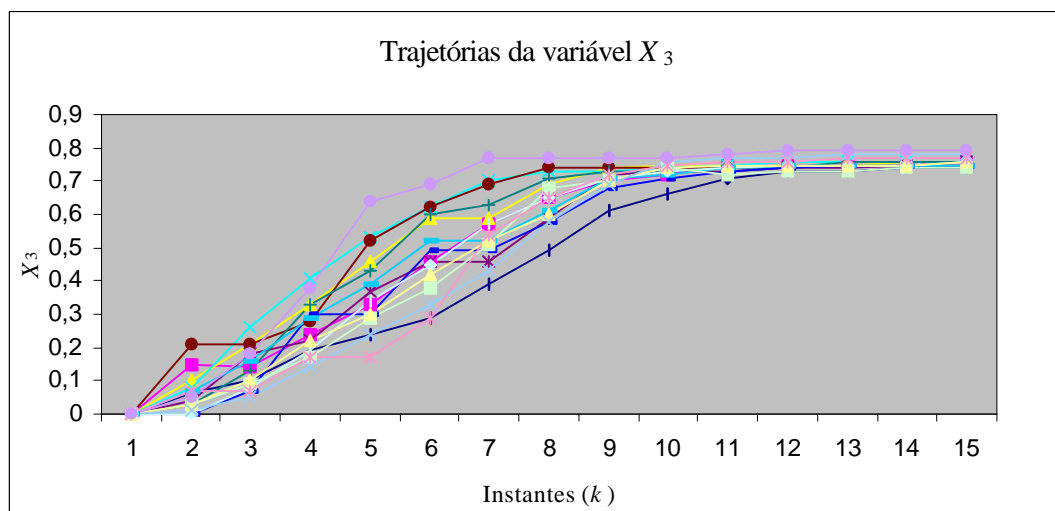


Figura 4.3 - Trajetórias da variável X_3 observadas em 15 instantes, nas 15 bateladas preliminares.

4.2. Elaboração de uma Distribuição de Referência para os dados de Processo

Esta etapa tem como objetivo selecionar as bateladas preliminares que tenham ocorrido de forma semelhante e aplicar a ACPM. A construção do CEP, na primeira fase do controle, deve utilizar a ACPM para elaborar a distribuição de referência para as variáveis de processo em estudo. Dessa forma, selecionam-se as bateladas nas quais as variáveis

apresentam variabilidade em torno das suas trajetórias médias devido apenas a causas comuns de variação.

Os dados resultantes dos cálculos efetuados para elaboração da distribuição de referência para este estudo de caso são apresentados no Anexo 1 desta dissertação.

As etapas para a elaboração da distribuição de referência são apresentadas a seguir.

4.2.1 Arranjo dos Dados das Bateladas

Os dados das bateladas devem ser dispostos de acordo com a matriz \mathbf{X} mostrada na Figura 2.6 (Nomikos & MacGregor, 1995).

A matriz $\underline{\mathbf{X}}$ (ver Figura 2.6), para os dados deste estudo de caso, possui dimensão $I \times J \times K$, tridimensional (15 bateladas $\times 3$ variáveis $\times 15$ instantes), sendo desdobrada na matriz \mathbf{X} , $I \times JK$, bidimensional (15 bateladas $\times 44$ variáveis). No primeiro instante de tempo ($k=1$) são computados os valores das variáveis X_1 e X_2 ; nos demais 14 instantes de tempo, são computados valores para as três variáveis X_1 , X_2 e X_3 (por isso tem-se a dimensão $JK = 44$). Em cada linha da matriz \mathbf{X} tem-se uma observação \mathbf{x}'_k que representa a realização completa das variáveis X_1 , X_2 e X_3 na batelada i ($i = 1, \dots, J$). Os dados para as variáveis de processo na matriz \mathbf{X} são apresentados na Tabela A1 no Anexo 1 desta dissertação.

4.2.2 Normalização dos Dados da Matriz X

A normalização dos dados da matriz \mathbf{X} é efetuada subtraindo a média das colunas da matriz e dividindo cada coluna de variáveis pelo seu desvio padrão. Ao subtrair a média de cada coluna na matriz, subtrai-se a trajetória média de cada variável e remove-se as principais não-linearidades e componentes dinâmicos dos dados. A divisão das colunas da matriz pelo seu desvio padrão elimina as diferenças de unidades entre as variáveis e pondera igualmente todas as variáveis em cada intervalo de tempo.

Os dados normalizados na matriz \mathbf{X} são apresentados na Tabela A2 do Anexo 1 desta dissertação.

4.2.3 Aplicação da ACPM nos Dados Normalizados da Matriz X

Na aplicação da primeira fase do CEP, a ACPM é utilizada para selecionar a distribuição de referência para as variáveis envolvidas. Desta forma, selecionam-se as

bateladas nas quais as variáveis X_1 , X_2 e X_3 apresentam variabilidade em torno das suas trajetórias médias devida apenas a causas comuns de variação.

Na matriz \mathbf{X} , com os dados normalizados, cada linha da matriz contém uma observação \mathbf{x}'_i que representa a realização completa das variáveis X_1 , X_2 e X_3 na batelada i ($i = 1, \dots, I$). Estimou-se a matriz de covariâncias amostrais \mathbf{S} , de dimensão $JK \times JK$ (44×44). Os 44 autovalores e respectivos autovetores foram calculados a partir da equação característica da matriz \mathbf{S} (ver Strang, 1988). A partir disso, obtiveram-se os CP através da equação (2.1).

O número R de componentes principais necessários no modelo de referência foi determinado utilizando-se a regra do *Scree Graph* e a regra da média dos autovalores, apresentadas na seção 2.1.1.1. Selecionou-se, a partir destes critérios, os oito primeiros CPs ($R=8$), considerando-se que estes CPs ofereciam uma caracterização suficiente da estrutura de correlação das 44 variáveis da matriz de dados \mathbf{X} . A Figura 4.4 apresenta o *Scree Graph* gerado a partir dos dados padronizados de \mathbf{X} .

A partir disso, selecionaram-se os oito maiores autovalores (λ_r , $r=1, \dots, 8$) e os autovetores associados (\mathbf{p}_r , $r=1, \dots, 8$). Estes oito primeiros autovalores explicam 94% da variabilidade existente nos dados padronizados da matriz \mathbf{X} . Cada escore, t_r , referente ao $r^{\text{ésimo}}$ CP pôde ser obtido através da equação (3.1). A partir da equação (3.4), obteve-se os escores padronizados $y_{r,k}$ ($r=1, \dots, 8$ e $k=1, \dots, 15$) para as 15 bateladas preliminares. A Tabela 4.1 apresenta a variabilidade total explicada pelos oito CP retidos no modelo ACPM elaborado.

Tabela 4.1 – Variabilidade total explicada pelos oito CP retidos no modelo ACPM elaborado.

Componente Principal	Autovalores	% Variância	% Acumulada da Variância
1	19,49	44,28	44,28
2	8,61	19,56	63,84
3	4,82	10,95	74,79
4	2,77	6,29	81,08
5	1,72	3,91	84,98
6	1,63	3,71	88,70
7	1,23	2,80	91,50
8	1,22	2,76	94,26

A matriz de cargas e os vetores de escores para os CP retidos no modelo elaborado, são apresentados, respectivamente, nas Tabelas A3 e A4 do Anexo 1 desta dissertação.

Os resíduos foram calculados de acordo com a equação (3.2) e o valor do somatório quadrado dos resíduos para cada batelada, Q_i , através da equação (2.77). Os valores de Q_i e da estatística de Hotelling calculados em relação aos CP retidos no modelo são apresentados na Tabela A5 do Anexo 1.

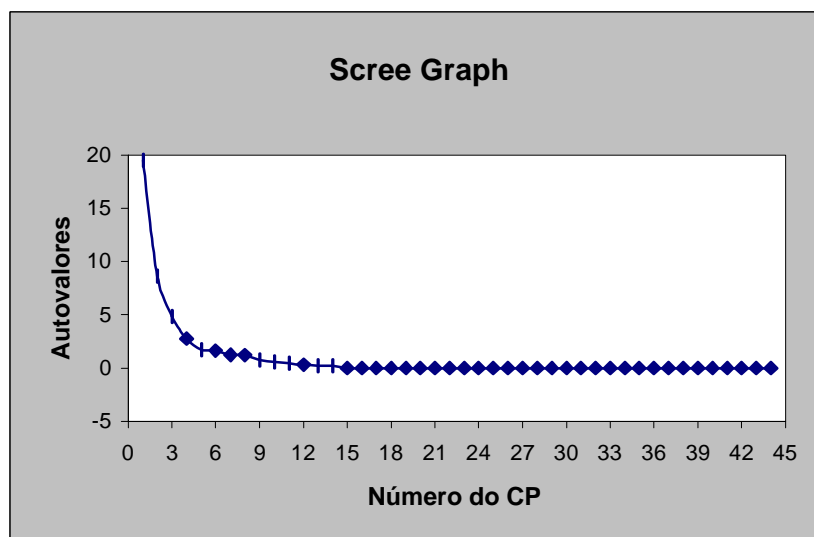


Figura 4.4 - *Scree Graph* para aplicação da ACPM nos dados padronizados da matriz **X**

4.2.4 Seleção de Bateladas no Conjunto de Dados Históricos

A análise da carta de resíduos, Q_i , e da carta de Hotelling deve ser efetuada para diferenciar as bateladas boas das ruins nos dados históricos. As bateladas ruins, ou que geraram produtos fora dos padrões de qualidade desejados, devem ser retiradas da base de dados históricos.

A carta de resíduos é construído comparando-se os valores Q_i , de cada batelada i , calculados pela equação (2.77), para cada intervalo de tempo k , com os limites de controle dados na equação (2.80).

Na Figura 4.5 é apresentado o gráfico dos resíduos para as 15 bateladas preliminares utilizadas na distribuição de referência deste estudo de caso. Analisando a Figura 4.5, observa-se que nenhuma das 15 bateladas preliminares apresentou valores de Q_i acima do LCS. Este fato indica que o modelo construído em função dos oito primeiros CPs oferece uma representação suficiente do comportamento das variáveis envolvidas dentro das 15 bateladas.

Caso houvesse resíduo significativo, algum valor de Q_i , para alguma batelada i , estaria acima do LCS na CC dos resíduos. Isto indicaria que o modelo construído em função dos oito primeiros CPs não ofereceu uma representação suficiente do comportamento das variáveis envolvidas dentro desta batelada i . Dessa forma, existiriam evidências de que algum evento especial ocorreu no processo durante essa batelada ou de que as variáveis X_1 , X_2 e/ou X_3 apresentaram variabilidade acentuada num determinado instante no processo.

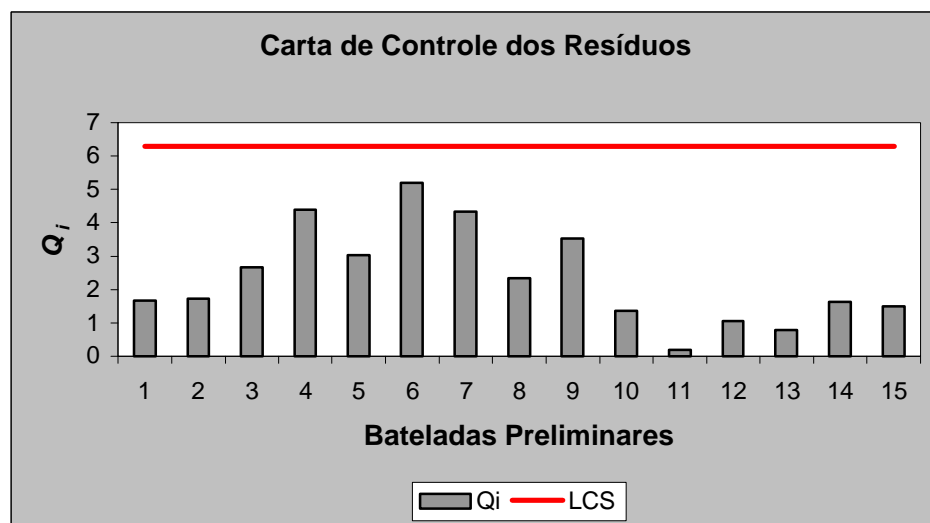


Figura 4.5 – Carta de Controle para os resíduos, Q_i para as 15 bateladas preliminares da distribuição de referência.

Além da análise da CC de resíduos, para garantir que não existem evidências de que algum evento incomum ocorreu no processo durante essas bateladas ou de que as variáveis X_1 , X_2 e/ou X_3 apresentaram variabilidade acentuada num determinado instante, deve ser avaliada a CCM de Hotelling para as bateladas do modelo de referência proposto.

A elaboração da CCM baseada em componentes principais para a distribuição de referência elaborada com os dados das 15 bateladas preliminares deste estudo de caso é apresentada a seguir.

4.3. Elaboração da CCM de Hotelling Baseada em Componentes Principais

A CCM baseada em componentes principais é elaborada utilizando como variáveis de entrada os escores t dos oito componentes principais retidos no modelo de referência elaborado.

