



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA



ENG07053 - TRABALHO DE DIPLOMAÇÃO EM ENGENHARIA QUÍMICA

Monitoramento Estatístico de Processos Baseado na Análise de Componentes Principais (PCA) e Cartas de Controle

Rafael Ramos Wagner

Orientador: Prof. Marcelo Farenzena

Porto Alegre, Julho de 2011

AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul e ao Departamento de Engenharia Química, pelo ensino de qualidade e pela infra-estrutura.

Ao Professor Marcelo Farenzena, pela disponibilidade sempre que necessário, pelas boas idéias que conduziram o trabalho e pelos conselhos.

Ao Péricles por me ceder os dados das suas simulações e da coluna de separação de Etilbenzeno.

Aos meus amigos de faculdade que sempre estiveram juntos comigo ajudando no trabalho e a relaxar nos churrascos da turma.

Aos meus amigos de Osório que sempre me deram força, e entenderam quando eu não podia ir para Osório fazer festa com eles.

E a minha família que sempre esteve do meu lado me dando suporte e torcendo pelo meu sucesso.

SUMÁRIO

1	Introdução.....	1
2	Revisão Bibliográfica	3
2.1	Análise de Componentes Principais (PCA)	3
2.2	Cartas de Controle.....	7
2.2.1	Cartas de Controle Univariadas (CC).....	7
2.2.2	Cartas de Controle Multivariadas (CCM).....	9
2.2.3	Cartas de Controle Multivariadas de Hotelling	9
2.2.4	Cartas de Controle Multivariadas de Hotelling Baseadas em Componentes Principais (CCP)	12
3	Metodologia de Obtenção e Tratamento de Dados.....	14
3.1	Caso de Estudo	14
3.2	Simulação do Processo	19
3.3	Tratamentos dos Dados.....	20
4	Resultados.....	21
4.1	Obtenção das Componentes Principais do Sistema.....	21
4.2	Detecção Descontroles do Processo Através da Carta de Controle Multivariada de Hotelling – T^2	25
5	Conclusão.....	28
	Referências.....	29
	Apendice A.....	30

LISTA DE FIGURAS

Figura 2.1: (a) Relação entre as variáveis X1 e X2; (b) Relação entre os componentes Z1 e Z2.....	4
Figura 2.2: Carta de Controle de Shewhart	8
Figura 3.1: Reação de alquilação do benzeno.....	14
Figura 3.2: Representação da coluna de separação de etilbenzeno.....	15
Figura 3.3: Números de desvios padrão em relação à média amostral para a vazão e temperatura de entrada coletadas da coluna de separação de etilbenzeno em operação.....	17
Figura 3.4: Números de desvios padrão em relação à média amostral para as temperaturas de fundo e topo coletadas da coluna de separação de etilbenzeno em operação.	17
Figura 3.5: Números de desvios padrão em relação à média amostral para a vazão e pressão de fundo coletadas da coluna de separação de etilbenzeno em operação.	18
Figura 3.6: Números de desvios padrão em relação à média amostral para as temperaturas do prato 55 e prato 30 coletadas da coluna de separação de etilbenzeno em operação.	18
Figura 3.7: Dados obtidos das simulações da coluna de separação de etilbenzeno.....	20
Figura 4.1: Comparação entre as variáveis reais normalizadas e as variáveis normalizadas obtidas através das componentes principais.	24
Figura 4.2: Carta de controle T^2 de Hotelling para a primeira fase dos calculos do limite de controle superior (LCS)	26
Figura 4.3: Carta de controle T^2 de Hotelling baseada em PCA para os dados simulados da coluna de separação de etilbenzeno	27
Figura A.1: Comando em MatLab para a obtenção dos PCs.....	30
Figura A.2: Comandos em MatLab para o cálculo da estatística T^2	31

LISTA DE TABELAS

Tabela 4.1: Desvio padrão das variáveis monitoradas do sistema	22
Tabela 4.2: Variância capturada para cada PC calculadas para dados normalizados e dados não normalizados	23
Tabela 4.3: Matriz de correlação (C^*) para os dados amostrais.....	23
Tabela 4.4: Autovetores obtidos a partir da matriz de correlação (C^*)	24
Tabela 4.5: Autovalores e proporção de variância do sistema associadas aos PCs obtidos	25

RESUMO

As cartas de controle multivariadas (CCMs) são amplamente utilizadas no monitoramento de desempenho de processos. Com a crescente informatização e automatização dos processos industriais, tem-se verificado um aumento sensível na quantidade de informações disponíveis sobre variáveis de processo. Este aumento no número de variáveis monitoradas dificulta o monitoramento com as CCMs. Nesses casos as cartas de controle multivariadas baseadas em componentes principais (CCPs) são utilizadas, pois, monitoram simultaneamente diversas variáveis controlando somente algumas combinações lineares independentes delas. As CCPs são utilizadas principalmente para monitorar variáveis que influenciam na qualidade de um produto final ou na estabilidade do sistema.

O objetivo deste trabalho é aplicar a técnica de Análise de Componentes Principais (PCA), reduzindo o sistema de variáveis, para a posterior utilização das CCPs no monitoramento de distúrbios no processo estudado. Mostrar-se-á a aplicabilidade e a eficiência desta técnica para processos que envolvam um número elevado de variáveis que devem ser monitoradas para garantir a sua estabilidade e a qualidade do produto final.

O processo estudado foi a coluna de separação de etilbenzeno, pertencente à Innova-PETROBRAS, que opera de forma a obter etilbenzeno com pureza igual ou maior que 99,6% na sua corrente de topo. A análise da composição da corrente de topo é feita a cada quatro horas gerando atrasos na correção de parâmetros e eventualmente produzindo produto fora da especificação. Para a análise de componentes principais e para a determinação do limite de controle da carta de controle, foram utilizados dados reais da planta operando dentro da especificação do produto de topo. Porém para a validação da carta de controle foram gerados dados de simulação da coluna, em *Aspen Plus Dynamics*, onde foram inseridos distúrbios em algumas variáveis.

A carta de controle multivariada T^2 de Hotelling baseada na análise de componentes principais foi utilizada para detectar as distúrbios simuladas na operação da coluna e os resultados obtidos foram satisfatórios visto que 100% das distúrbios foram detectadas pela carta de controle gerando apenas dois alarmes falsos, totalizando 3% dos dados da mostra.

1 INTRODUÇÃO

A partir das medidas usuais de processos industriais automatizados uma grande quantidade de informações sobre seu desempenho pode ser gerada. Nestes processos são feitas medições e em tempo real de diversas variáveis de desempenho como temperatura, pressão, vazão, etc. Com esta quantidade elevada de dados podemos utilizar ferramentas para o monitoramento do desempenho destes processos. Estas ferramentas têm a finalidade de transformar estes dados de processo em conhecimento que serão de suma importância para o entendimento do funcionamento do processo.

As cartas de controle (CCs) univariadas de Shewhart são cartas de controle que monitoram uma única variável limitada pelos limites de controle superior (LCS) e inferior (LCI) e são comumente utilizadas no controle estatístico de processos industriais (Montgomery, 1996; Duncan, 1986). Essas cartas utilizam como dados de entrada medições de variáveis que influenciam o desempenho do processo em questão. A cada medição, compara-se o resultado obtido com limites de controle. Medições fora dos limites indicam a presença de causas especiais de variabilidade, anômalas ao processo, que prejudicam seu desempenho. Uma vez identificadas as causas especiais, pode-se atuar sobre elas, melhorando continuamente o desempenho do processo. As CCs permitem a redução sistemática da variabilidade nas características críticas de processo.

A estratégia de controle univariado de Shewhart, acrescida de algumas heurísticas, pode ser estendida para o caso multivariado. Cartas de Controle Multivariadas (CCMs) são indicadas em situações onde variáveis de processo são monitoradas simultaneamente. O tipo de CCM de utilização mais comum é a carta de Hotelling, ou carta T^2 (Hotelling, 1947).

Algumas situações especiais dificultam ou impossibilitam a utilização das CCMs tradicionais. Com a crescente automatização de processos, um grande número de variáveis é medida em tempo real e em pontos temporais próximos; nessas condições, a suposição de independência dificilmente é verificada. Outra situação especial ocorre quando a variação normal de uma variável monitorada ao longo do tempo é descrita por um perfil. As CCMs não podem ser aplicadas nessa situação, já que pressupõem variáveis com média constante.

Neste trabalho, apresenta-se uma ferramenta 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 (CCPs). Essas cartas são operacionalizadas utilizando a análise de componentes principais e as CCMs. As CCPs admitem pontos amostrais sucessivos dependentes, além de permitir o monitoramento de variáveis que não apresentam valores fixos de média ao longo do tempo.

A Análise de Componentes Principais (PCA) 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. Matematicamente, as Componentes Principais (PC) são combinações lineares independentes das variáveis aleatórias (Montgomery, 1996).

O objetivo das CCPs consiste em utilizar uma estratégia de controle similar àquela das CCMs utilizando, ao invés das variáveis de processo, PCs delas derivadas. Desta forma, simplifica-se a tarefa de monitoramento de processos, já que este passa a ser feito sobre um número menor de variáveis. A principal vantagem deste método de monitoramento é que não se faz necessário o conhecimento de nenhum modelo matemático do sistema, apenas dados são necessários para a criação da carta de controle e seus limites de controle.

O presente trabalho está dividido em 5 capítulos. O capítulo 2 apresenta uma revisão bibliográfica do método PCA e das cartas de controle multivariadas. O capítulo 3 descreve o estudo de caso escolhido, a metodologia de obtenção dos dados para análise, o tratamento destes dados e os softwares envolvidos. No capítulo 4 é aplicada a técnica de PCA e posteriormente é aplicada a CCP de Hotelling, apresentando resultados e discussões sobre os mesmo. No capítulo 5 as principais conclusões deste trabalho são elucidadas.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo, é feita uma breve introdução sobre o procedimento Análise de Componentes Principais (PCA) aplicadas e redução de variáveis do processo. Apresenta-se uma revisão das cartas de controle tradicionais (CCs) e das cartas de controle multivariadas (CCMs), mais especificamente a carta de controle T^2 de Hotelling.

2.1 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

A Análise de Componentes Principais, também conhecida como *Principal Component Analysis* (PCA) é uma técnica estatística utilizada para reduzir a dimensionalidade de um conjunto de dados onde há um grande número de variáveis inter-relacionadas (Warne et al. 2004). Isto é feito de forma que o máximo de variância presente nos dados seja mantido. Essa redução se dá pela obtenção de um novo conjunto de variáveis não correlacionadas, chamadas Componentes Principais (PCs) (Mingoti 2005). As PCs são ortogonais entre si, ou seja, a primeira PC capta as informações na direção onde existe a maior variância, a segunda PC irá captar as informações em um sentido ortogonal a primeira PC e com a maior variância possível e assim sucessivamente, de modo que a PC $i+1$ seja ortogonal a sua anterior i . Tais PCs são então ordenadas de forma que os primeiros guardam a maior parte da variância presente nas variáveis originais.

Segundo Jolliffe (2002), esta técnica é, provavelmente, a mais antiga técnica de análise de dados multivariados. As primeiras descrições da técnica que hoje é conhecida como PCA, foram dadas por Pearson (1901) e Hotelling (1933). Mesmo utilizando abordagens diferentes, Pearson e Hotelling chegaram à derivação de PCA: Pearson, na tentativa de encontrar retas e planos que melhor se ajustavam a um conjunto de pontos em um espaço p -dimensional e, mais tarde, Hotelling ao procurar um menor conjunto fundamental de variáveis independentes que determinassem os valores das p variáveis originais. Pearson notou que o PCA só era viável para um número de p menor ou igual a quatro, pois demandava muitos cálculos matriciais. Por isso, novos estudos sobre a técnica foram publicados somente em 1963 com a utilização de computadores para os cálculos.

Muitas pesquisas ainda estão sendo realizadas na área de PCA. Além disso, PCA tem sido amplamente utilizado e aplicado em diversas áreas com, biologia, química, engenharia, economia e outros (Hair, Jr. et al., 1995).

A PCA é uma técnica estatística utilizada para resumir informações contidas em conjuntos de dados multivariados. Considere um conjunto de dados composto de

realizações de p variáveis aleatórias. Tal conjunto de dados possui p PCs, sendo cada PC formado por uma combinação linear distinta e ortogonal entre si das p variáveis aleatórias originais.

Essas combinações são determinadas através de manipulações algébricas na matriz de covariâncias ou de correlação das p variáveis, conforme apresentado adiante. A opção por uma dessas duas matrizes normalmente é feita de acordo com a discrepância causada pelas diferentes unidades de medida das variáveis do sistema. Cada PC captura uma direção de variabilidade do conjunto de dados originais. As direções capturadas por cada PC são ortogonais entre si, o que equivale a dizer que os PCs são variáveis aleatórias independentes. A Figura 2.1-a mostra um gráfico de duas variáveis altamente correlacionadas X_1 e X_2 , onde uma variação considerável pode ser vista em ambas. Transformando para os PCs Z_1 e Z_2 , obtemos o gráfico da Figura 2.1-b onde fica claro que há mais variação na direção de Z_1 do que em qualquer uma das variáveis originais, mas pouca variação na direção Z_2 . De forma geral, se um conjunto de p (>2) variáveis tem correlações substanciais entre si, então os primeiros PCs esclarecerão a maior parte da variação das variáveis originais. Da mesma forma, os últimos PCs identificam direções em que há pouca variação, ou seja, eles identificam relações lineares quase constantes entre as variáveis originais.

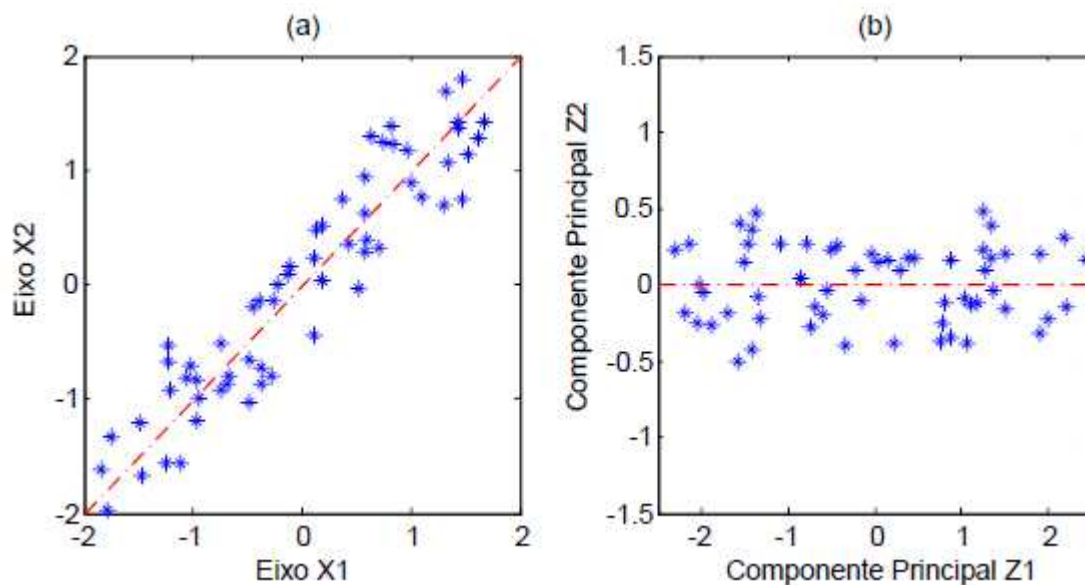


Figura 2.1: (a) Relação entre as variáveis X_1 e X_2 ; (b) Relação entre os componentes Z_1 e Z_2 .

O objetivo da determinação de PCs 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 PCs. Assim, pode-se reduzir a dimensionalidade do conjunto de

dados (isto é, o número de variáveis aleatórias 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.

Conforme mencionado acima, PCs são combinações lineares de variáveis aleatórias, identificadas por X_i , $i = 1, \dots, p$, com observações designadas por $(x_{i1}, x_{i2}, \dots, x_{in})$. Essas combinações são obtidas a partir da matriz de covariâncias ou de correlação, associadas às p variáveis aleatórias. A determinação algébrica dos PCs é apresentada na seqüência.

A análise de componentes principais começa com dados de p variáveis para n observações. O primeiro PC é então a combinação linear das variáveis X_1, X_2, \dots, X_p , mostrada na equação (2.1).

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (2.1)$$

Desta forma o segundo PC é dado pela equação (2.2), assim como os demais componentes principais.

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (2.2)$$

Se existem p variáveis então existiriam no máximo p PCs. De fato para obter uma PCA é necessário determinar os autovalores de uma matriz de covariância amostral, C . A matriz de covariâncias é simétrica e tem a forma:

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{pmatrix}$$

Em que o elemento c_{ij} na diagonal é a variância de X_i e o termo fora da diagonal c_{ij} é a covariância entre as variáveis X_i, X_j . As variâncias dos PCs são os autovalores da matriz C . Existem p autovalores, alguns dos quais podem ser zero. Autovalores negativos não são possíveis para uma matriz de covariância. Nos casos em que as PCs obtidas são influenciadas pelas variáveis de maior variância, costuma-se utilizar a matriz de correlação à matriz de covariância, pois a matriz de correlação normaliza os dados para a sua obtenção, eliminando os problemas de variáveis com unidades diferentes e ordens de grandeza muito diferentes. Para isso é necessário realizar a normalização dos dados amostrais. Esta normalização é feita através da equação (2.3), tornando os dados com media zero e variância unitária, gerando a Matriz de Correlação, C^* .

$$x_{ij}^* = \frac{x_{ij} - x_m}{\sigma} \quad (2.3)$$

Onde x_{ij}^* é o valor normalizado da variável $i, i=1, 2 \dots p$, na sua j -ésima observação, $j=1, 2, \dots, n$; x_m e σ são a média e o desvio padrão das n observações de x_i respectivamente.