Calculado os valores da estatística de Hotelling, através da equação (2.73) e o LCS, calculado através da equação (2.74), para as 15 bateladas preliminares, obtém-se então a carta de Hotelling, em relação aos oito CP retidos, ilustrada na Figura 4.6

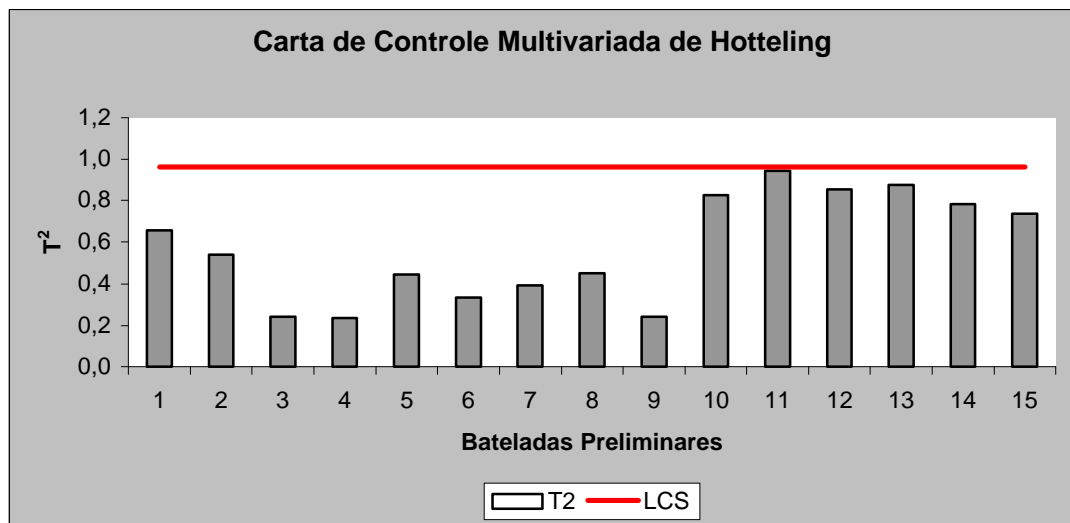


Figura 4.6 – CCM de Hotelling para as 15 bateladas preliminares, em relação aos oito CP retidos no modelo ACPM elaborado.

Analisando a Figura 4.6, observa-se que nenhuma das 15 bateladas preliminares apresentou escores acima do LCS. Isto indica que a partir da projeção destas bateladas no plano definido pelos oito CPs, não existem evidências de que a variação ocorrida nas variáveis de processo, nestas bateladas, esteve acima do tolerado; ou seja, apenas causas comuns estão presentes no processo. Isto significa que, nestas bateladas, as trajetórias das variáveis X_1 , X_2 e/ou X_3 não tiveram variabilidade acentuada, não apresentando desvios significativos em relação as suas respectivas trajetórias, conforme dadas pela distribuição de referência.

4.4. Procedimento de Monitoramento *on line* de Novas Bateladas

No monitoramento *on line*, a evolução de uma nova batelada em tempo real é monitorada no espaço reduzido definido pelos CP do modelo ACPM elaborado. Os dados resultantes dos cálculos efetuados para o procedimento de monitoramento *on line* de novas bateladas neste estudo de caso são apresentados no Anexo 2 desta dissertação.

Considere, por exemplo, o início de uma nova batelada. O modelo ACPM, construído em função de 15 bateladas preliminares, será então utilizado para monitorar a variabilidade apresentada nesta batelada. Como existem 15 bateladas na distribuição de referência e assume-se que a nova batelada, denominada por batelada 16, é a primeira batelada a ser monitorada na fase II do controle estatístico. As Tabela A1 e A2 do Anexo 2 apresentam os dados para as variáveis de processo coletados para as 15 bateladas da distribuição de referência e para a batelada 16 nos 15 instantes de tempo.

Para o monitoramento *off-line*, bastaria projetar a observação \mathbf{X}_{16}^t no plano definido pelos oito primeiros CPs após o término da batelada 16 e verificar, através da carta de Hotelling e a carta Q se a referente batelada apresentou variabilidade e resíduo além do tolerado (isto é, $T_{8,16}^2 > LSC$ e $Q_{16} > LSC$).

No monitoramento *on line* de uma nova batelada o vetor de dados \mathbf{x}_{16}^t não está completo antes do término da batelada. Quando a batelada está no $k^{\text{ésimo}}$ intervalo de tempo, faltam os dados correspondentes aos demais intervalos até o término da batelada no vetor \mathbf{x}_{16}^t . Para aplicar a ACPM o vetor de dados deve estar completo. Para que o monitoramento *on line* da nova batelada seja possível, deve-se preencher os dados futuros no vetor de forma que os valores preditos para os escores t , a cada intervalo de tempo sejam tão próximos quanto possível dos valores verdadeiros ao final da batelada.

A forma escolhida para o preenchimento dos dados faltantes em \mathbf{x}_{16}^t a cada instante é apresentada na seção a seguir.

4.4.1 Preenchimento dos Dados que Faltam no vetor \mathbf{x}_{16}^t entre o Tempo Corrente k e o Término da Batelada.

Pode ser utilizada uma das três abordagens sugeridas por Nomikos & MacGregor (1995), apresentadas na seção 2.4.1. desta dissertação.

Neste estudo de caso a batelada 16 será projetada no plano definido pelos oito primeiros CPs, realizando-se o preenchimento gradativo da referente observação \mathbf{x}_{16}^t através do segundo critério apresentado na seção 2.4.1. Completa-se então o vetor \mathbf{x}_{16}^t utilizando os resultados padronizados de X_1 , X_2 e X_3 obtidos no último instante transcorrido na batelada.

O vetor \mathbf{x}_{16}^t , para o instante $k=9$, é apresentado na Tabela A3 do Anexo 2 desta dissertação. Observa-se que a partir do instante $k=10$ até o instante final $k=15$, o vetor é preenchido com os resultados padronizados de X_1 , X_2 e X_3 obtidos no último instante $k=9$. A Tabela A4 do Anexo 2 desta dissertação apresenta o vetor \mathbf{x}_{16}^t completo com todos os dados para a batelada 16 concluída.

Os cálculos dos escores, da estatística de Hotelling e dos resíduos são efetuados a cada instante k , considerando-se o vetor \mathbf{x}_{16}^t totalmente preenchido de acordo com o critério escolhido.

4.4.2 Cálculos dos Escores e da Estatística de Hotelling para cada Instante k

A predição dos escores t para cada um dos 8 CP e dos resíduos \mathbf{E} para uma nova batelada $\mathbf{X}_{16}(15 \times 3)$, desdobrada em $\mathbf{x}_{16}^t(1 \times 44)$, são dadas na equação (2.75) e (2.76), respectivamente. A variável $y_{r,k}$, representa a padronização do escore $t_{r,k}$, é dada na equação (3.4). Os limites de controle para monitorar a média de cada componente y_r , na carta de Shewhart, são dados na equação (3.6).

A Tabela A5 do Anexo 2 desta dissertação apresenta os valores dos escores padronizados calculados para cada instante k e o cálculo dos limites de controle para a carta de Shewhart destes escores.

A estatística de Hotelling para um novo vetor de escores independentes \mathbf{t} é dada na equação (3.3). Os limites de controle para esta estatística de Hotelling são dados na equação (3.5). A Tabela A6 do Anexo 2 desta dissertação apresenta os valores da estatística de Hotelling para cada batelada e o cálculo do limites de controle para a CCM.

A Figura 4.7 apresenta os valores parciais da estatística de Hotelling obtidos a cada instante dentro da batelada 16 (equação 3.3) e o LSC da CCM (equação 3.5).

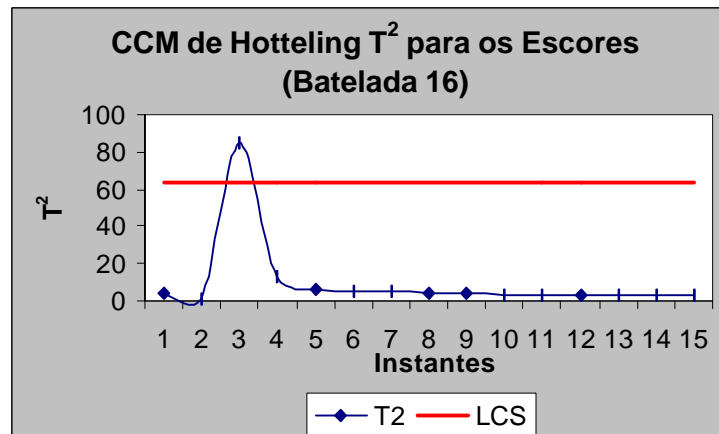


Figura 4.7 – CCM de Hotelling, em relação aos escores dos CP, para a batelada 16.

Na Figura 4.7 observa-se que, no instante $k=3$, a batelada apresenta variabilidade excessiva (isto é, $T_{3,16}^2 > LCS$). Este fato indica que, neste instante, as trajetórias das variáveis X_1 , X_2 e/ou X_3 apresentaram desvios significativos em relação as suas respectivas trajetórias médias.

As Figuras 4.8 a 4.15 apresentam respectivamente as cartas de Shewhart para os oito CP retidos. Os LC para estas cartas foram obtidos com nível de significância de 5% ($\alpha=5\%$). Analisando as figuras, observa-se que, no instante $k=3$ do monitoramento *on line*, os escores padronizados referentes ao segundo, terceiro, sexto, sétimo e oitavo CP apresentam valores superiores aos seus respectivos LCS. Isto indica que o valor parcial significativo encontrado na carta de Hotelling é reflexo dos escores elevados encontrados nestes CP. Dessa forma, observa-se que no instante $k=3$, a variabilidade excessiva verificada na batelada 16 foi provocada pelas fontes de variabilidade do processo, evidenciadas por estes CP. A partir disso, devem ser analisados estes CP na busca de um diagnóstico *on-line* para o sinal de processo fora de controle, gerado neste instante pela carta de Hotelling.

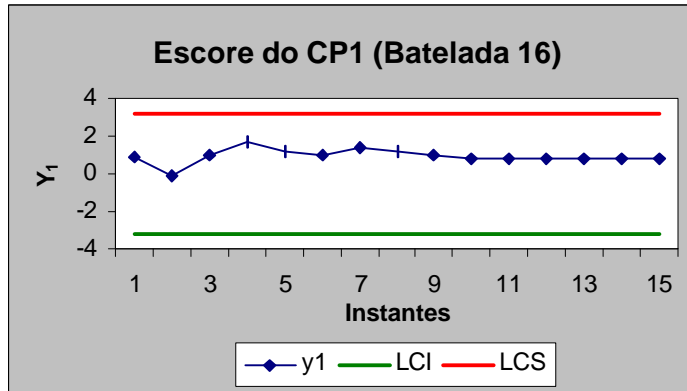


Figura 4.8 – Carta de Shewhart para o CP1 referente a Batelada 16.

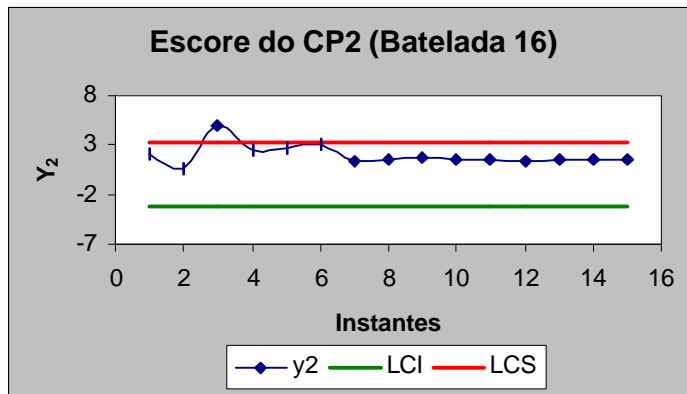


Figura 4.9 – Carta de Shewhart para o CP2 referente a Batelada 16.

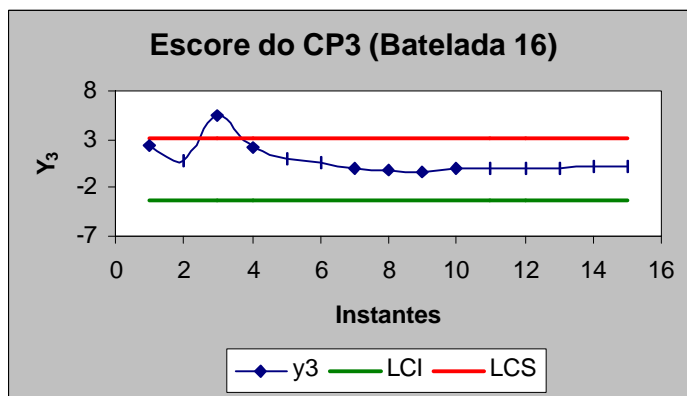


Figura 4.10 – Carta de Shewhart para o CP3 referente a Batelada 16.

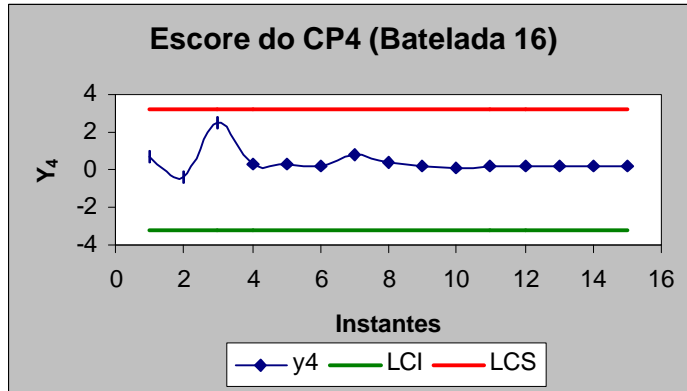


Figura 4.11 – Carta de Shewhart para o CP4 referente a Batelada 16.

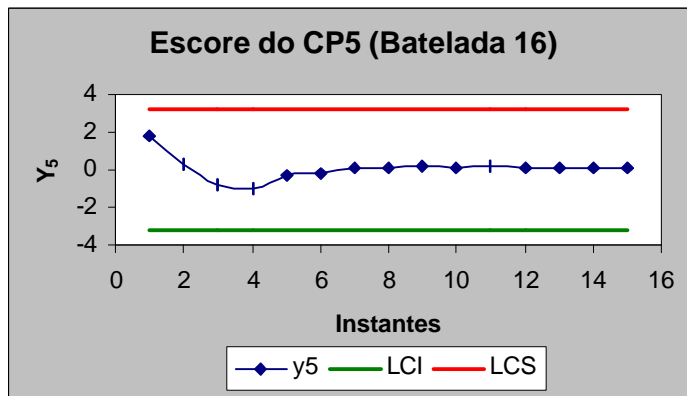


Figura 4.12 – Carta de Shewhart para o CP5 referente a Batelada 16.