A matriz de correlação também é simétrica e tem a forma:

$$C^* = \begin{vmatrix} 1 & c_{12}^* & \dots & c_{1p}^* \\ c_{21}^* & 1 & \dots & c_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1}^* & c_{p2}^* & \dots & 1 \end{vmatrix}$$

Com a matriz de correlação podemos calcular os autovalores utilizando a equação característica (2.4) da matriz C^* .

$$|C^* - \lambda I| = 0 \quad (2.4)$$

Onde λ é o conjunto de escalares, designados por $(\lambda_1, \lambda_2 \dots \lambda_p)$, que tornam a equação (2.4) verdadeira e são chamados de autovalores da matriz C^* e I é a matriz identidade da matriz C^* . De posse dos autovalores calculam-se os autovetores e também a proporção da variância descrita por cada PC aplicando a equação (2.5) e (2.6) respectivamente para cada um dos autovalores encontrados em (2.4).

$$|C^* - \lambda_i I| \gamma_i = 0 \quad (2.5)$$

$$R_i = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (2.6)$$

Onde γ_i são os i autovetores, com $i = 1, \dots, p$, da matriz C^* e R_i é a porção da variância representada pelo autovetor relacionado ao $i^{\text{és}}$ autovalor. Cada autovetor representa os coeficientes dos PCs, $(a_{i1}, a_{i2}, \dots, a_{ip})$, descritos na equação (2.1) e (2.2) acima.

Segundo Mingoti (2005), para se reduzir a dimensão de p para k PCs, deve-se adotar um critério de escolha. Uma forma comum de se escolher as k PCs é através da porcentagem K da variância total, em que $0 < K < 1$. Assim sendo, busca-se o menor valor de k que satisfaça a equação (2.7) para o valor de K desejado. Normalmente na literatura usam-se valores de K em torno de 90 a 95%.

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^p \lambda_j} \geq K \quad (2.7)$$

2.2 CARTAS DE CONTROLE

2.2.1 CARTAS DE CONTROLE UNIVARIADAS (CC)

As cartas de controle são registros dos dados mensurados em um ponto crítico do processo, construídos num sistema de coordenadas cartesianas (Siqueira, 1997). Assim, o eixo das ordenadas é representado pelas mensurações feitas de uma determinada característica do produto e o eixo das abscissas representado pelos subgrupos. A Figura 2.2 ilustra a carta de controle de Shewhart, o qual apresenta três linhas paralelas ao eixo da abscissa, denominadas de linha central (LC) que é comumente definida a partir do valor médio, o limite superior de controle (LSC) e o limite inferior de controle (LIC).

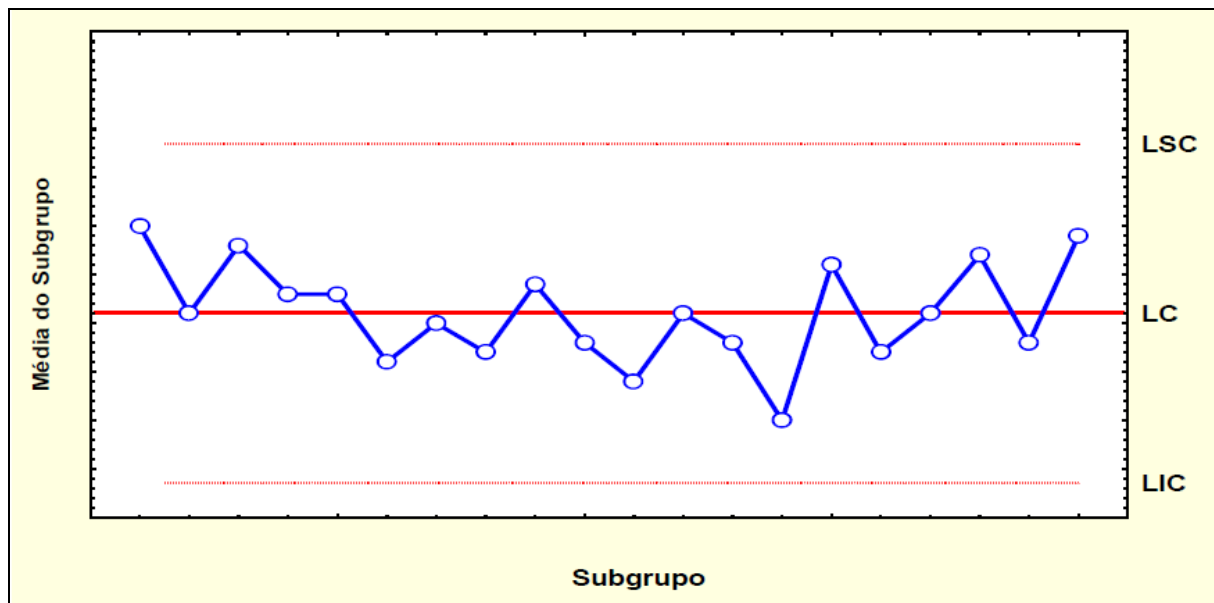


Figura 2.2: Carta de Controle de Shewhart.

As cartas de controle devem ser implementadas para monitorar as variáveis que influenciam a qualidade dos itens produzidos. Dessa forma, uma determinada variável de qualidade ou característica de qualidade é monitorada por meio de sucessivas amostras espaçadas no tempo (Siqueira, 1997). As medições obtidas em cada amostra são representadas em ordem cronológica num gráfico e comparadas com os limites de controle obtidos da amostra. Caso alguma medição ultrapasse os limites de controle, o referido processo é considerado fora dos limites de controle estatístico. Se uma seqüência de medições apresentarem um comportamento aleatório, mesmo que nenhuma medição ultrapasse os limites de controle do processo, o referido processo também poderá ser considerado fora dos limites de controle (Siqueira, 1997).

Para uma primeira fase de implantação do controle estatístico de processos, deve-se monitorar uma amostragem que sirva de referência para as etapas posteriores. Isso significa que, após verificar as condições da matéria-prima, equipamentos, treinamento de pessoal, coleta-se uma amostra e verifica-se, se esta apresenta um comportamento sob controle. Caso o processo já esteja sob controle, pode-se passar para a segunda fase, na qual são realizadas outras coletas de amostras para proceder-se o monitoramento contínuo, caso apresente um estado fora dos limites de controle, deve-se investigar o processo até que a amostra do processo se normalize.

2.2.2 CARTAS DE CONTROLE MULTIVARIADAS (CCM)

As cartas de controle multivariadas monitoram processos onde um grande número de variáveis apresenta correlação. As cartas de controle multivariadas (CCM) são utilizadas no monitoramento conjunto de duas ou mais variáveis do processo ou produto. Nestas cartas os pontos p -dimensionais (ou seja, são valores das p variáveis aleatórias ou estatísticas de interesse delas derivadas) são representados uni-dimensionalmente em cartas similares as cartas de controle de Shewhart, simplificando a tarefa de controle simultâneo de variáveis.

Em situações onde exista correlação significativa entre as variáveis a serem monitoradas as CCM são particularmente recomendados, pois, em processos multivariados, as variáveis interferem e sofrem interferências entre si, possuindo assim uma forte correlação. No caso em que as variáveis não sejam fortemente correlacionadas, e seja aplicada a CCM, há possibilidade de surgirem alarmes falsos no processo investigado, já que o operador pode constatar problemas em uma determinada variável que não esteja necessariamente interferindo no processo; ou também pode considerar que o processo está sob controle sem que o mesmo esteja, em função desta estar trabalhando com variáveis que não interferem no mesmo, mas que influenciam na média e desvio-padrão. Assim, a carta T^2 de Hotelling, descrita na seqüência, é uma carta de controle do tipo Shewhart; no que diz respeito à detecção de grandes variações no processo, bem como na verificação do comportamento do processo através da média. Em processos químicos é comum a existência de variáveis correlacionadas como, por exemplo, as pressões, temperaturas e composições de uma coluna de destilação que são fortemente correlacionadas.

2.2.3 CARTAS DE CONTROLE MULTIVARIADAS DE HOTELLING

A carta de controle T^2 foi desenvolvida por Hotelling (1947), o pioneiro na pesquisa sobre as cartas de controle multivariado. Ele utilizou-se da abordagem multivariada de controle em dados contendo informações sobre localizações de bombardeios na segunda guerra mundial. O desenvolvimento teórico proposto por Hotelling é descrito a seguir.

Em muitos casos a média (X_m) e a Matriz de Covariância (C) das variáveis envolvidas no processo não são conhecidas, nestes casos freqüentemente estas são estimadas através de amostras preliminares recolhidas do processo, quando este está sob controle estatístico.

Considerando as estimativas dos parâmetros, a equação será expressa na seguinte forma:

$$T^2 = n(X - X_m)'C^{-1}(X - X_m) \quad (2.8)$$

Onde X_m e C representam, respectivamente, as estimativas para o vetor de médias e matriz de covariâncias do processo, sendo obtidos na seguinte forma:

$$X_m = \frac{1}{c} \sum_{k=1}^c X_k \quad (2.9)$$

$$C = \frac{1}{c} \sum_{k=1}^c C_k \quad (2.10)$$

Com

$$C_k = \frac{1}{n-1} \sum_{i=1}^n (X_i - X_k)(X_i - X_k) \quad (2.11)$$

Onde c representa o número de amostras preliminares de tamanho n retiradas do processo e X_i é a i -ésima observação da p variável referente à k -ésima amostra. A expressão estatística (2.8) é utilizada como base para o gráfico de controle T^2 de Hotelling (Lowry & Montgomery, 1995). Os limites de controles para esta carta de controle são definidos em duas fases. Na primeira fase são utilizados os dados do processo sob controle estatístico para determinação do primeiro limite. A equação (2.12) é utilizada nesta primeira fase para calcular o limite superior de controle (LSC). O limite inferior de controle (LIC) é igual a zero, pois a estatística T^2 representa a distância quadrada entre os dados reais e a posição estatística calculada para estes dados, desta forma o menor valor possível para T^2 é zero.

$$LSC = \frac{p(c-1)(n-1)}{cn-c-p+1} F_{\alpha, p, (cn-c-p-1)} \quad (2.12)$$

$$LIC = 0$$

Onde $F_{\alpha, p, (cn-c-p-1)}$ representa o percentil da distribuição F com $p=v_1$ e $(cn-c-p+1)=v_2$ graus de liberdade, obtido a partir da probabilidade α de alarme falso adotada. Após o

cálculo dos limites de controle da primeira fase as estatísticas T^2 são calculadas para os dados sobre controle estatístico e as curvas dispostas em um gráfico (carta de controle). Os dados que ultrapassarem o limite superior de controle são eliminados para um melhor ajuste no cálculo dos limites de controle da segunda fase. Na segunda fase novos limites são estabelecidos apenas para monitorar as observações futuras, utilizando os limites de controle mostrados na equação (2.13).

$$LSC = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,(mn-m-p-1)} \quad (2.13)$$

$$LIC = 0$$

Onde m é o número de observações da amostra que representa o processo sob controle estatístico, filtradas a partir das c amostras. Nesta fase $c-m$ amostras foram descartadas por não caracterizarem o padrão das variáveis no processo, recalcula-se X_m e C para monitoramento das amostras futuras substituindo c por m nas expressões (2.9) e (2.10).

Lowry e Montgomery (1995) apresentaram tabelas indicando o número mínimo recomendado de m amostras preliminares para tamanhos de $n=3, 5$ e 10 e para $p=2, 3, 4, 5, 10$ e 20 variáveis, para que os limites de controle da segunda fase sejam bem aproximados, através dos limites de controle obtidos. Os valores recomendados para m são sempre maiores que 50 . Logo, quando o número de variáveis p aumenta, maior o valor de m necessário para uma boa aproximação dos limites de controle da segunda fase.

Em algumas situações, o processo é monitorado utilizando amostras de tamanho unitário ($n=1$). Nestes casos, não é possível a seleção de subgrupos homogêneos de tamanho grande. Portanto, não faz sentido agrupar uma seqüência de observações consecutivas das variáveis, pois se verifica, a cada instante, se as variáveis estão seguindo as trajetórias especificadas para um determinado processo. Neste caso os limites de controle da primeira fase são determinados pela equação (2.14) para $n=1$ (Tracy et AL, 1992).

$$LSC = \frac{(c-1)^2}{c} \beta_{\alpha,p/2,(c-p-1)/2} \quad (2.14)$$

$$LIC = 0$$

Onde $\beta_{(\alpha, p/2, (c-p-1)/2)}$ representa o percentil de uma distribuição β com $p/2$ e $(c-p-1)/2$ graus de liberdade, obtido a partir da probabilidade α de alarme falso adotado. Valores na distribuição *Beta* podem ser aproximados pela distribuição *F*. Assim, os limites de controle podem ser determinados pela equação (2.15) da seguinte forma (Lowry & Montgomery, 1995):

$$LSC = \frac{(c-1)^2}{c} \frac{\frac{p}{(c-p-1)} F_{\alpha, p, (c-p-1)}}{1 + \frac{p}{(c-p-1)} F_{\alpha, p, (c-p-1)}} \quad (2.15)$$

$$LIC = 0$$

O procedimento para a determinação dos limites de controle da segunda fase permanece o mesmo, porém utiliza-se a equação (2.16)

$$LSC = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, (m-p)} \quad (2.16)$$

$$LIC = 0$$

Na engenharia de processos, este procedimento é bastante comum, pois os dados são analisados em tempo real, e suas estatísticas calculadas imediatamente, para que se possa verificar a tendência do processo em relação aos seus limites de controle.

2.2.4 CARTAS DE CONTROLE MULTIVARIADAS DE HOTELLING BASEADAS EM COMPONENTES PRINCIPAIS (CCP)

A estatística de Hotelling, descrita anteriormente, não deve ser aplicada para conjuntos muito grandes de variáveis correlacionadas, pois foi concebida para monitorar um número limitado de variáveis de qualidade do produto final. Uma estrutura muito grande de variáveis favorece o surgimento de variáveis colineares, isto é, variáveis que praticamente duplicam informações contidas nas demais e trazem pouca informação. A estrutura colinear gerada a partir dessas redundâncias faz com que a estatística forneça sinalizações

distorcidas acerca do estado do processo, bem como diagnósticos imprecisos de causas especiais. Em algumas situações, tal estatística não poderá ser calculada devido à dificuldade de inversão da matriz de covariâncias das variáveis de processo (Kourti & McGregor, 1996). Para eliminar estas redundâncias, surge então a necessidade de utilizar cartas de controle baseadas em métodos de projeção de dados, dentre os quais se destaca a Análise de Componentes Principais (PCA). O método de PCA está descrito no capítulo 2.1.

Pode-se escrever a estatística de Hotelling tradicional na equação (2.17) em função das q PCs, selecionadas para representar o sistema, da seguinte forma (Kourti & McGregor, 1996):

$$T^2 = \sum_{i=1}^q \frac{Z_i^2}{\lambda_i} \quad (2.17)$$

Onde Z_i é o escore da i -ésima PC e q ($< p$) é o número de PCs selecionados na etapa de redução de variáveis do procedimento de obtenção das PCs. O procedimento para o cálculo dos limites de controle das cartas permanece o mesmo.

3 METODOLOGIA DE OBTENÇÃO E TRATAMENTO DE DADOS

3.1 CASO DE ESTUDO

O caso estudado neste trabalho, é a coluna de separação de etilbenzeno pertencente a Innova-PETROBRAS. O etilbenzeno é utilizado para a obtenção do monômero estireno, que por sua vez é utilizado para a produção de poliestireno. A reação de obtenção do etilbenzeno é feita em fase líquida através da reação de alquilação do benzeno com eteno. Esta reação está representada na Figura 3.1.

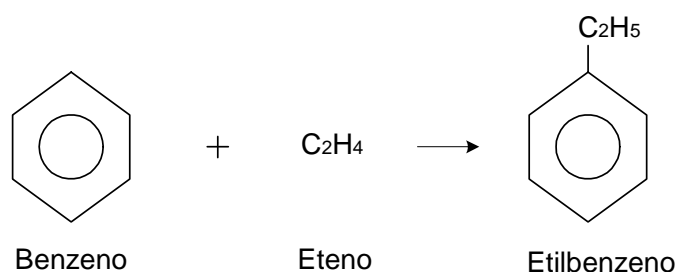


Figura 3.1: Reação de alquilação do benzeno.

O sistema de reação é formado por dois reatores em série que utilizam zeólitas como catalisador. Os reagentes são previamente tratados para a remoção de contaminantes e venenos para o catalisador. Além desta reação o etilbenzeno ainda pode sofrer novas alquilações formando compostos chamados de polietilbenzenos. Os produtos da reação de alquilação, que são basicamente benzeno não reagido, etilbenzeno, compostos polialquilados e compostos de alto peso molecular são separados através de um conjunto formado por 4 colunas de destilação.

A primeira coluna separa o benzeno excedente no topo e os produtos da reação pela corrente de fundo. A segunda coluna retira a água do benzeno que volta para a reação em uma corrente de reciclo. A terceira coluna separa o etilbenzeno dos demais compostos formados nas reações de alquilação e transalquilação, esta será a coluna onde o procedimento proposto neste trabalho será estudado. Por fim a quarta coluna separa os polietilbenzenos dos demais compostos pesados formados na reação, estes pesados são utilizados como combustível nos fornos.