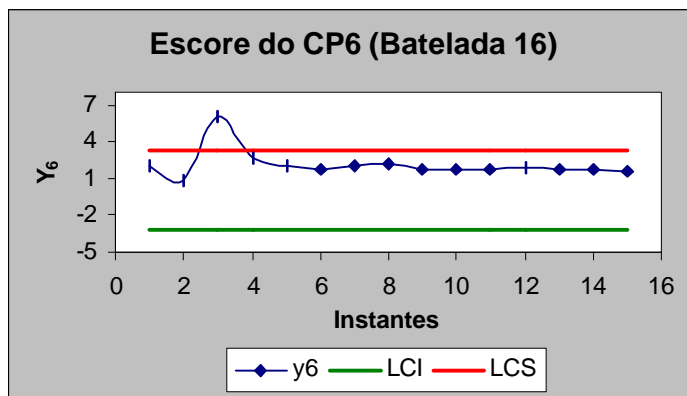


Figura 4.13 – Carta de Shewhart para o CP6 referente a Batelada 16.

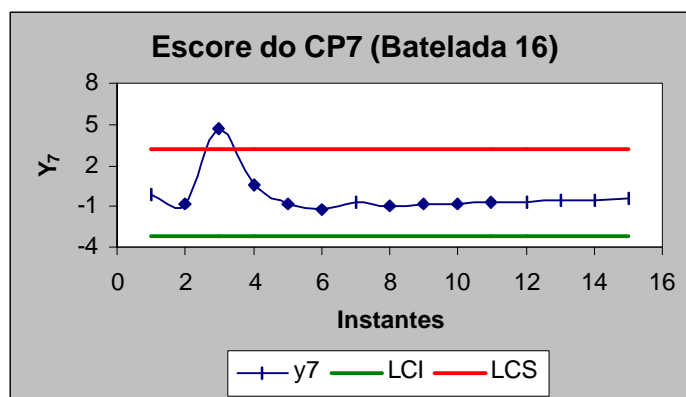


Figura 4.14 – Carta de Shewhart para o CP7 referente a Batelada 16.

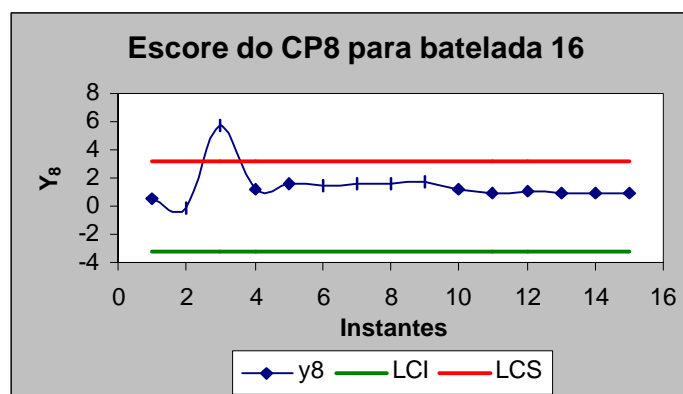


Figura 4.15 – Carta de Shewhart para o CP8 referente a Batelada 16.

4.4.3 Verificar se uma Nova Batelada apresenta desvios do Modelo ACPM Elaborado para o Processo

A verificação de algum desvio do modelo ACPM elaborado, no decorrer de uma nova batelada, é realizada através da análise das CC de Shewhart para os escores dos CP retidos no modelo e do gráfico dos resíduos (*QEP*).

Os valores calculados para *QEP* (equações (2.81) e (2.86)), baseados no modelo de distribuição de referência elaborado com as 15 bateladas preliminares, para o monitoramento *on line* da batelada 16, são também apresentados na Tabela A5 do Anexo 2.

A Figura 4.16 apresenta a carta QEP , a cada instante k no monitoramento *on line* para batelada 16, com LCS obtido com nível de significância de 5% ($\alpha=5\%$). A Figura 4.17 apresenta uma ampliação da Figura 4.16, permitindo uma melhor visualização dos dados na carta QEP .

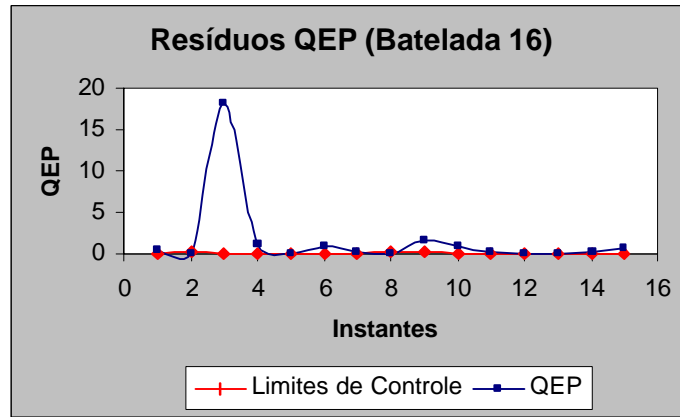


Figura 4.16 – Carta QEP , monitoramento *on line*, para a batelada 16.

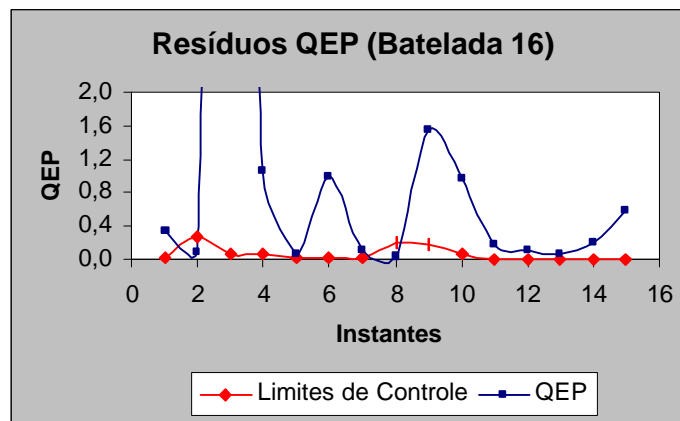


Figura 4.17 – Ampliação da carta QEP , monitoramento *on line*, para a batelada 16.

As Figuras 4.16 e 4.17 apresentam os resíduos parciais $QEP_{k,16}$ obtidos a cada instante dentro da batelada 16 e os respectivos valores para LCS da carta de controle. Observa-se que já no primeiro instante $k=1$ a batelada apresenta um resíduo significativo (valor de $QEP_{1,16} > LCS$). No entanto, é a partir do instante $k=3$ que a carta efetivamente começa a sinalizar resíduos com valores significativamente elevados ($QEP_{3,16}=18,28$ e $LCS=0,077$). Conclui-se que, a partir deste instante, o modelo ACPM de referência construído em função dos oito primeiros CP não está descrevendo suficientemente o comportamento das variáveis X_1 , X_2 e X_3 , dentro da batelada 16.

4.5. Procedimento para Diagnóstico das Variáveis que Ocasionaram Causa Especial

A finalidade do procedimento de diagnóstico é identificar quais variáveis contribuíram mais efetivamente para a ocorrência dos altos valores nas cartas de controle. A partir disso, pode-se 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 posteriores continuem apresentando variabilidade excessiva.

4.5.1 Identificação de Sinal Fora de Controle na CCM de Hotelling e/ou na Carta QEP

Neste estudo de caso, a CCM de Hotelling para batelada 16 (Figura 4.7), baseada nos oito CP do modelo de referência elaborado, apresentou sinal fora de controle ($T_{3,16}^2 > LSC$) no instante $k=3$.

A carta *QEP*, apresentada nas Figuras 4.16 e 4.17, também apresenta sinais fora de controle para o monitoramento *on line* da batelada 16. Da mesma forma, deve-se realizar o diagnóstico de quais as variáveis contribuíram para o sinal fora de controle e depois identificar o que ocorreu no processo para ocasionar o(s) desvio(s) nesta(s) variável(s).

4.5.2 Análise das CC de Shewhart dos Escores Normalizados dos CP

Neste estudo de caso observa-se que, no instante $k=3$ do monitoramento *on line*, os escores padronizados referentes ao segundo, terceiro, sexto, sétimo e oitavo CP apresentam valores superiores aos seus respectivos LCS, nas suas respectivas cartas de Shewhart (Figuras 4.9, 4.10, 4.13, 4.14 e 4.15). Isto indica que o sinal fora de controle no instante $k=3$, evidenciado na CCM de Hotelling baseada nos CP para batelada 16, é confirmado nas cartas de Shewhart dos escores destes CP. A partir disso, devem ser analisados os CP de forma a diagnosticar qual(s) variável(s) contribuiu para o sinal fora de controle e depois identificar o que ocorreu no processo para ocasionar o(s) desvio(s) nesta(s) variável(s).

4.5.3 Cálculo da Contribuição das Variáveis

As Figuras 4.18, 4.19 e 4.20 apresentam a contribuição das variáveis X_1 , X_2 e X_3 para os valores elevados dos escores y_2 , y_3 , y_6 , y_7 e y_8 no instante $k=3$ e para os valores elevados de QEP nos instantes $k=1$ e $k=3$, respectivamente, no monitoramento *on line* da batelada 16.

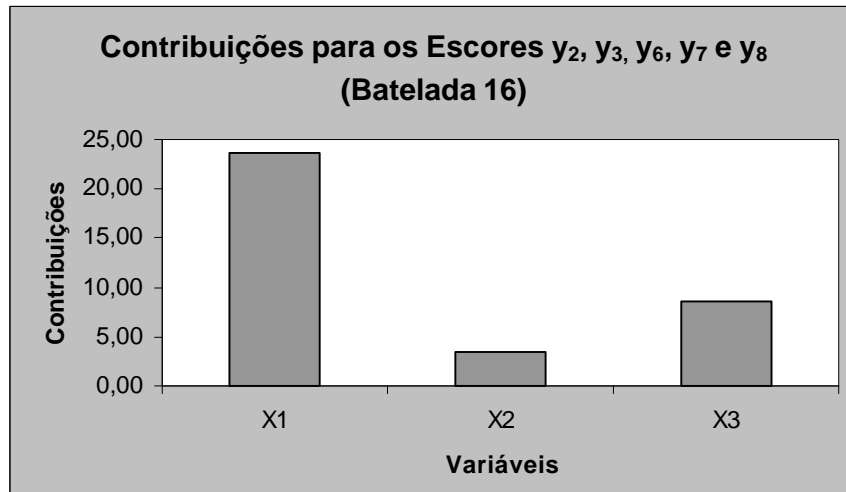


Figura 4.18 – Gráfico de Contribuição das variáveis X_1 , X_2 e X_3 para os valores elevados dos escores y_2 , y_3 , y_6 , y_7 e y_8 no instante $k=3$ do monitoramento *on line* da batelada 16.

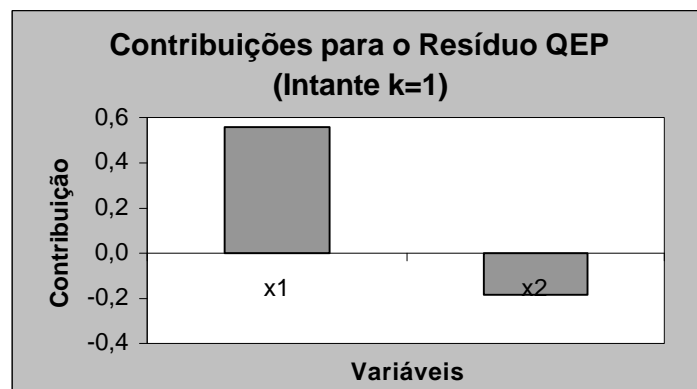


Figura 4.19 – Gráfico de Contribuição das variáveis X_1 , X_2 para o valor elevado de QEP no instante $k=1$ do monitoramento *on line* da batelada 16.

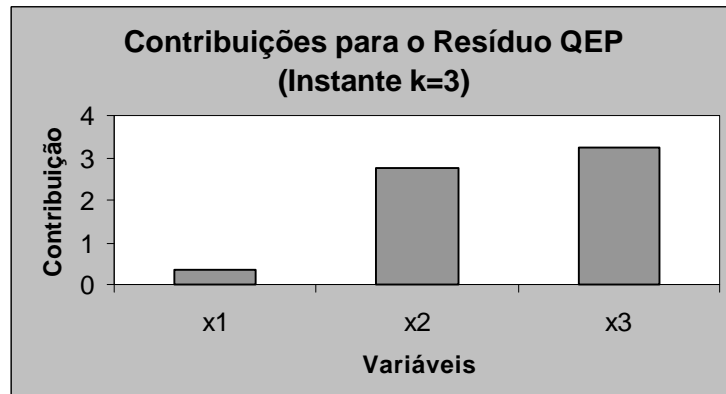


Figura 4.20 – Gráfico de Contribuição das variáveis X_1 , X_2 e X_3 para o valor elevado de QEP no instante $k=3$ do monitoramento *on line* da batelada 16.

O gráfico de contribuição identifica o grupo de variáveis de processo que não estão consistentes com as condições de operação normal da batelada. Analisando-se a Figura 4.18 observa-se que todas as variáveis X_1 , X_2 e X_3 apresentam valores de contribuição elevados sendo candidatas a investigação detalhada. Analisando-se a Tabela A1 do Anexo 2, a qual contém os dados completos da matriz \mathbf{X}_{16} , verifica-se valores elevados destas variáveis na batelada 16, no instante $k=3$, quando comparados aos respectivos valores das outras 15 bateladas preliminares da distribuição de referência ($X_{16,1,3}=15$ quando $\bar{X}_{i,1,3}=11,94$ e $S_{i,1,3}=0,59$; $X_{16,2,3}=16$ quando $\bar{X}_{i,2,3}=12,79$ e $S_{i,2,3}=1,15$ e $X_{16,3,3}=0,40$ quando $\bar{X}_{i,3,3}=0,14$ e $S_{i,3,3}=0,06$).

Na Figura 4.19, foram calculados apenas os valores das contribuições das variáveis X_1 e X_2 , uma vez que a variável X_3 começa a ser coletada no processo a partir do instante $k=1$. Nesta Figura, verifica-se o valor de contribuição elevado da variável X_1 para o resíduo QEP no instante $k=1$.