O etilbenzeno é o produto final deste processo e é a matéria prima para a planta de estireno. A terceira coluna, ou coluna de separação de etilbenzeno deve ter composição de topo maior que 99,6% de etilbenzeno em massa. Esta coluna tem sessenta pratos e a

alimentação é feita no prato doze. Um esquema da coluna com seus mecanismos de controle está representado na Figura 3.2.

A composição da coluna não pode ser monitorada em tempo real pois a análise das correntes é feita em intervalos de tempo elevados. Então o controle da composição é feita pelo controle da temperatura do prato 55, onde é mantida constante através do aumento ou redução da vazão de vapor no refeedor. Além do controle de temperatura a coluna também é equipada com controle de inventário do vaso de refluxo e do fundo da coluna. O controle do nível do vaso de refluxo é feito pelo controle da vazão de retirada de condensado do topo da coluna. O controle do nível do fundo da coluna é feito através do controle da vazão de fundo da coluna.

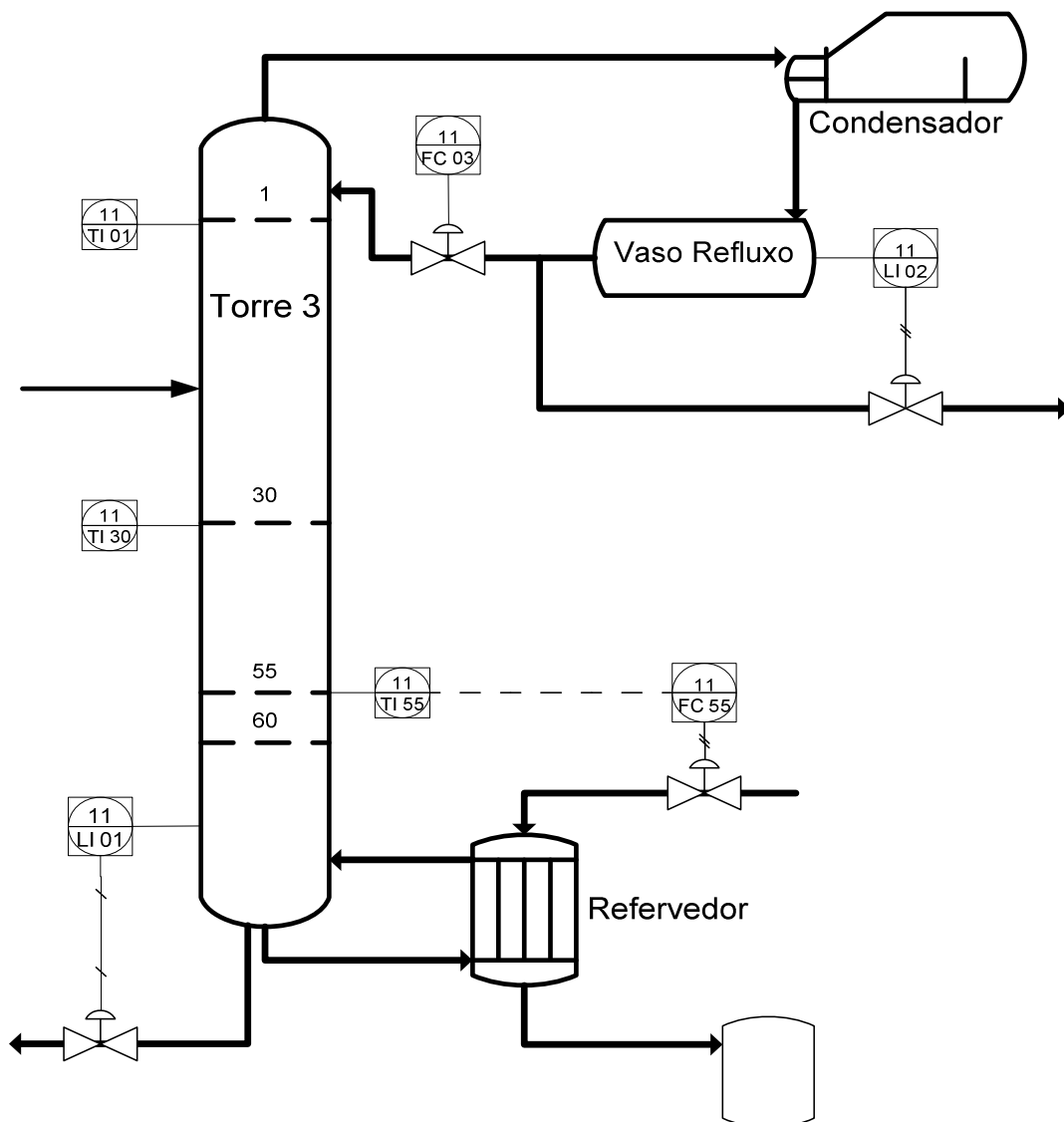


Figura 3.2: Representação da coluna de separação de etilbenzeno.

Para a Análise de Componentes Principais e para a determinação dos Limites Críticos de Controle da carta T^2 de Hotelling foram coletados dados reais da coluna de separação de etilbenzeno operando dentro da especificação, ou seja, sem distúrbios e com suas composições de topo e fundo dentro das especificações. A coleta dos dados foi realizada com o auxílio do software *Aspen Process Explorer*, um software que permite monitorar em tempo real as variáveis de processo, assim como o histórico das mesmas. A coleta foi realizada entre os dias 05 a 15 de abril de 2011.

A aquisição de dados foi realizada entre intervalos de 1 minuto, de forma a garantir uma reprodução confiável do processo, já às análises laboratoriais, que foram coletadas para o mesmo período de tempo, apresentam intervalos variáveis de tempo entre cada análise, variando de 2 a 5 horas. O número de análises feitas neste período de tempo foi 63. As variáveis monitoradas para a obtenção das componentes principais, para o ajuste da carta de controle e para a posterior verificação do método foram selecionadas conforme a disponibilidade dos instrumentos existentes na coluna de destilação, e são:

- Composição de topo da coluna, Y_t ;
- Vazão da carga de alimentação, V_e ;
- Temperatura da carga de alimentação, T_e ;
- Temperatura do fundo da coluna, T_f ;
- Temperatura do 55º prato, T_{55} ;
- Temperatura do 30º prato, T_{30} ;
- Temperatura do topo da coluna, T_t ;
- Vazão de retirada do fundo, V_f ;
- Pressão do fundo da coluna, P_f .

Os dados coletados da coluna em operação estão representados nas Figuras 3.3, 3.4, 3.5, e 3.6 onde os dados foram normalizados e representados através do número de desvios padrão deslocados da sua média amostral. A partir destes dados foram realizadas a Análise de Componentes Principais e a aplicação da Carta de Controle Multivariada de Hotelling.

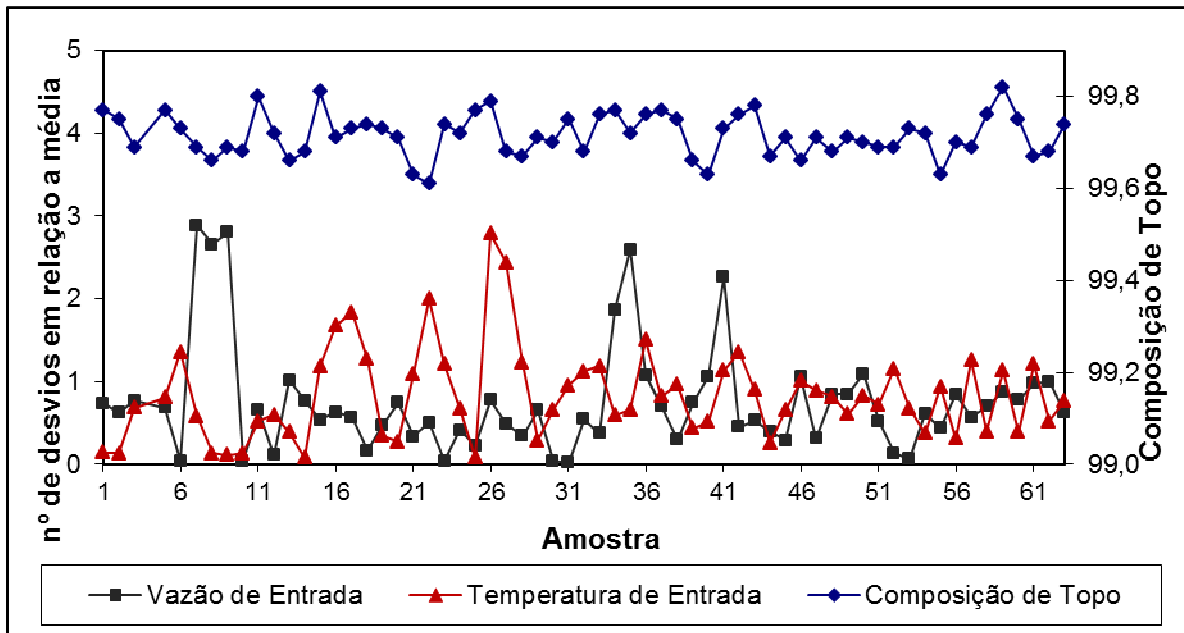


Figura 3.3: Números de desvios padrão em relação à média amostral para a vazão e temperatura de entrada coletadas da coluna de separação de etilbenzeno em operação.