Na Figura 4.20, as variáveis X_2 e X_3 apresentaram valores elevados de contribuição para o QEP , no instante $k=3$ do monitoramento *on line* da batelada 16.

Conclui-se que o modelo ACPM de referência construído em função dos oito primeiros CPs gerou um elevado valor de resíduo para a variável X_1 , no instante $k=1$, e para as variáveis X_2 e X_3 , no instante $k=3$. Deve-se concentrar esforços nestas variáveis, durante estes períodos, na busca de uma causa para o alto resíduo encontrado.

A partir da identificação as variáveis responsáveis pelos sinais fora de controle nas cartas monitoradas, os operadores e engenheiros devem utilizar seus conhecimentos sobre o processo produtivo para encontrar as causas que provocaram as alterações nas respectivas variáveis nos k instantes considerados.

4.6. Considerações Finais

No estudo de caso apresentado, foram selecionadas 15 bateladas preliminares utilizadas na elaboração da distribuição de referência, as quais geraram produto final dentro das especificações de qualidade. A seleção das bateladas preliminares para elaboração do modelo ACPM de referência deve contemplar apenas bateladas que geraram produtos com variáveis de qualidade Y dentro das especificações de qualidade requeridas. Assim, as bateladas que apresentam, através da análise ACPM, variação excessiva nas variáveis de processo ou produto final com qualidade fora das especificações, não devem ser utilizadas no modelo ACPM de referência para o monitoramento do desempenho de bateladas futuras.

Analisando os dados acerca das três variáveis de processo envolvidas para as 15 bateladas preliminares, durante os 15 intervalos de tempo, através da abordagem ACPM, observou-se que nenhuma das bateladas apresentou variação excessiva nas trajetórias destas variáveis, não sendo necessária a exclusão de nenhuma das bateladas preliminares do modelo ACPM de referência elaborado.

O modelo ACPM de referência é um modelo empírico construído a partir de dados históricos com relação às trajetórias das variáveis de processo, obtidos em bateladas passadas. Este modelo empírico considera a estrutura de correlação das variáveis obtida a partir do processo em controle estatístico para discriminar bateladas que ocorram dentro dos padrões requeridos.

Assim, a ACPM é uma ferramenta para a identificação de eventos atípicos que ocorram no processo produtivo através de alterações nas variáveis de processo X . As variáveis de qualidade do produto Y , não são utilizadas diretamente na elaboração do modelo proposto. Estas variáveis de qualidade do produto Y são utilizadas de forma a fornecer informações adicionais sobre a qualidade do produto final gerado nas bateladas. Os resultados obtidos através do monitoramento *on line* do processo com a ACPM em cada batelada, juntamente com os resultados obtidos da análise das variáveis de qualidade do produto final Y fornecem uma caracterização completa do processo produtivo e da qualidade do produto final gerado.

Dessa forma, conclui-se que a ACPM é uma ferramenta eficiente na detecção da presença de eventos especiais que provoquem alterações nas variáveis de processo. Eventos atípicos que provoquem alterações apenas nas variáveis de qualidade do produto final não são detectados na ACPM.

Quando houver alterações estruturais no processo em estudo, deve-se construir um novo modelo ACPM de referência, capaz de capturar novos eventos comuns que possam estar presentes no processo. Isto se deve ao fato de que estes novos eventos comuns presentes no processo, provavelmente ocasionam alteração na estrutura de correlação padrão das variáveis envolvidas. A partir disso, um novo modelo ACPM deve ser gerado através da obtenção de dados históricos com relação às variáveis de processo após as alterações terem sido efetuadas, gerando-se, assim, um novo modelo que constitua uma nova estrutura de correlação para as variáveis em estudo.

Finalmente, ressalta-se que para execução de um efetivo monitoramento e controle do processo produtivo, através do CEP, quando se detecta um sinal fora de controle, tendo este uma causa especial associada, deve-se utilizar a ACPM juntamente com os gráficos de contribuição para a identificação da causa especial atuante no processo produtivo, seguida de uma ação corretiva sobre o processo.

CAPÍTULO 5

5. CONCLUSÃO

A realização desta dissertação de mestrado permitiu a formulação de algumas conclusões a respeito da elaboração de um método para a implantação de Cartas de Controle Multivariadas Baseadas em Componentes Principais para Processos em Bateladas, assim como sugestões para possíveis trabalhos futuros.

Esta dissertação de Mestrado teve como objetivo principal a apresentação de um método para a implantação das Cartas de Controle Multivariadas Baseadas em Componentes Principais, para o monitoramento *on line* de processos industriais em batelada. O método elaborado é constituído da CC de Hotelling, utilizando como dados de entrada as novas variáveis definidas na ACP. Para tanto, fez-se necessário um desenvolvimento teórico detalhado sobre a ferramenta de projeção multivariada de dados escolhida: a Análise de Componentes Principais (ACP) e também o desenvolvimento teórico das CC multivariadas de Hotelling.

A dissertação apresentou também, 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 (MCUSUM, MEWMA) foram detalhadas. Foi realizada, ainda, uma revisão das cartas de controle baseadas em métodos de projeção de dados, entre as quais a ACP faz parte. Além disso, foi apresentada brevemente uma revisão sobre as estratégias tradicionais de diagnóstico

para as causas de distúrbios ocorridos no processo. Finalmente, discorreu-se brevemente sobre a interface entre o controle de processos na engenharia e CEP.

O método proposto mostrou-se eficiente no monitoramento *on line* de processos em batelada, onde um grande número de variáveis de processo esteja envolvido. Isto se deve ao fato de que métodos de projeção de dados, como a ACP, são técnicas estatísticas bastante úteis em processos onde existem informações sobre um número muito grande de variáveis fortemente correlacionadas. Estas técnicas permitem reduzir tais variáveis para um número menor de variáveis independentes denominadas Componentes Principais. Os CP trazem grande parte da informação contida nas variáveis iniciais.

Além disso, o método elaborado é indicado para situações especiais onde não seja possível a utilização das CC tradicionais. Tais situações ocorrem em processos automatizados, nos quais as variáveis podem ser medidas em tempo real e em pontos temporais próximos; condições onde a suposição de independência entre pontos amostrais dificilmente é verificada. Uma outra situação especial, na qual as CC univariadas não podem ser aplicadas, ocorre quando a variação normal de uma variável monitorada ao longo do tempo é descrita por um perfil; ou seja, a média da variável, em condições normais de processo, muda ao longo do tempo.

O método proposto na dissertação foi ilustrado através de um estudo de caso, apresentando a elaboração da CCM baseada em componentes principais em uma etapa do processo de fabricação de cerveja. A metodologia mostrou-se eficiente na incorporação das 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. As cartas de controle construídas forneceram sinalizações *on-line* sobre o estado do processo, em vários instantes da batelada, o que permitiria a correção dos distúrbios apresentados durante a batelada em tempo real.

5.1. Sugestões para Trabalhos Futuros

O método proposto nesta dissertação utiliza a ACPM como técnica estatística para projeção de dados, para a elaboração de um modelo de referência que capture o padrão de variação normal das variáveis de processo em estudo. Dessa forma, a ACPM é uma ferramenta para a identificação de eventos atípicos que ocorram no processo produtivo através de alterações nas variáveis de processo X . As variáveis de qualidade do produto Y não são utilizadas diretamente na elaboração do modelo proposto. Portanto, é importante ressaltar que esta abordagem considera somente a estrutura de correlação das variáveis de processo X na elaboração das cartas de controle e, conseqüentemente, monitora o processo através de amostras futuras das variáveis de processo. Como técnica alternativa à ACPM para monitoramento de processos em bateladas tem-se a PELM (Projeção de Estruturas Latentes *Multiway*), que permite a utilização simultânea das informações contidas nas variáveis de qualidade das matérias primas Z , nas variáveis de processo X , e nas variáveis de qualidade do produto final Y . A exploração teórica detalhada desta técnica e sua aplicação prática em uma indústria brasileira apresenta-se como um tema relevante para o desenvolvimento de trabalho futuro.

Os métodos de projeção de dados ACPM e PELM constituem-se em modelos lineares aplicáveis no monitoramento de processos. Entretanto, existem processos industriais complexos cujos dados apresentam fortes características não-lineares que não podem ser monitorados através de métodos lineares. Para o efetivo monitoramento destes processos, técnicas não lineares vêm sendo apresentadas na literatura, incluindo ACP não-linear, PEL não-linear, análise de correlação não-linear e procedimentos envolvendo redes neurais. O desenvolvimento de um trabalho que utilize um modelo não-linear para monitoramento de tais processos apresenta-se como tema relevante.

O método proposto nesta dissertação aplica-se somente aos processos em bateladas que se realizem com tempo fixo de duração. A extensão deste método proposto para processos industriais que ocorram em bateladas com tempo variável de duração constitui-se em assunto interessante para a realização de trabalho futuro.

Muitos processos industriais são realizados de maneira contínua. O desenvolvimento de uma metodologia de controle estatístico multivariado de processos que utilize técnicas de

projeção de dados, como ACPM e PEL, elaborada para aplicação em processos industriais contínuos, apresenta-se como tema interessante para realização de trabalho futuro.

Finalmente, aplicou-se o método proposto nesta dissertação, em um estudo de caso, apresentado no Capítulo 4. Entretanto, para viabilizar a efetiva execução deste método de controle estatístico de processos, necessita-se de ferramentas computacionais que incorporem a teoria proposta. O desenvolvimento de ferramentas computacionais faz-se necessário para que os operadores tenham acesso, em tempo real, às informações sobre o processo através da geração de cartas e gráficos de controle. A criação de *software* que viabilizem a implantação do método proposto se constitui num tema bastante relevante para realização de trabalhos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AKEY, T. M., GREEN, S. B. & SALKIND, N. J. 2^a Ed. **Using SPSS for Windows: analysing and understanding data**. Prentice Hall, Upper Saddle River, New Jersey, 2000.
2. ALT, F. B. Multivariate Quality Control, in **The Encyclopedia of Statistical Sciences**. John Wiley & Sons, Inc., New York, 1984.
3. APARISI, F. Sampling Plans for the Multivariate T^2 Control Chart. *Quality Engineering*, v. 10, n° 1 p. 141-147, 1997-98.
4. BOX, G. E. P., HUNTER, W. G. & HUNGER, J. S. **Statistics for Experiments**. John Wiley & Sons, Inc., New York, 1978.
5. BOX, G. E. P., JENKINS, G. M. & REINSEL, G. C. **Time Series Analysis – Forecasting and Control**. 3^a Ed., Prentice Hall, Inc., New Jersey, 1994.
6. BOX, G. E. P. & KRAMER, T. Statistical Process Control and Feedback Adjustment – A Discussion. *Technometrics*, v.36, p. 251-285, 1992.
7. BOX, G. E. P. & RAMIREZ, J. **Cumulative Scores Charts**. **Center for Quality and Productivity Improvement**, University of Wisconsin, Report n.58, Wisconsin, 1991.
8. BROOK, D. & EVANS, D. A. An approach to the probability distribution of cusum run length. *Biometrika*, v.59, n° 3, p. 539-544, 1972.
9. CATEN, C. S., RIBEIRO, J. L. D. & FOGLIATTO, F. S. Implantação de Controle Integrado de Processos – Etapas da Implantação e Estudo de Caso. *Produto & Produção*, v.4, n° 1, p. 22-39, 2000.
10. CHANG, T. Statistical Control of Correlated Variables. *ASQC Quality Congress Transactions – Milwaukee*, p. 298-305, 1991.
11. CHEN, G. & MCAVOY, T. Process Control Data Based Multivariate Statistical Models. *The Canadian Journal of Chemical Engineering*, v. 74, p. 1010-1024, 1996.

12. CHEN, J., BANDONI, J. A. & ROMAGNOLI, J. A. Robust PCA and Normal Region in Multivariate Statistical Process Monitoring. *AIChE Journal*, v. 42, n° 12, p. 3563-3570, 1996.
13. CLARKE-PRINGLE, T. & MACGREGOR, J.F. Optimization of Molecular-Weight Distribution Using Batch-to-Batch Adjustments. *Ind. Eng. Chem. Res.*, v. 37, p. 3660-3669, 1998.
14. CROWDER, S. V., HAWKINS, D. M., REYNOLDS, M. R. Jr. & YASHCHIN, E. Process Control and Statistical Inference. *Journal of Quality Technology*, v. 29, n° 2, p. 134-139, 1997.
15. CROWDER, V. C. Design of Exponentially Weighted Moving Average Schemes. *Journal of Quality Technology*, v. 21, n° 3, p. 155-162, 1989.
16. CROSIER, R. B. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, v. 30, n° 3, p. 291-303, 1988.
17. DAVIS, S. M. "From future perfect": Mass Customizing. *Planning Review*, p. 16-21, 1989.
18. 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.
19. DOWNS, J. & VOGEL, E. A Plant-Wide Industrial Process Control Problem. *Computers in Chemical Engineering*, v 17, p. 245-255, 1993.
20. DUNCAN, A. J. **Quality Control and Industrial Statistics**. 5^a Ed. Irwin, Chicago, 1986.
21. ELSAYED, E. A., RIBEIRO, J. L. & LEE, M. K. Automated Process Control and Quality Engeneering for Process with Damped Controllers. *International Journal of Production Research*, v.33, n° 10, p. 2923-2932, 1995.
22. FREUND, J. E. & SIMON, G. A. **Modern and Elementary Statistics**. 9^a Ed. Revisada, Prentice Hall, Inc., New Jersey, 1997.
23. FUCHS, C. & BENJAMINI, Y. Multivariate Profile Charts for Statistical Process Control. *Technometrics*, n° 36, p.182-195, 1994.
24. 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.
25. GELADI, P. & KOWALSKI, B. R. Partial Least Squares Regression: a Tutorial. *Analytica Chimica Acta*, v. 185, p.1-17, 1986a.
26. GNANADESIKAN, R. **Methods for Statistical Data Analysis of Multivariate Observation**. 2^a Ed., John Wiley & Sons, Inc., New York, 1997.

27. HAIR, J. F. Jr., ANDERSON, R. E., TATHAM, R. L. & BLACK, W. C. **Multivariate Data Analysis with Readings**. 4^a Ed., Prentice Hall, Inc., New Jersey, 1995.
28. HAWKINS, D. M. Multivariate Quality Control Based on Regression-Adjusted Variables. *Technometrics*, v.33, n°1, p. 61-75, 1991.
29. HAWKINS, D. M. Regression Adjustment for Variables in Multivariate Quality Control. *Journal of Quality Technology*, v. 25, n° 3, p. 170-182, 1993.
30. HAYER, A. & TSUI, K. Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Technology*, v. 26, n° 3, p. 197-207, 1994.
31. HEALY, J. D. A Note on Multivariate CUSUM Procedures. *Technometrics*, v. 29, n° 4, p. 409-412, 1987.
32. HIMES, D. M., STORER, R. H. & GEORGAKIS, C. Determination of the Number of Principal Components for Disturbance Detection and Isolation. *Proceedings of the American Control Conference Baltimore, Maryland*, p. 1279-1283, 1994.
33. HOLMES, D. S. & MERGEN, A. E. Improving the Performance of the T^2 Control Chart. *Quality Engineering*, v.5, p.619-625, 1993.
34. HOSKULDSSON, A. PLS Regression Methods. *Journal of chemometrics*, v. 2, p. 211-228, 1988.
35. HOTELLING, H. "Multivariate Quality Control" in **Techniques of Statistical Analysis**. McGraw Hill, New York, 1947.
36. JACKSON, J. E. **A User's Guide to Principal Components**. John Wiley & Sons, New York, 1991.
37. JACKSON, J. E. Multivariate Quality Control. *Communications in Statistics – Theory and Methods*, v. 14, p. 657-2688, 1985.
38. JACKSON, J. E. Principal Component and factor Analysis: Part I – Principal Components. *Journal of Quality Technology*, v. 12, n° 4, p. 201-213, 1980.
39. JACKSON, J. E. Principal Component and factor Analysis: Part II – Additional Topics Related to Principal Components. *Journal of Quality Technology*, v. 13, n° 1, p. 46-58, 1981.
40. JACKSON, J. E. Quality Control Methods for Two Related Variables. *Industrial Quality Control*, v. 12, n° 7, p. 4-8, 1956.
41. JACKSON, J. E. & MUDHOLKAR, G. S. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, v. 21, n° 3, p. 341-349, 1979.
42. JOHNSON, R.A. & WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 5^a Ed. Prentice Hall, Englewood Cliffs, New Jersey, 1992.
43. KIPARISSIDES, C., VEROS, G. & MAcGREGOR, J. F. Mathematical Modelling, Optimization and Quality Control of High Pressure Ethylene Polymerization Reactors.

Journal of Macromolecular Science, Reviews in Macromolecular Chemistry and Physics C33, v. 4, p. 437-527, 1993.

44. KOURTI, T., LEE, J. & MACGREGOR, J. F. Experiences with Industrial Applications on Projection Methods for Multivariate Statistical Process Control. *Computers in Chemical Engineering*, v 20 Supplementary, p. S745-S750, 1996
45. 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
46. KOURTI, T., NOMIKOS, P. & MACGREGOR, J. F. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal of process Control*, v. 5, n° 4, 1995.
47. 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.
48. LEHNINGER, A. L. **Princípios de Bioquímica**. 3ª Ed. Sarvier, São Paulo, 1986.
49. LORBER, A. WANGEN, L. E. & KOWALSKI, B. R. A theoretical Foundation for the PLS Algorithm. *Journal of Chemometrics*, v. 1, p. 19-31, 1987.
50. LOWRY, C. A. & MONTGOMERY, D. C. A Review of Multivariate Control Charts. *IIE Transactions*, v. 27, p. 800-810, 1995.
51. LOWRY C. A., WOODALL, W. H., CHAMP, C. W. & RIGDON, S. E. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, v. 34, n° 1, p. 46-53, 1992.
52. 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.
53. LUCAS, J. M. & SACCUCCI, M. S. Exponentially Weighted Moving Average Control Schemes: properties and enhancements. *Technometrics*, v. 32, n° 1, p. 1-12, 1990.
54. MACGREGOR, J. F. Using On-line Process Data to Improve Quality. *39ª Annual Fall Conference*, p. 277-284, 1995.
55. MACGREGOR, J. F. & HARRIS, T. J. Discussion of “Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements”, by J. M. Lucas e M. S. Saccucci. *Technometrics*, v. 32, n° 1, p. 23-26, 1990.
56. 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.
57. MACGREGOR, J. F. & KOURTI, T. Statistical Process Control of Multivariate Processes. *Control Engineering Practice*, v. 3, n° 3, p. 403-414, 1995.

58. MARSH, C. E. & TUCKER, T. W. Application of SPC Techniques to Batch Units. *ISA Transactions*, v. 30, n° 1, p. 39-47, 1991.
59. MARTENS, H. & NAES, T. **Multivariate Calibration**. John Wiley & Sons, Inc., New York, 1989.
60. MARTIN, E. B. & MORRIS, A. J. An overview of multivariate statistical process control in continuous and batch process performance monitoring. *Trans Inst MC*, v. 18, n° 1, p. 51-60, 1996.
61. 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.
62. 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.
63. MASON, R. L., TRACY, N. D. & YOUNG, J. C. Decomposition of T^2 for Multivariate Control Chart Interpretation, *Journal of Quality Technology*, v. 27, n° 2, p. 99-108, 1995.
64. MASON, R. L. & YOUNG, J. C. Why Multivariate Statistical Process Control? *Quality Progress*, v. 31, n° 12, p. 88-93, 1998.
65. MICHEL, R. & FOGLIATTO, F.S. Projeto Econômico de Cartas Adaptativas para Monitoramento de Processos. Artigo submetido à revista Produto & Produção para publicação, 2000.
66. 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.
67. MONTGOMERY, D. C. **Introduction to Statistical Quality Control**. 3^a Ed. John Wiley & Sons, Inc., New York, 1996.
68. 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.
69. MONTGOMERY, D. C., KEATS, B. J., RUNGER, G. C. & MESSINA, W. S. Integrating Statistical Process Control and Engineering Process Control. *Journal of Quality Technology*, v. 26, n° 2, p. 79-87, 1994.
70. 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.
71. MONTGOMERY, D. C., MASTRANGELO, C. M. & LOWRY, C. M. Process Monitoring for Dynamic Systems. *Proceedings of the Institute of Industrial Engineers Research Conference*, p. 559-563, 1993.

72. MONTGOMERY, D. C. & WADSWORTH, H. M. Jr. Some Techniques for Multivariate Quality Control Applications. *ASDC Technical Conference Transactions*, p. 427-435, 1972.
73. MOOD, A. M., GRAYBILL, F. A. & BOES, D. C. **Introduction to the Theory of Statistics**. MacGraw-Hill Publishing Company, New York, 1974.
74. MURPHY, B. J. Selecting out of control variables with the T^2 multivariate quality control procedure. *The Statistician*, v. 36, p. 571-583, 1987.
75. NEGIZ, A. & ÇINAR, A. Statistical Monitoring of Multivariable Dynamic Processes with State-Space Models. *AIChE Journal*, v. 43, n° 8, p. 2002-2019, 1997.
76. NELSON, L. S. Control Charts for Individual Measurements. *Journal of Quality Technology*, v. 14, n° 2, p. 172-173, 1982.
77. 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.
78. NOMIKOS, P. & MACGREGOR, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, v. 37, n° 1, p. 41-59, 1995.
79. 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.
80. PIGNATIELLO, J. J. Jr. & RUNGER, G. C. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, v. 22, n° 3, p. 173-186, 1990.
81. PIOSOVO, M. J. & KOSANOVICH, K. Applications of Multivariate Statistical Methods to Process Monitoring and Controller Design. *International Journal Control*, v. 59, n° 3, p. 743-765, 1994.
82. PRINS, J. & MADER, D. Multivariate Control Charts for Grouped and Individual Observations. *Quality Engeneering*, v. 10, n° 1, p. 49-57, 1997-98.
83. PROCEP. Software para o Controle Estatístico de Processos – Versão para Windows 98. Maxi Gestão Empresarial. Porto Alegre, RS, 2000.
84. RENCHER, A. C. **Methods of Multivariate Analysis**. John Wiley & Sons, Inc., New York, 1995.
85. RENCHER, A. C. The Contribution of Individual Variables to Hotelling's T^2 , Wilks' Λ , and R^2 . *Biometrics*, v. 49, p. 479-489, 1993.
86. ROBERTS, S. W. Control Chart Tests Based on Geometrics Moving Averages. *Technometrics*, v. 1, p. 239-250, 1959.

87. ROESCH, S. M. A. *A Dissertação de Mestrado em Administração: Proposta de uma Tipologia*. Série Documentos para Estudo 14/94. Programa de Pós Graduação em Administração, Administração, UFRGS, Porto Alegre, 1994.
88. RYAN, T. P. **Statistical Methods for Quality Improvement**. 1ª Ed. John Wiley & Sons, Inc., New York, 1989.
89. SAUERS, D. G. Hotelling's T^2 Atatistic for Multivariate Statistical Process Control: A Nonrigorous Approach. *Quality Engeneering*, v. 9, n° 4, p. 627-634, 1997.
90. SKAGERBERG, B., MACGREGOR, J. F. & KIPARISSIDES, C. Multivariate Data Analysis Applied to Low Density Polyethylene Reactors. *Chemometrics and Intelligent Laboratory Systems*, v. 14, p. 341-356, 1992.
91. STRANG. G. **Linear Algebra and its Applications**. 3ª Ed. Harcourt Brace Jovanovich, Inc., Tokio, 1988.
92. 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.
93. 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.
94. TURCKER, W.T., FALTIN, F. W. & VANDER WIEL, S. A. Algorithmic Statistical Process Control: An Elaboration. *Technometrics*, v. 35, p. 363-375, 1993.
95. 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.
96. WANGEN, L. E. & KOWALSKI, B. A Multi-block Partial Least Squares Algorithm for Investigating Complex Chemical Systems. *Journal of Chemometrics*, n° 3, p. 3-20, 1988.
97. WIERDA, S. J. Multivariate Statistical Process Control – Recent Results and Directions for Future Research. *Statistica Neerlandica*, v.48. p. 147-168, 1994.
98. 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.
99. WOODALL, W. H. Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology*, v. 32, n° 4, p. 341-350, 2000.
100. WOODALL, W. H. & NCUBE, M. M. Multivariate CUSUM Quality Control Procedures. *Technometrics*, v. 27, n° 3, p. 285-292, 1985.
101. WOODALL, W. H. & THOMAS, E. V. Statistical Process Control with Several Components of Common Cause Variability. *IIE Transactions*, v. 27, p. 757-764, 1995.
102. YASHCHIN, E. Some aspects of the theory of statistical control schemes. *IBM Journal of Research and Development*, v. 31, p. 199-205, 1987.
103. YASHCHIN, E. Performance of CUSUM Control Schemes for Serially Correlated Observations. *Technometrics*, v. 35, p. 37-52, 1993a.

104. YASHCHIN, E. Statistical Control Schemes: Methods, Applications and Generalizations. *International Statistical Review*, v. 61, p. 41-66, 1993b.
105. YIN, R. K. **Case study Research – de sign and methods**. 2^a Ed. Stage Publications, London, 1994.
106. ZHANG, G. X. A New Type of Control Charts and a Theory of Diagnosis with Control Charts. *World Quality Congress Transactions*, p. 175-185, 1984.
107. ZHANG, G. X. Cause-Selecting Control Charts – A new type of quality control charts. *The QR Journal*, v.12, p. 221-225, 1985.

ANEXO 1

Este anexo contém os dados das variáveis de processo coletados, nos 15 intervalos de tempo, para as 15 bateladas utilizadas na distribuição de referência do estudo de caso realizado.

Além disso, são apresentados aqui os dados referentes a ACPM aplicada aos dados das bateladas utilizadas na distribuição de referência do estudo de caso realizado.