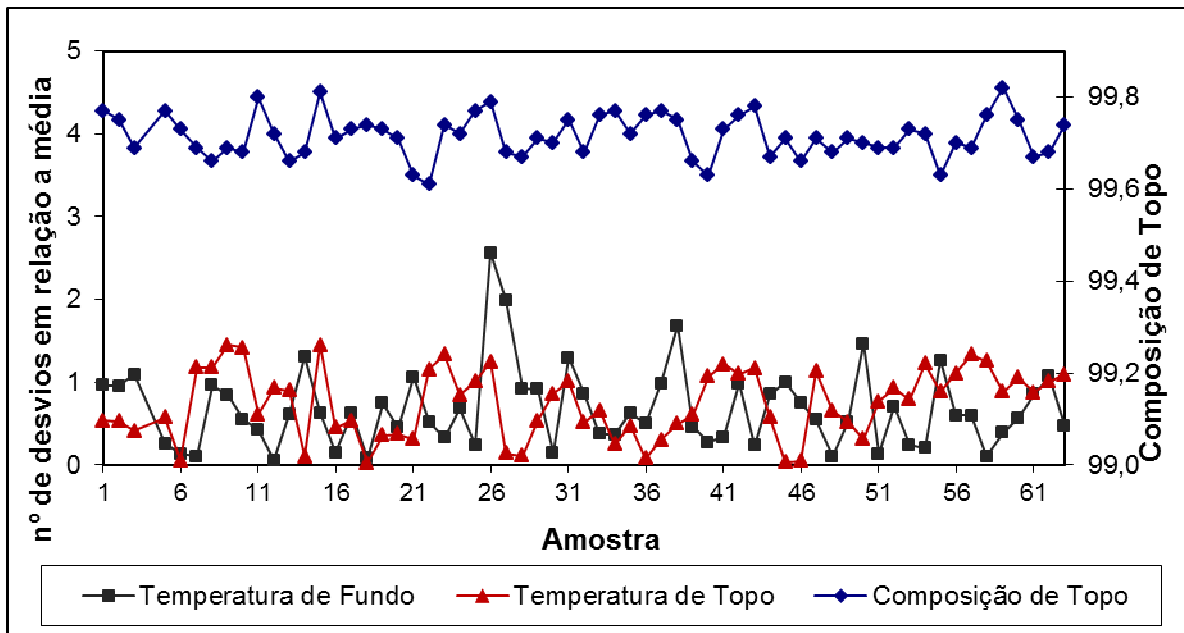


Figura 3.4: Números de desvios padrão em relação à média amostral para as temperaturas de fundo e topo coletadas da coluna de separação de etilbenzeno em operação.

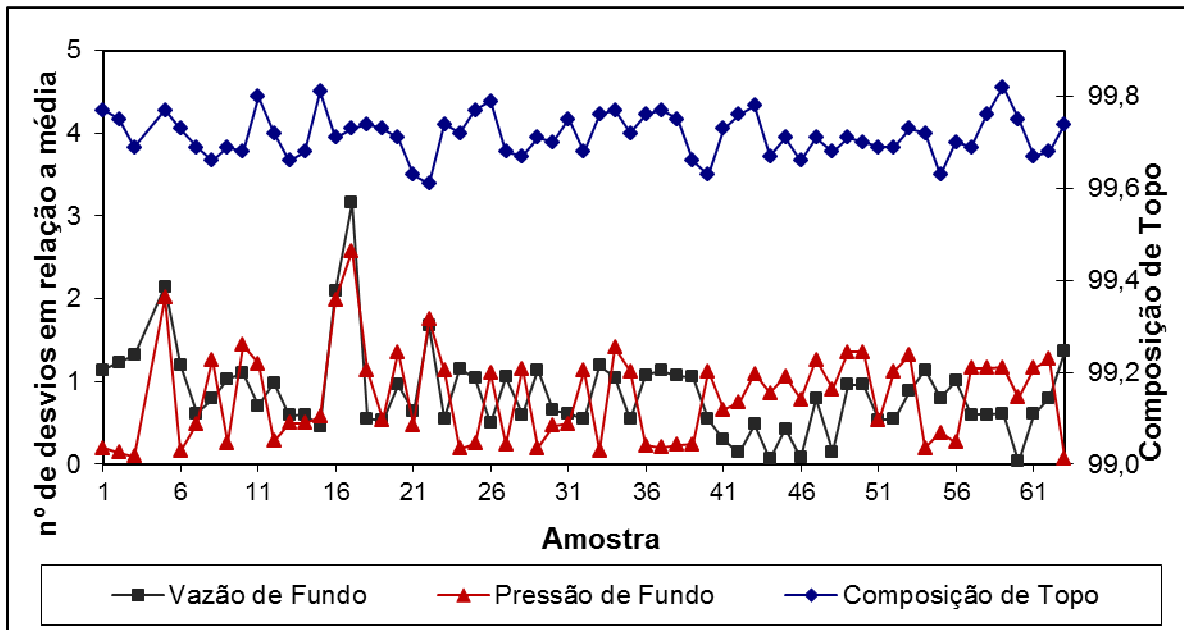


Figura 3.5: Números de desvios padrão em relação à média amostral para a vazão e pressão de fundo coletadas da coluna de separação de etilbenzeno em operação.

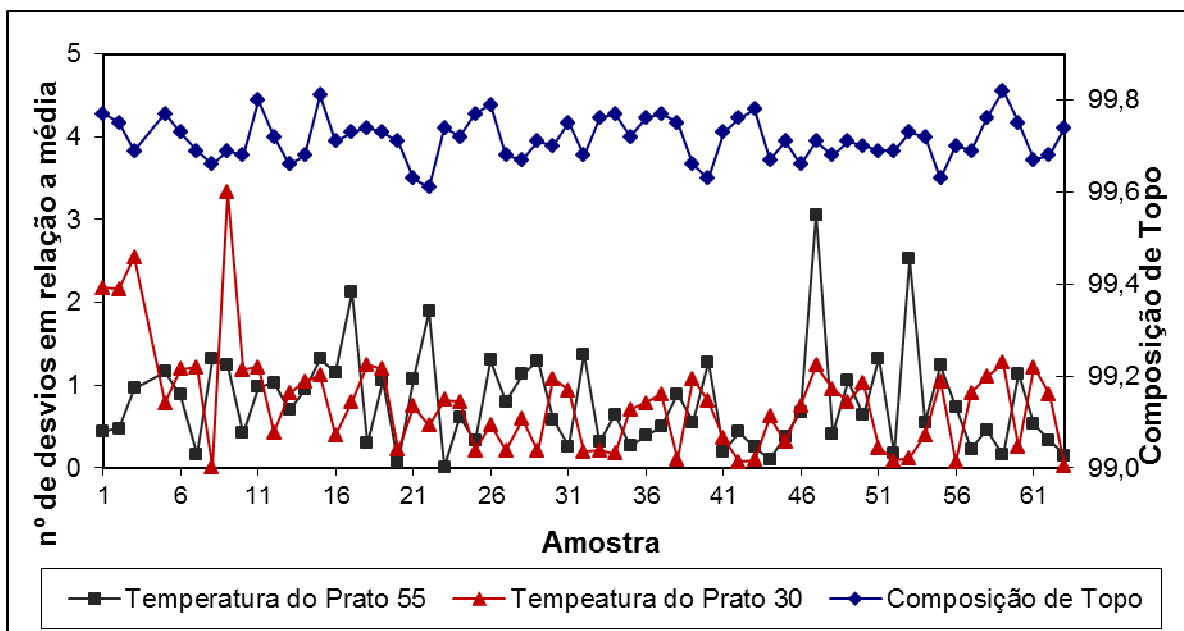


Figura 3.6: Números de desvios padrão em relação à média amostral para as temperaturas do prato 55 e prato 30 coletadas da coluna de separação de etilbenzeno em operação.

Estes dados foram obtidos da coluna operando sobre controle estatístico, e como se pode ver, não existem variações maiores do que três desvios padrão, o que mostra que o processo estava sobre controle estatístico.