Tabela A1 - Observações das variáveis de processo X1, X2 e X3, coletadas nos 15 intervalos de tempo, do estudo de caso realizado

Bateladas	k=1			k=2			k=3			k=4			k=5			k=6			k=7			k=8			k=9			k=10			k=11			k=12			k=13			k=14			k=15		
	X1 ₁	X2 ₁	X3 ₁	X1 ₂	X2 ₂	X3 ₂	X1 ₃	X2 ₃	X3 ₃	X1 ₄	X2 ₄	X3 ₄	X1 ₅	X2 ₅	X3 ₅	X1 ₆	X2 ₆	X3 ₆	X1 ₇	X2 ₇	X3 ₇	X1 ₈	X2 ₈	X3 ₈	X1 ₉	X2 ₉	X3 ₉	X1 ₁₀	X2 ₁₀	X3 ₁₀	X1 ₁₁	X2 ₁₁	X3 ₁₁	X1 ₁₂	X2 ₁₂	X3 ₁₂	X1 ₁₃	X2 ₁₃	X3 ₁₃	X1 ₁₄	X2 ₁₄	X3 ₁₄	X1 ₁₅	X2 ₁₅	X3 ₁₅
1	9,90	14,81	11,00	13,94	0,06	12,10	13,30	0,10	12,40	12,00	0,19	12,50	11,28	0,24	12,50	10,45	0,29	14,00	9,06	0,39	14,00	7,51	0,49	14,00	5,81	0,61	13,90	4,97	0,66	14,00	4,30	0,71	13,90	4,05	0,73	14,00	3,94	0,73	14,00	3,92	0,74	14,00	3,80	0,74	
2	10,30	15,64	11,50	13,25	0,15	12,20	13,50	0,14	12,00	11,90	0,24	13,30	10,43	0,33	12,40	8,51	0,46	13,60	6,68	0,57	14,00	5,51	0,65	13,90	4,72	0,70	13,90	4,44	0,72	14,00	3,95	0,75	13,70	3,91	0,75	14,00	3,87	0,75	13,90	3,85	0,75	13,90	3,76	0,76	
3	11,10	14,35	12,40	12,98	0,10	12,50	11,39	0,21	12,50	9,67	0,33	13,50	7,76	0,46	14,10	5,87	0,59	14,10	5,87	0,59	14,00	4,40	0,69	13,80	3,76	0,74	13,90	3,54	0,75	13,90	3,53	0,75	14,00	3,53	0,75	13,90	3,53	0,75	13,90	3,42	0,76	13,90	3,39	0,76	
4	10,40	15,06	11,40	13,85	0,08	12,40	11,07	0,26	12,50	8,84	0,41	14,00	7,08	0,53	14,00	5,68	0,62	14,00	4,55	0,70	14,00	4,14	0,73	14,00	4,02	0,73	14,00	3,90	0,74	14,00	3,72	0,75	14,00	3,63	0,76	14,00	3,63	0,76	14,00	3,63	0,76	14,00	3,63	0,76	
5	12,00	15,12	12,00	14,49	0,04	12,30	12,35	0,18	12,50	11,72	0,22	12,50	9,47	0,37	14,00	8,22	0,46	14,00	8,22	0,46	14,00	6,15	0,59	14,00	4,23	0,72	14,00	4,05	0,73	14,00	4,05	0,73	14,00	3,88	0,74	14,00	3,88	0,74	14,00	3,74	0,75	14,00	3,65	0,76	
6	12,00	14,06	12,10	14,06	0,21	12,50	11,08	0,21	12,50	10,12	0,28	14,00	6,73	0,52	14,00	5,32	0,62	14,00	4,34	0,69	14,00	3,72	0,74	14,00	3,64	0,74	14,00	3,61	0,74	14,00	3,61	0,74	14,00	3,48	0,75	14,00	3,48	0,75	14,00	3,48	0,75	14,00	3,48	0,75	
7	10,80	14,62	11,90	14,62	0,03	12,40	13,50	0,13	12,80	10,09	0,33	12,60	8,60	0,43	13,80	6,05	0,60	13,60	5,59	0,63	13,70	4,38	0,71	14,00	4,04	0,73	14,00	4,04	0,73	14,00	3,87	0,74	14,00	3,87	0,74	13,90	3,71	0,76	13,60	3,71	0,76	13,60	3,71	0,76	
8	10,00	14,38	10,80	14,38	0,00	11,90	13,33	0,07	12,90	10,09	0,30	14,10	10,09	0,30	13,90	7,36	0,49	14,10	7,36	0,49	14,00	6,08	0,58	13,90	4,64	0,68	14,00	4,16	0,71	14,00	3,86	0,73	14,00	3,80	0,74	13,90	3,65	0,75	13,90	3,64	0,75	14,00	3,64	0,75	
9	9,90	15,35	11,20	14,24	0,07	12,40	12,90	0,16	12,40	10,97	0,29	12,30	9,34	0,39	12,80	7,39	0,52	13,90	7,39	0,52	13,90	6,05	0,61	13,80	4,46	0,71	13,70	4,26	0,72	14,00	4,01	0,74	13,70	3,87	0,75	13,80	3,83	0,75	13,80	3,82	0,75	13,60	3,79	0,75	
10	9,70	15,23	11,00	15,23	0,00	12,50	13,94	0,08	13,80	12,29	0,19	13,70	10,10	0,34	13,70	8,34	0,45	13,70	6,54	0,57	13,80	5,40	0,65	14,50	4,36	0,71	13,70	4,14	0,73	13,80	4,00	0,74	13,70	3,88	0,75	13,80	3,82	0,75	13,80	3,75	0,75	13,60	3,68	0,76	
11	9,60	15,11	10,50	14,69	0,03	11,00	13,72	0,09	12,30	12,61	0,17	12,30	10,73	0,29	12,40	9,38	0,38	13,80	7,43	0,51	13,80	4,89	0,68	13,70	4,53	0,70	13,80	4,00	0,74	13,80	4,24	0,72	13,80	4,01	0,73	13,80	4,01	0,73	13,80	4,00	0,74	13,80	4,00	0,74	
12	9,80	15,20	11,10	14,79	0,03	11,30	13,61	0,10	11,30	11,92	0,22	11,40	10,68	0,30	11,30	8,89	0,42	12,80	7,29	0,52	13,60	6,14	0,60	14,20	4,49	0,70	14,20	3,99	0,74	14,20	3,87	0,75	14,30	3,78	0,75	14,10	3,74	0,75	14,10	3,74	0,75	14,10	3,67	0,76	
13	11,00	14,31	11,00	14,16	0,01	11,00	13,58	0,05	11,00	12,33	0,14	11,00	10,93	0,24	11,60	9,56	0,33	13,70	8,16	0,43	14,30	6,03	0,58	14,10	4,50	0,69	14,20	3,39	0,76	13,50	3,25	0,77	13,60	3,23	0,77	13,60	3,17	0,78	13,60	3,12	0,78	13,60	3,15	0,78	
14	11,00	14,99	11,40	13,87	0,07	11,30	13,87	0,07	11,40	12,44	0,17	11,30	12,44	0,17	12,40	10,67	0,29	14,30	6,84	0,54	14,30	5,21	0,65	14,30	4,14	0,72	14,20	3,81	0,75	14,20	3,64	0,76	14,20	3,57	0,76	14,20	3,52	0,77	14,20	3,52	0,77	14,20	3,45	0,77	
15	12,30	13,06	11,30	12,46	0,05	11,30	10,74	0,18	11,20	8,06	0,38	13,70	4,66	0,64	14,30	3,99	0,69	14,20	3,23	0,77	14,00	3,00	0,77	14,00	2,98	0,77	14,00	2,98	0,77	14,00	2,87	0,78	14,00	2,76	0,79	14,00	2,76	0,79	14,00	2,76	0,79	14,00	2,76	0,79	

Tabela A2 - Observações normalizadas das variáveis de processo X1, X2 e X3, coletadas nos 15 intervalos de tempo, do estudo de caso realizado

Bateladas	k=1			k=2			k=3			k=4			k=5			k=6			k=7			k=8			k=9			k=10			k=11			k=12			k=13			k=14			k=15		
	X1 ₁	X2 ₁	X3 ₁	X1 ₂	X2 ₂	X3 ₂	X1 ₃	X2 ₃	X3 ₃	X1 ₄	X2 ₄	X3 ₄	X1 ₅	X2 ₅	X3 ₅	X1 ₆	X2 ₆	X3 ₆	X1 ₇	X2 ₇	X3 ₇	X1 ₈	X2 ₈	X3 ₈	X1 ₉	X2 ₉	X3 ₉	X1 ₁₀	X2 ₁₀	X3 ₁₀	X1 ₁₁	X2 ₁₁	X3 ₁₁	X1 ₁₂	X2 ₁₂	X3 ₁₂	X1 ₁₃	X2 ₁₃	X3 ₁₃	X1 ₁₄	X2 ₁₄	X3 ₁₄	X1 ₁₅	X2 ₁₅	X3 ₁₅
1	-0,84	0,09	-0,71	-0,17	-0,04	0,27	0,44	-0,56	0,22	0,70	-0,82	-0,30	0,94	-1,02	-0,65	1,38	-1,52	0,41	1,56	-1,61	0,21	1,95	-2,12	-0,07	2,46	-2,79	-0,42	2,18	-2,85	0,24	1,39	-1,88	-0,14	1,08	-1,35	0,46	0,94	-1,46	0,56	0,98	-1,19	0,57	0,76	-1,45	
2	-0,39	1,37	0,24	-1,12	1,54	0,44	0,62	0,07	-0,31	0,63	-0,21	0,47	0,52	-0,32	-0,75	0,40	-0,16	-0,70	0,07	0,11	0,21	0,23	0,03	-0,56	0,70	-0,28	-0,42	1,05	-0,50	0,24	0,45	0,33	-1,17	0,67	-0,04	0,46	0,72	-0,24	-0,04	0,76	-0,48	0,07	0,63	0,05	
3	0,50	-0,62	1,95	-1,49	0,67	0,96	-1,22	1,19	0,36	-0,93	0,89	0,66	-0,78	0,71	0,96	-0,93	0,87	0,68	-0,44	0,30	0,21	-0,72	0,56	-1,05	-0,85	0,84	-0,42	-0,88	0,68	-0,36	-0,69	0,33	0,38	-0,45	-0,04	-0,23	-0,33	-0,24	-0,04	-0,58	0,24	0,07	-0,60	0,05	
4	-0,28	0,48	0,05	-0,30	0,32	0,79	-1,50	1,99	0,36	-1,51	1,87	1,14	-1,11	1,26	0,86	-1,02	1,11	0,41	-1,26	1,35	0,21	-0,95	1,10	-0,07	-0,43	0,56	0,21	-0,11	0,29	0,24	-0,17	0,33	0,38	-0,16	0,61	0,46	-0,02	0,37	0,56	0,07	0,24	0,57	0,20	0,05	
5	1,50	0,57	1,19	0,58	-0,39	0,62	-0,38	0,71	0,36	0,50	-0,46	-0,30	0,06	0,00	0,86	0,26	-0,16	0,41	1,03	-0,94	0,21	0,78	-0,78	-0,07	-0,09	0,28	0,21	0,21	-0,10	0,24	0,71	-0,78	0,38	0,58	-0,70	0,46	0,75	-0,85	0,56	0,42	-0,48	0,57	0,26	0,05	
6	1,50	-1,07	1,38	-0,01	2,59	0,96	-1,49	1,19	0,36	-0,62	0,28	1,14	-1,28	1,18	0,86	-1,20	1,11	0,41	-1,40	1,25	0,21	-1,31	1,23	-0,07	-1,05	0,84	0,21	-0,73	0,29	0,24	-0,47	-0,22	0,38	-0,60	-0,04	0,46	-0,48	-0,24	0,56	-0,40	-0,48	0,57	-0,30	-0,70	
7	0,16	-0,21	1,00	0,76	-0,56	0,79	0,62	-0,09	0,76	-0,64	0,89	-0,21	-0,37	0,47	0,66	-0,84	0,95	-0,70	-0,61	0,68	-1,38	-0,74	0,83	-0,07	-0,40	0,56	0,21	0,19	-0,10	0,24	0,23	-0,22	0,38	0,55	-0,70	-0,23	0,23	0,37	-1,84	0,32	0,24	-1,44	0,46	0,05	
8	-0,73	-0,58	-1,09	0,43	-1,09	-0,07	0,47	-1,04	0,89	-0,64	0,52	1,24	0,36	-0,55	0,76	-0,18	0,07	0,68	0,49	-0,66	0,21	0,72	-0,91	-0,56	0,57	-0,84	0,21	0,45	-0,89	0,24	0,20	-0,78	0,38	0,34	-0,70	-0,23	0,04	-0,24	-0,04	0,10	-0,48	0,57	0,23	-0,70	
9	-0,84	0,92	-0,33	0,24	0,14	0,79	0,09	0,39	0,22	-0,02	0,40	-0,49	-0,01	0,16	-0,35	-0,16	0,31	0,13	0,51	-0,37	-0,32	0,70	-0,51	-1,05	0,28	0,00	-1,68	0,66	-0,50	0,24	0,61	-0,22	-1,17	0,55	-0,04	-0,92	0,60	-0,24	-0,64	0,67	-0,48	-1,44	0,73	-0,70	
10	-1,06	0,74	-0,71	1,60	-1,09	0,96	1,00	-0,88	2,10	0,90	-0,82	0,85	0,36	-0,24	0,56	0,32	-0,24	-0,43	-0,02	0,11	-0,85	0,14	0,03	2,40	0,12	0,00	-1,68	0,40	-0,10	-0,95	0,58	-0,22	-1,17	0,58	-0,04	-0,92	0,57	-0,24	-0,64	0,45	-0,48	-1,44	0,36	0,05	
11	-1,17	0,55	-1,66	0																																									

Tabela A3 - Matriz de cargas dos CP retidos no modelo ACPM elaborado.

Tempo	Variáveis	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
k=1	X1 ₁	0,1704	-0,0266	0,0810	-0,1621	0,1862	0,1459	-0,2136	-0,2482
	X2 ₁	-0,1702	0,0565	0,0394	0,1615	0,2592	0,0752	0,0671	0,2693
k=2	X1 ₂	0,1044	0,1286	0,0999	-0,0667	0,3725	0,2060	-0,4235	-0,0706
	X2 ₂	-0,1487	0,0136	-0,0937	0,2900	-0,0715	0,2164	-0,0858	-0,2488
	X3 ₂	0,0654	0,1352	0,1502	-0,1448	0,4553	-0,0823	0,0851	0,1875
k=3	X1 ₃	0,0021	0,2731	-0,0383	-0,0869	0,0941	0,3103	-0,2556	0,2295
	X2 ₃	-0,1792	-0,1359	-0,0540	0,1649	-0,0242	0,1101	0,0693	0,0378
	X3 ₃	0,1296	0,2203	0,0663	-0,0660	0,1900	-0,0909	-0,1203	0,0976
k=4	X1 ₄	-0,0660	0,2522	-0,1470	0,0117	-0,1275	0,3608	0,0538	-0,0717
	X2 ₄	-0,1822	-0,1405	0,0022	0,0589	0,2404	0,1200	0,1040	-0,0524
	X3 ₄	0,1531	0,1918	0,0078	0,0123	-0,1928	-0,1300	-0,1070	0,1576
k=5	X1 ₅	0,0891	0,2469	-0,0546	-0,1090	-0,2037	0,1128	0,1889	0,0846
	X2 ₅	-0,1925	-0,1483	0,0599	0,0126	0,0774	0,1030	0,0755	0,0027
	X3 ₅	0,1830	0,1731	-0,0587	0,0315	-0,0454	-0,0996	-0,0696	0,0514
k=6	X1 ₆	0,1259	0,2158	-0,0523	-0,1200	-0,1569	0,2791	0,0002	-0,1889
	X2 ₆	-0,1846	-0,1716	0,0778	-0,0412	0,0898	0,0712	0,1167	0,0032
	X3 ₆	0,1732	0,1950	-0,0711	0,0849	-0,0540	-0,0678	-0,1147	0,0375
k=7	X1 ₇	0,0783	0,0429	0,0166	-0,4252	-0,0409	0,2223	0,3315	-0,1876
	X2 ₇	-0,1983	-0,1043	-0,0086	-0,1372	0,0145	-0,0109	-0,1815	-0,0802
	X3 ₇	0,1891	0,1200	0,0139	0,1741	0,0047	0,0108	0,1979	0,1229
k=8	X1 ₈	0,0493	-0,1686	0,0423	-0,4162	0,1735	0,1865	0,1596	0,0253
	X2 ₈	-0,1936	-0,0791	0,0346	-0,1443	-0,1040	0,0138	-0,2496	0,1241
	X3 ₈	0,1776	0,1005	-0,0348	0,2051	0,1534	-0,0087	0,2796	-0,0686
k=9	X1 ₉	-0,0067	-0,1291	0,0678	0,2151	-0,1338	0,5688	-0,0142	0,1921
	X2 ₉	-0,2069	-0,0229	0,0127	-0,1526	-0,0989	-0,0799	-0,0666	0,1825
	X3 ₉	0,1881	0,0469	-0,0191	0,2133	0,2028	0,0970	0,0748	-0,1233
k=10	X1 ₁₀	0,0552	-0,2098	0,2280	0,0566	0,0019	-0,0291	-0,2882	-0,1425
	X2 ₁₀	-0,1926	0,1169	0,0596	-0,0801	-0,0797	0,0212	-0,0386	0,2602
	X3 ₁₀	0,1612	-0,1247	-0,0549	0,2264	0,2248	0,0062	0,1202	-0,1927
k=11	X1 ₁₁	0,0072	0,1030	0,3819	0,1382	-0,1090	0,0061	-0,0114	0,1231
	X2 ₁₁	-0,2003	0,1477	0,0285	0,0392	0,0296	0,0091	0,0433	-0,0369
	X3 ₁₁	0,1692	-0,1834	-0,0099	0,1050	0,0933	0,0306	-0,0040	0,2597
k=12	X1 ₁₂	0,0496	0,0199	0,3710	0,1803	-0,1745	0,0034	-0,1543	-0,2550
	X2 ₁₂	-0,1987	0,1466	0,0183	0,0709	0,0585	0,0380	-0,0155	0,0002
	X3 ₁₂	0,1813	-0,1612	-0,0186	0,0170	-0,0301	0,0324	0,0720	0,2806
k=13	X1 ₁₃	0,0291	0,0531	0,4307	0,0650	-0,0317	0,0809	0,0630	0,0840
	X2 ₁₃	-0,1924	0,1567	0,0298	0,0750	0,1260	0,0181	0,0268	-0,0097
	X3 ₁₃	0,1694	-0,1887	-0,0439	0,0474	-0,0684	0,0926	-0,0198	0,1923
k=14	X1 ₁₄	0,0245	0,0126	0,4154	-0,0389	-0,0595	0,0569	0,1994	0,0408
	X2 ₁₄	-0,1920	0,1559	0,0414	0,0977	0,0994	-0,0237	0,0712	0,0230
	X3 ₁₄	0,1741	-0,1971	-0,0418	-0,0233	-0,0482	0,0641	-0,0412	0,0828
k=15	X1 ₁₅	0,0341	-0,0067	0,4216	-0,0870	-0,0968	-0,0485	0,1264	-0,1000
	X2 ₁₅	-0,1867	0,1617	0,0096	0,1135	0,0998	-0,0575	0,0871	-0,0187
	X3 ₁₅	0,1593	-0,2036	-0,0226	0,0588	-0,0110	0,1420	-0,0978	0,1653

Tabela A4 - Matriz de escores dos CP retidos no modelo ACPM elaborado.

Bateladas	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
1	-7,0950	0,5400	1,5767	-3,1188	-1,4846	-0,4524	-0,5045	0,6364
2	-1,8579	0,7160	0,4451	-0,3492	1,8615	-0,6200	0,5885	2,2597
3	3,7250	2,0598	-0,2519	-1,2600	1,0536	-0,1906	-0,7241	-0,5685
4	3,6074	3,0182	0,6469	0,3073	-0,3829	-0,2109	0,4863	1,0548
5	-1,4019	1,0519	1,3004	-0,9088	0,8981	1,2710	-1,4613	-1,4624
6	4,2208	3,1267	1,0567	-0,5125	1,5034	0,3212	0,0894	-0,8603
7	0,7246	2,0016	-1,7055	2,1690	-0,1921	0,3761	-1,5844	-0,4494
8	-1,8649	0,7216	-0,0733	-1,2025	-2,8022	0,0452	0,2162	-0,8082
9	-2,0394	1,6453	-1,9471	-0,2709	0,5764	-0,8501	-0,1124	1,1881
10	-2,5894	0,7947	-3,0934	1,7764	-0,8495	3,0640	0,9140	0,6851
11	-4,6057	0,4375	-1,7183	1,0134	0,6119	-2,0519	2,2567	-2,0005
12	-2,5305	-2,1230	3,2534	3,7820	-0,6455	-1,4289	-1,3000	0,3028
13	0,4202	-7,7203	-3,6860	-1,1453	0,8201	-0,2645	-0,9877	-0,0331
14	-0,1003	-4,4805	4,2495	0,0138	0,7792	1,8021	1,4744	-0,2227
15	11,3870	-1,7894	-0,0532	-0,2941	-1,7473	-0,8103	0,6488	0,2781

Tabela A5- Valores calculados de Q_i , T^2 e seus respectivos LCS para o modelo ACPM elaborado com as 15 bateladas preliminares.

Bateladas	Q_i	LCS	T^2	LCS
1	1,658822	6,302613	0,65792	0,962088
2	1,727211	6,302613	0,53987	0,962088
3	2,663908	6,302613	0,24118	0,962088
4	4,375465	6,302613	0,23470	0,962088
5	3,027089	6,302613	0,44602	0,962088
6	5,189046	6,302613	0,33448	0,962088
7	4,333584	6,302613	0,39085	0,962088
8	2,342443	6,302613	0,45217	0,962088
9	3,534029	6,302613	0,24096	0,962088
10	1,351493	6,302613	0,82468	0,962088
11	0,191703	6,302613	0,94226	0,962088
12	1,04066	6,302613	0,85400	0,962088
13	0,793636	6,302613	0,87673	0,962088
14	1,631929	6,302613	0,78271	0,962088
15	1,492402	6,302613	0,73790	0,962088

ANEXO 2

Este anexo contém os dados das variáveis de processo coletados, nos 15 intervalos de tempo, para a realização do monitoramento *on line* da nova batelada , a batelada 16, para o estudo de caso realizado.

Além disso, são apresentados aqui os dados referentes à aplicação do método proposto ao monitoramento *on line* da batelada 16 no estudo de caso realizado.

Tabela A1 - Observações das variáveis de processo X₁, X₂ e X₃, coletadas nos 15 intervalos de tempo, do estudo de caso realizado incluindo as observações para Batelada 16.

Bateladas	k=1			k=2			k=3			k=4			k=5			k=6			k=7			k=8			k=9			k=10			k=11			k=12			k=13			k=14			k=15				
	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃								
1	9,90	14,81	11,00	13,94	0,06	12,10	13,30	0,10	12,40	12,00	0,19	12,50	11,28	0,24	12,50	10,45	0,29	14,00	9,06	0,39	14,00	7,51	0,49	14,00	5,81	0,61	13,90	4,97	0,66	14,00	4,30	0,71	13,90	4,05	0,73	14,00	3,94	0,73	14,00	3,92	0,74	14,00	3,80	0,74			
2	10,30	15,64	11,50	13,25	0,15	12,20	13,50	0,14	12,00	11,90	0,24	13,30	10,43	0,33	12,40	8,51	0,46	13,60	6,68	0,57	14,00	5,51	0,65	13,90	4,72	0,70	13,90	4,44	0,72	14,00	3,95	0,75	13,70	3,91	0,75	14,00	3,87	0,75	13,90	3,85	0,75	13,90	3,76	0,76			
3	11,10	14,35	12,40	12,98	0,10	12,50	11,39	0,21	12,50	9,67	0,33	13,50	7,76	0,46	14,10	5,87	0,59	14,10	5,87	0,59	14,00	4,40	0,69	13,80	3,76	0,74	13,90	3,54	0,75	13,90	3,53	0,75	14,00	3,53	0,75	13,90	3,53	0,75	13,90	3,42	0,76	13,90	3,39	0,76			
4	10,40	15,06	11,40	13,85	0,08	12,40	11,07	0,26	12,50	8,84	0,41	14,00	7,08	0,53	14,00	5,68	0,62	14,00	4,55	0,70	14,00	4,14	0,73	14,00	4,02	0,73	14,00	3,90	0,74	14,00	3,72	0,75	14,00	3,63	0,76	14,00	3,63	0,76	14,00	3,63	0,76	14,00	3,63	0,76			
5	12,00	15,12	12,00	14,49	0,04	12,30	12,35	0,18	12,50	11,72	0,22	12,50	9,47	0,37	14,00	8,22	0,46	14,00	8,22	0,46	14,00	6,15	0,59	14,00	4,23	0,72	14,00	4,05	0,73	14,00	4,05	0,73	14,00	3,88	0,74	14,00	3,88	0,74	14,00	3,74	0,75	14,00	3,65	0,76			
6	12,00	14,06	12,10	14,06	0,21	12,50	11,08	0,21	12,50	10,12	0,28	14,00	6,73	0,52	14,00	5,32	0,62	14,00	4,34	0,69	14,00	3,72	0,74	14,00	3,64	0,74	14,00	3,61	0,74	14,00	3,61	0,74	14,00	3,48	0,75	14,00	3,48	0,75	14,00	3,48	0,75	14,00	3,48	0,75			
7	10,80	14,62	11,90	14,62	0,03	12,40	13,50	0,13	12,80	10,09	0,33	12,60	8,60	0,43	13,80	6,05	0,60	13,60	5,59	0,63	13,70	4,38	0,71	14,00	4,04	0,73	14,00	4,04	0,73	14,00	3,87	0,74	14,00	3,87	0,74	14,00	3,87	0,74	13,90	3,71	0,76	13,60	3,71	0,76	13,60	3,71	0,76
8	10,00	14,38	10,80	14,38	0,00	11,90	13,33	0,07	12,90	10,09	0,30	14,10	10,09	0,30	13,90	7,36	0,49	14,10	7,36	0,49	14,00	6,08	0,58	13,90	4,64	0,68	14,00	4,16	0,71	14,00	3,86	0,73	14,00	3,80	0,74	13,90	3,65	0,75	13,90	3,64	0,75	14,00	3,64	0,75			
9	9,90	15,35	11,20	14,24	0,07	12,40	12,90	0,16	12,40	10,97	0,29	12,30	9,34	0,39	12,80	7,39	0,52	13,90	7,39	0,52	13,90	6,05	0,61	13,80	4,46	0,71	13,70	4,26	0,72	14,00	4,01	0,74	13,70	3,87	0,75	13,80	3,83	0,75	13,80	3,82	0,75	13,60	3,79	0,75			
10	9,70	15,23	11,00	15,23	0,00	12,50	13,94	0,08	13,80	12,29	0,19	13,70	10,10	0,34	13,70	8,34	0,45	13,70	6,54	0,57	13,80	5,40	0,65	14,50	4,36	0,71	13,70	4,14	0,73	13,80	4,00	0,74	13,70	3,88	0,75	13,80	3,82	0,75	13,80	3,75	0,75	13,60	3,68	0,76			
11	9,60	15,11	10,50	14,69	0,03	11,00	13,72	0,09	12,30	12,61	0,17	12,30	10,73	0,29	12,40	9,38	0,38	13,80	7,43	0,51	13,80	4,89	0,68	13,70	4,53	0,70	13,80	4,00	0,74	13,80	4,24	0,72	13,80	4,01	0,73	13,80	4,01	0,73	13,80	4,00	0,74	13,80	4,00	0,74			
12	9,80	15,20	11,10	14,79	0,03	11,30	13,61	0,10	11,30	11,92	0,22	11,40	10,68	0,30	11,30	8,89	0,42	12,80	7,29	0,52	13,60	6,14	0,60	14,20	4,49	0,70	14,20	3,99	0,74	14,20	3,87	0,75	14,30	3,78	0,75	14,10	3,74	0,75	14,10	3,67	0,76	14,10	3,67	0,76			
13	11,00	14,31	11,00	14,16	0,01	11,00	13,58	0,05	11,00	12,33	0,14	11,00	10,93	0,24	11,60	9,56	0,33	13,70	8,16	0,43	14,30	6,03	0,58	14,10	4,50	0,69	14,20	3,39	0,76	13,60	3,25	0,77	13,60	3,23	0,77	13,60	3,17	0,78	13,60	3,12	0,78	13,60	3,15	0,78			
14	11,00	14,99	11,40	13,87	0,07	11,30	13,87	0,07	11,40	12,44	0,17	11,30	12,44	0,17	12,40	10,67	0,29	14,30	6,84	0,54	14,30	5,21	0,65	14,30	4,14	0,72	14,20	3,81	0,75	14,20	3,64	0,76	14,20	3,57	0,76	14,20	3,52	0,77	14,20	3,52	0,77	14,20	3,45	0,77			
15	12,30	13,06	11,30	12,46	0,05	11,30	10,74	0,18	11,20	8,06	0,38	13,70	4,66	0,64	14,30	3,99	0,69	14,20	3,23	0,77	14,00	3,00	0,77	14,00	2,98	0,77	14,00	2,98	0,77	14,00	2,87	0,78	14,00	2,76	0,79	14,00	2,76	0,79	14,00	2,76	0,79	14,00	2,76	0,79			
16	11,70	15,26	11,70	14,30	0,06	15,00	16,00	0,40	14,00	9,15	0,40	14,00	7,25	0,52	14,00	5,94	0,61	14,00	4,39	0,71	14,00	4,13	0,73	14,00	4,01	0,74	14,00	3,93	0,74	14,00	3,83	0,75	14,00	3,68	0,76	14,00	3,68	0,76	14,00	3,68	0,76						

Tabela A2 - Observações normalizadas das variáveis de processo X₁, X₂ e X₃, coletadas nos 15 intervalos de tempo, do estudo de caso realizado incluindo as observações para Batelada 16.