3.2 SIMULAÇÃO DO PROCESSO

Para a construção de um modelo que representasse a coluna de separação de etilbenzeno, foi utilizado o simulador *Aspen Plus* na forma estática, para fazer a representação da mesma em estado estacionário, utilizando os dados reais da coluna como parâmetros de comparação do modelo. Como simplificação apenas as composições de Dietilbenzeno e de Etilbenzeno foram consideradas. Isso pode ser feito porque os outros dois produtos com quantidades mensuráveis (Trietilbenzeno e o Benzeno) são facilmente separáveis, ou seja, todo benzeno da alimentação é retirado pela corrente de topo e todo o trietilbenzeno é retirado pela corrente de fundo.

Após a implementação em estado estacionário foi utilizado o *Aspen Plus Dynamics* para inserir distúrbios no processo e coletar os dados da resposta da coluna. Estes dados foram utilizados para aplicar a carta de controle multivariada e detectar estes distúrbios inseridos. As variáveis monitoradas foram as mesmas monitoradas no processo real, para que a informação obtida pela planta simulada fosse a mesma obtida na planta real. Foram realizadas, ao todo, 20 horas de simulação, inserindo distúrbios na vazão da carga de alimentação (V_e) e na temperatura da carga de alimentação (T_e).

A modelagem da coluna de separação de etilbenzeno assim como os dados das simulações foram obtidos de Pavoni (2011). Os resultados das simulações estão representados na Figura 3.8, onde os valores foram normalizados para manterem uma mesma base de comparação e por fins de sigilo imposto pela empresa. Na Figura 3.8 estão representadas as variações inseridas nas variáveis vazão da carga de alimentação (V_e) e temperatura da carga de alimentação (T_e), juntamente com a porção de etilbenzeno retirado pelo topo da coluna.

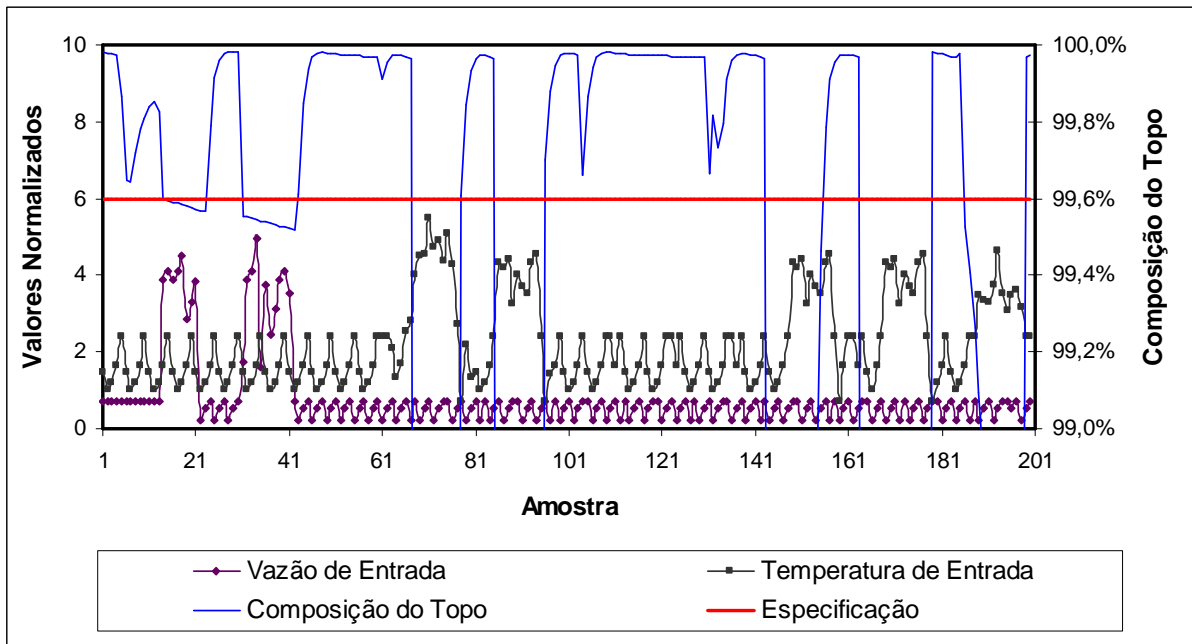


Figura 3.7: Dados obtidos das simulações da coluna de separação de etilbenzeno.

Foram feitos sete distúrbios na coluna para que a mesma transgredisse a especificação, dois na vazão da corrente de entrada e cinco na temperatura da corrente de entrada. O objetivo deste trabalho, como mencionado anteriormente, é detectar estes distúrbios inseridas no sistema sem que a composição de topo da torre seja informada, utilizando a carta de controle multivariada de Hotelling baseada em componentes principais (CCP).

3.3 TRATAMENTOS DOS DADOS

Os procedimentos descritos no capítulo 2 demandam muitos cálculos matriciais e trabalham com bancos de dados muito grandes. Para este tratamento dos dados foi utilizado o software MATLAB.

O MATLAB permite a realização de aplicações ao nível da análise numérica, de análise de dados, cálculo matricial, processamento de sinais e construção de gráficos, entre outras, abordando uma banda larga de problemas científicos e de engenharia.

No Apêndice A estão representadas as funções criadas em MatLab para o cálculo dos PCs e da estatística T^2 e os comandos utilizados para cada etapa do processo de obtenção.

4 RESULTADOS

4.1 OBTENÇÃO DAS COMPONENTES PRINCIPAIS DO SISTEMA

Para a análise de componentes principais (PCA), foram utilizados dados reais da planta de dez dias de operação da coluna de recuperação de Etilbenzeno, em condições normais onde a composição de Etilbenzeno foi monitorada em intervalos variados de tempo e manteve-se sempre dentro da faixa de especificação da unidade (>99,6%). Nestas condições foi gerado um banco de dados com 63 amostras onde foram realizadas as análises da composição do topo da coluna, onde esta composição foi utilizada para determinar se a coluna estava operando dentro das especificações do produto final. O objetivo deste trabalho é determinar um modelo estatístico que detecte distúrbios do processo, que façam o produto do topo da coluna sair da faixa de especificação, somente utilizando informações da planta em tempo real sem que a análise da corrente de topo da coluna seja necessária. As análises da composição de topo são utilizadas apenas para confirmar que os dados utilizados na modelagem estatística estão em operação estável e dentro de especificação do produto.

Por se tratar de variáveis de diferentes tipos de unidades e ordens de grandeza, tem-se uma discrepância na variância de cada uma das variáveis devido a este fato, como mostra a Tabela 4.1, desta forma as variáveis de maior desvio padrão tendem a dominar a variância do sistema omitindo as relevâncias das variáveis de menor desvio padrão. Este fato pode ser facilmente visualizado na Tabela 4.2 que mostra a variância dos PC determinados com dados não normalizados e dados normalizados. Como ilustrado na Tabela 4.2 utilizando os dados sem normalização apenas os dois primeiros PCs já seriam suficientes para representar o sistema, porém, utilizando os dados normalizados os cinco primeiros PCs seriam necessários para representar a variância do sistema.

Tabela 4.1: Desvio padrão das variáveis monitoradas do sistema.

	V_e (kg/h)	T_e (°C)	T_f (°C)	T_{55} (°C)	T_{30} (°C)	T_t (°C)	V_f (kg/h)	P_f (bar)
Desv. Pad.	370,20	0,18	0,39	4,00	1,22	0,22	199,43	0,01

Tabela 4.2: Variâncias capturadas por cada PC, calculadas para dados normalizados e dados não normalizados.

PC	Variância para dados não normalizados (λ_p)	Variância para dados normalizados (λ_p)
1	6,2394	2,8471
2	0,2817	2,0562
3	0,0001	1,2416
4	0,0000	0,9560
5	0,0000	0,5254
6	0,0000	0,1616
7	0,0000	0,1327
8	0,0000	0,0795

Para eliminar este problema foi realizada a normalização dos dados, através da equação (2.3), para que os mesmos sejam avaliados pelo número de desvios padrão deslocados em relação a sua média amostral.

O próximo passo para a obtenção dos PCs é a determinação da Matriz de Covariância (C). Porém, com o a normalização dos dados a Matriz de Covariância (C) torna-se a Matriz de Correlação (C^*) dos dados amostrais e pode ser obtida sem se fazer necessária a normalização dos dados amostrais. A Matriz de Correlação (C^*) obtida está representada na Tabela 4.3. Nela podemos ver as correlações entre as variáveis existentes no processo, como por exemplo, a pressão de fundo (P_f) e a temperatura de fundo (T_f) que é uma correlação óbvia.

Tabela 4. 3: Matriz de Correlação (C^*) para os dados amostrais.