Bateladas	k=1			k=2			k=3			k=4			k=5			k=6			k=7			k=8			k=9			k=10			k=11			k=12			k=13			k=14			k=15		
	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃	X ₁	X ₂	X ₃
1	-0,84	0,09	-0,71	-0,17	-0,04	0,27	0,44	-0,56	0,22	0,70	-0,82	-0,30	0,94	-1,02	-0,65	1,38	-1,52	0,41	1,56	-1,61	0,21	1,95	-2,12	-0,07	2,46	-2,79	-0,42	2,18	-2,85	0,24	1,39	-1,88	-0,14	1,08	-1,35	0,46	0,94	-1,46	0,56	0,98	-1,19	0,57	0,76	-1,45	
2	-0,39	1,37	0,24	-1,12	1,54	0,44	0,62	0,07	-0,31	0,63	-0,21	0,47	0,52	-0,32	-0,75	0,40	-0,16	-0,70	0,07	0,11	0,21	0,23	0,03	-0,56	0,70	-0,28	-0,42	1,05	-0,50	0,24	0,45	0,33	-1,17	0,67	-0,04	0,46	0,72	-0,24	-0,04	0,76	-0,48	0,07	0,63	0,05	
3	0,50	-0,62	1,95	-1,49	0,67	0,96	-1,22	1,19	0,36	-0,93	0,89	0,66	-0,78	0,71	0,96	-0,93	0,87	0,68	-0,44	0,30	0,21	-0,72	0,56	-1,05	-0,85	0,84	-0,42	-0,88	0,68	-0,36	-0,69	0,33	0,38	-0,45	-0,04	-0,23	-0,33	-0,24	-0,04	-0,58	0,24	0,07	-0,60	0,05	
4	-0,28	0,48	0,05	-0,30	0,32	0,79	-1,50	1,99	0,36	-1,51	1,87	1,14	-1,11	1,26	0,86	-1,02	1,11	0,41	-1,26	1,35	0,21	-0,95	1,10	-0,07	-0,43	0,56	0,21	-0,11	0,29	0,24	-0,17	0,33	0,38	-0,16	0,61	0,46	-0,02	0,37	0,56	0,07	0,24	0,57	0,20	0,05	
5	1,50	0,57	1,19	0,58	-0,39	0,62	-0,38	0,71	0,36	0,50	-0,46	-0,30	0,06	0,00	0,86	0,26	-0,16	0,41	1,03	-0,94	0,21	0,78	-0,78	-0,07	-0,09	0,28	0,21	0,21	-0,10	0,24	0,71	-0,78	0,38	0,58	-0,70	0,46	0,75	-0,85	0,56	0,42	-0,48	0,57	0,26	0,05	
6	1,50	-1,07	1,38	-0,01	2,59	0,96	-1,49	1,19	0,36	-0,62	0,28	1,14	-1,28	1,18	0,86	-1,20	1,11	0,41	-1,40	1,25	0,21	-1,31	1,23	-0,07	-1,05	0,84	0,21	-0,73	0,29	0,24	-0,47	-0,22	0,38	-0,60	-0,04	0,46	-0,48	-0,24	0,56	-0,40	-0,48	0,57	-0,30	-0,70	
7	0,16	-0,21	1,00	0,76	-0,56	0,79	0,62	-0,09	0,76	-0,64	0,89	-0,21	-0,37	0,47	0,66	-0,84	0,95	-0,70	0,61	0,68	-1,38	-0,74	0,83	-0,07	-0,40	0,56	0,21	0,19	-0,10	0,24	0,23	-0,22	0,38	0,55	-0,70	-0,23	0,23	0,37	-1,84	0,32	0,24	-1,44	0,46	0,05	
8	-0,73	-0,58	-1,09	0,43	-1,09	-0,07	0,47	-1,04	0,89	-0,64	0,52	1,24	0,36	-0,55	0,76	-0,18	0,07	0,68	0,49	-0,66	0,21	0,72	-0,91	-0,56	0,57	-0,84	0,21	0,45	-0,89	0,24	0,20	-0,78	0,38	0,34	-0,70	-0,23	0,04	-0,24	-0,04	0,10	-0,48	0,57	0,23	-0,70	
9	-0,84	0,92	-0,33	0,24	0,14	0,79	0,09	0,39	0,22	-0,02	0,40	-0,49	-0,01	0,16	-0,35	-0,16	0,31	0,13	0,51	-0,37	-0,32	0,70	-0,51	-1,05	0,28	0,00	-1,68	0,66	-0,50	0,24	0,61	-0,22	-1,17	0,55	-0,04	-0,92	0,60	-0,24	-0,64	0,67	-0,48	-1,44	0,73	-0,70	
10	-1,06	0,74	-0,71	1,60	-1,09	0,96	1,00	-0,88	2,10	0,90	-0,82	0,85	0,36	-0,24	0,56	0,32	-0,24	-0,43	-0,02	0,11	-0,85	0,14	0,03	2,4																					

Tabela A3 - Observações normalizadas das variáveis de processo X1, X2 e X3 , coletadas até o instante k=9 para a Batelada 16.

Batelada	k=1		k=2			k=3			k=4			k=5			k=6			k=7			k=8		
	X1 ₁	X2 ₁	X1 ₂	X2 ₂	X3 ₂	X1 ₃	X2 ₃	X3 ₃	X1 ₄	X2 ₄	X3 ₄	X1 ₅	X2 ₅	X3 ₅	X1 ₆	X2 ₆	X3 ₆	X1 ₇	X2 ₇	X3 ₇	X1 ₈	X2 ₈	X3 ₈
16	1,1652	0,7850	0,6198	0,3196	-0,0350	5,2281	2,7937	4,2313	2,3709	-1,2926	1,7458	1,1427	-1,0266	1,1816	0,8615	-0,8923	1,0282	0,4071	-1,3642	1,4459	0,2125	-0,9538	1,0983
			k=9			k=10			k=11			k=12			k=13			k=14			k=15		
			X1 ₉	X2 ₉	X3 ₉	X1 ₁₀	X2 ₁₀	X3 ₁₀	X1 ₁₁	X2 ₁₁	X3 ₁₁	X1 ₁₂	X2 ₁₂	X3 ₁₂	X1 ₁₃	X2 ₁₃	X3 ₁₃	X1 ₁₄	X2 ₁₄	X3 ₁₄	X1 ₁₅	X2 ₁₅	X3 ₁₅
			-0,0657	-0,4487	0,8367	-0,0657	-0,4487	0,8367	-0,0657	-0,4487	0,8367	-0,0657	-0,4487	0,8367	-0,0657	-0,4487	0,8367	-0,0657	-0,4487	0,8367	-0,0657	-0,4487	0,8367

Tabela A4 - Observações normalizadas das variáveis de processo X1, X2 e X3 , coletadas até o instante k=15 para a Batelada 16.

Batelada	k=1		k=2			k=3			k=4			k=5			k=6			k=7			k=8		
	X1 ₁	X2 ₁	X1 ₂	X2 ₂	X3 ₂	X1 ₃	X2 ₃	X3 ₃	X1 ₄	X2 ₄	X3 ₄	X1 ₅	X2 ₅	X3 ₅	X1 ₆	X2 ₆	X3 ₆	X1 ₇	X2 ₇	X3 ₇	X1 ₈	X2 ₈	X3 ₈
16	1,1652	0,7850	0,6198	0,3196	-0,0350	5,228	2,7937	4,2313	2,3709	-1,2926	1,7458	1,1427	-1,0266	1,1816	0,8615	-0,8923	1,0282	0,4071	-1,3642	1,4459	0,2125	-0,9538	1,0983
			k=9			k=10			k=11			k=12			k=13			k=14			k=15		
			X1 ₉	X2 ₉	X3 ₉	X1 ₁₀	X2 ₁₀	X3 ₁₀	X1 ₁₁	X2 ₁₁	X3 ₁₁	X1 ₁₂	X2 ₁₂	X3 ₁₂	X1 ₁₃	X2 ₁₃	X3 ₁₃	X1 ₁₄	X2 ₁₄	X3 ₁₄	X1 ₁₅	X2 ₁₅	X3 ₁₅
			-0,0657	-0,4487	0,8367	0,2098	-0,0472	0,2877	0,2378	0,1221	0,3325	0,3771	-0,0099	0,6085	0,4606	0,1359	0,3661	0,5597	0,2288	0,2386	0,5680	0,3635	0,0500

Tabela A5 - Valores calculados dos escores y_k , QEP, T^2 e seus respectivos LC até o instante $k=15$ para a Batelada 16.

k	y_1	LCI	LCS	y_2	LCI	LCS	y_3	LCI	LCS	y_4	LCI	LCS	y_5	LCI	LCS	y_6	LCI	LCS	y_7	LCI	LCS	y_8	LCI	LCS	T^2	LCS	QEP	LCS
1	0,9500	-3,21	3,21	2,0799	-3,21	3,21	2,3055	-3,21	3,21	0,6608	-3,21	3,21	1,8368	-3,21	3,21	1,9686	-3,21	3,21	-0,1495	-3,21	3,21	0,46949	-3,21	3,21	4,3976	63,66	0,3431	0,0177
2	-0,1107	-3,21	3,21	0,6703	-3,21	3,21	0,7681	-3,21	3,21	-0,4406	-3,21	3,21	0,3029	-3,21	3,21	0,8418	-3,21	3,21	-0,8001	-3,21	3,21	-0,18902	-3,21	3,21	1,4823	63,66	0,0970	0,2732
3	0,9844	-3,21	3,21	4,9222	-3,21	3,21	5,5562	-3,21	3,21	2,4719	-3,21	3,21	-0,8438	-3,21	3,21	5,9551	-3,21	3,21	4,6613	-3,21	3,21	5,78042	-3,21	3,21	84,8051	63,66	18,2759	0,0765
4	1,7234	-3,21	3,21	2,4229	-3,21	3,21	2,0808	-3,21	3,21	0,2763	-3,21	3,21	-0,9897	-3,21	3,21	2,6113	-3,21	3,21	0,4984	-3,21	3,21	1,13914	-3,21	3,21	13,5961	63,66	1,0672	0,0679
5	1,1556	-3,21	3,21	2,6441	-3,21	3,21	0,8987	-3,21	3,21	0,3230	-3,21	3,21	-0,3326	-3,21	3,21	2,0303	-3,21	3,21	-0,7820	-3,21	3,21	1,59382	-3,21	3,21	6,1091	63,66	0,0662	0,0121
6	1,0288	-3,21	3,21	3,0800	-3,21	3,21	0,5970	-3,21	3,21	0,2018	-3,21	3,21	-0,2284	-3,21	3,21	1,7204	-3,21	3,21	-1,2626	-3,21	3,21	1,46638	-3,21	3,21	5,1839	63,66	0,9951	0,0123
7	1,3597	-3,21	3,21	1,3507	-3,21	3,21	-0,0826	-3,21	3,21	0,7823	-3,21	3,21	0,0851	-3,21	3,21	1,9818	-3,21	3,21	-0,6972	-3,21	3,21	1,60419	-3,21	3,21	4,7539	63,66	0,1227	0,0209
8	1,1628	-3,21	3,21	1,4765	-3,21	3,21	-0,1795	-3,21	3,21	0,4498	-3,21	3,21	0,0527	-3,21	3,21	2,1139	-3,21	3,21	-0,9102	-3,21	3,21	1,59035	-3,21	3,21	4,2226	63,66	0,0396	0,2007
9	0,9534	-3,21	3,21	1,7251	-3,21	3,21	-0,4171	-3,21	3,21	0,1522	-3,21	3,21	0,1942	-3,21	3,21	1,7176	-3,21	3,21	-0,8403	-3,21	3,21	1,70708	-3,21	3,21	3,7025	63,66	1,5521	0,1863
10	0,7912	-3,21	3,21	1,5627	-3,21	3,21	-0,0667	-3,21	3,21	0,1419	-3,21	3,21	0,1414	-3,21	3,21	1,7007	-3,21	3,21	-0,7995	-3,21	3,21	1,25510	-3,21	3,21	2,7653	63,66	0,9625	0,0738
11	0,7777	-3,21	3,21	1,4744	-3,21	3,21	-0,0388	-3,21	3,21	0,1917	-3,21	3,21	0,1785	-3,21	3,21	1,7316	-3,21	3,21	-0,7005	-3,21	3,21	0,94735	-3,21	3,21	2,8178	63,66	0,1859	0,0025
12	0,8406	-3,21	3,21	1,4124	-3,21	3,21	0,0445	-3,21	3,21	0,1886	-3,21	3,21	0,0577	-3,21	3,21	1,8247	-3,21	3,21	-0,7454	-3,21	3,21	1,02319	-3,21	3,21	3,1979	63,66	0,1054	0,0027
13	0,8046	-3,21	3,21	1,4610	-3,21	3,21	0,1095	-3,21	3,21	0,1971	-3,21	3,21	0,1071	-3,21	3,21	1,7171	-3,21	3,21	-0,5844	-3,21	3,21	0,96221	-3,21	3,21	2,8703	63,66	0,0738	0,0008
14	0,7873	-3,21	3,21	1,4871	-3,21	3,21	0,1574	-3,21	3,21	0,1957	-3,21	3,21	0,1159	-3,21	3,21	1,6964	-3,21	3,21	-0,4969	-3,21	3,21	0,95159	-3,21	3,21	2,7317	63,66	0,1911	0,0066
15	0,7769	-3,21	3,21	1,5052	-3,21	3,21	0,1618	-3,21	3,21	0,1958	-3,21	3,21	0,1272	-3,21	3,21	1,6391	-3,21	3,21	-0,4701	-3,21	3,21	0,89357	-3,21	3,21	2,6197	63,66	0,5928	0,0051