	V_e	T_e	T_f	T_{55}	T_{30}	T_t	V_f	P_f
V_e	1	0,3946	-0,117	0,0242	-0,114	-0,041	0,8131	0,1128
T_e	0,3946	1	-0,216	0,0013	-0,008	-0,269	0,674	-0,33
T_f	-0,117	-0,216	1	0,6624	0,1353	0,6131	-0,152	0,5062
T_{55}	0,0242	0,0013	0,6624	1	0,1513	0,1029	-0,01	0,064
T_{30}	-0,114	-0,008	0,1353	0,1513	1	0,2552	0,0123	0,2771
T_t	-0,041	-0,269	0,6131	0,1029	0,2552	1	-0,151	0,8373
V_f	0,8131	0,674	-0,152	-0,01	0,0123	-0,151	1	-0,033
P_f	0,1128	-0,33	0,5062	0,064	0,2771	0,8373	-0,033	1

Para a determinação dos PCs do sistema é necessário calcular os autovalores e os autovetores associados à Matriz de Correlação (C^*) utilizando as equações (2.4) e (2.5) respectivamente. Para facilitar o cálculo foi utilizado o comando EIG(X) do MatLab, que calcula os autovetores e os autovalores associados a matriz X, que neste caso é a Matriz de Correlação (C^*). Os autovetores, ou os índices a_{pp} de cada PC, obtidos para os dados normalizados estão representados na Tabela 4.4, com estes dados pode-se representar os PCs da forma apresentada nas equações (2.1) e (2.2).

Tabela 4.4: Autovetores obtidos a partir da Matriz de Correlação (C^*).

PC	a_{p1}	a_{p2}	a_{p3}	a_{p4}	a_{p5}	a_{p6}	a_{p7}	a_{p8}
1	0,1645	0,6688	-0,28	0,0351	-0,079	0,4022	0,3033	0,4311
2	-0,57	0,2502	0,4416	0,0538	0,0175	0,0567	0,5519	-0,326
3	-0,401	-0,56	-0,164	-0,344	-0,068	0,3014	0,2531	0,4705
4	0,1132	-0,112	0,0949	0,239	0,9198	0,0443	0,1499	0,1952
5	-0,278	0,0699	-0,453	0,2414	-0,02	-0,746	0,225	0,2175
6	0,3832	-0,026	0,6167	-0,192	-0,182	-0,372	0,2437	0,4521
7	0,2704	0,0585	-0,287	-0,715	0,2198	-0,163	0,3431	-0,37
8	0,4202	-0,394	-0,152	0,4625	-0,247	0,1456	0,5377	-0,244

Como dito no Capítulo 2, um conjunto com p variáveis terá p PC, porém grande parte da variância do sistema pode ser representada com menos que a totalidade dos mesmos. Para se determinar a quantidade de PCs necessárias para se obter uma fração K da variância do sistema, foi utilizada a equação (2.7), com os autovalores determinados pelo comando EIG(X) do MatLab e $K = 0,95$ (95% da variância). Os autovalores e a porção de variância representada por cada PC, assim como a variância acumulada pelos primeiros PCs estão representados na Tabela 4.5. Como se pode ver na Tabela 4.5 os primeiros cinco PCs representam 95,33% da variância do sistema, sendo o suficiente para representá-lo dentro do critério adotado e reduzindo o número de variáveis em três dimensões, reduzindo também em 35,7% a memória necessária para armazenar os dados do sistema.

Tabela 4.5: Autovalores e proporção da variância do sistema associadas aos PCs obtidos.

PC	Autovalores (λ_p)	Proporção da Variância (%)	Proporção da Variância Acumulada (%)
1	2,8471	35,59	35,59
2	2,0562	25,70	61,29
3	1,2416	15,52	76,81
4	0,956	11,95	88,76
5	0,5254	6,57	95,33
6	0,1616	2,02	97,35
7	0,1327	1,66	99,01
8	0,0795	0,99	100,00

Então escritos da forma apresentada nas equações (2.1) e (2.2) os PCs obtidos para uma representação maior que 95% da variância do sistema são:

$$Z_1 = 0,1645*V_e + 0,6688*T_e - 0,28*T_f + 0,0351*T_{55} - 0,079*T_{30} + 0,4022*T_t + 0,3033*V_f + 0,4311*P_f$$

$$Z_2 = -0,57*V_e + 0,2502*T_e + 0,4416*T_f + 0,0538*T_{55} + 0,0175*T_{30} + 0,0567*T_t + 0,5519*V_f - 0,326*P_f$$

$$Z_3 = -0,401*V_e - 0,56*T_e - 0,164*T_f - 0,344*T_{55} - 0,068*T_{30} + 0,3014*T_t + 0,2531*V_f + 0,4705*P_f$$

$$Z_4 = 0,1132*V_e - 0,112*T_e + 0,0949*T_f + 0,239*T_{55} + 0,9198*T_{30} + 0,0443*T_t + 0,1499*V_f + 0,1952*P_f$$

$$Z_5 = -0,278*V_e + 0,0699*T_e - 0,453*T_f + 0,2414*T_{55} - 0,02*T_{30} - 0,746*T_t + 0,225*V_f + 0,2175*P_f$$

Em seguida as cinco PCs obtidas são aplicadas aos dados reais para testar a sua capacidade de representar o sistema. Para isso é feita uma comparação entre os valores das variáveis reais normalizadas (X_n) originais e as variáveis normalizadas obtidas pelo sistema formado pelas cinco PCs selecionadas (X_c). Essa comparação está exemplificada na Figura 4.1 para algumas das variáveis.

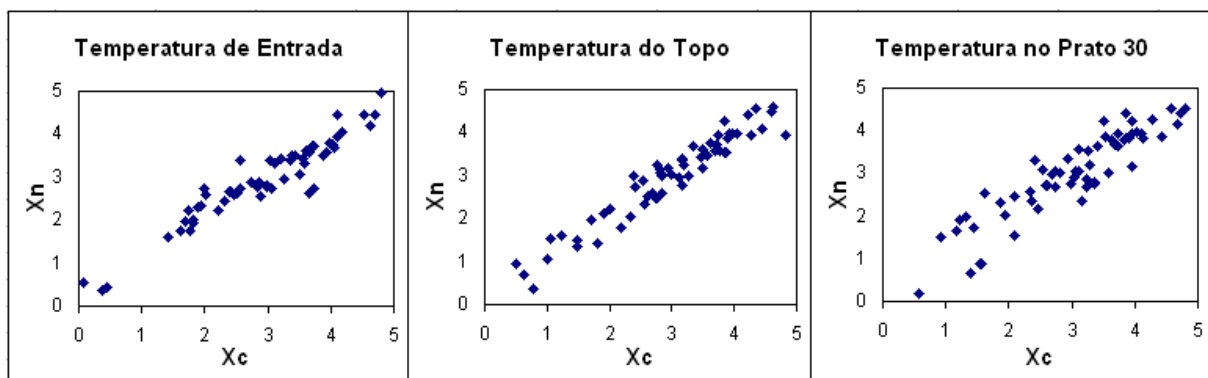


Figura 4.1: Comparação entre as variáveis reais normalizadas e as variáveis normalizadas obtidas através das componentes principais.

4.2 DETECÇÃO DESCONTROLES DO PROCESSO ATRAVÉS DA CARTA DE CONTROLE MULTIVARIADA DE HOTELLING – T²

Como dito anteriormente neste trabalho foi aplicada a carta de controle multivariada T² de Hotelling baseada em PCA (CCP) para monitorar a estabilidade da coluna de separação de etilbenzeno pertencente à Innova-PETROBRAS. Esta técnica utiliza componentes principais (PCs) para a redução do número de variáveis monitoradas pela carta de controle multivariada, facilitando os cálculos matriciais envolvidos na obtenção da estatística T² e eliminando as correlações entre as variáveis.

Neste trabalho os dados foram adquiridos com tamanho de amostra igual a um, ou seja, os dados são analisados em tempo real, sem que médias sejam calculadas para uma amostragem das variáveis. Este fato leva a dizer que $n = 1$ e implica na utilização das equações (2.15) e (2.16) para o cálculo dos limites de controle para primeira e segunda fase, respectivamente.

O limite para a primeira fase é então determinado utilizando a equação (2.15), onde a distribuição de probabilidade F é determinada para uma probabilidade de alarmes falsos $\alpha = 10\%$, com os graus de liberdade $v_1 = p = 5$, pois cinco variáveis (componentes principais) são utilizadas nesta carta e $v_2 = (c - p - 1) = 57$, pois tem-se 63 amostra de dados. A porção da distribuição de probabilidade F é determinada a partir de tabelas (Dean e Voss, 1999), onde α , v_1 e v_2 são necessários para a obtenção dos seus valores. Para a primeira fase o valor de F encontrado foi 1,96 e o LCS da 1ª fase foi 8,94.

Após a obtenção do limite de controle da primeira fase, foi calculada a estatística T² para os dados reais da coluna em operação estável através da equação (2.17) que calcula a estatística especialmente para PCs. Em seguida foi construída a carta de controle para estes valores, conforme mostrado na Figura 4.2. Os picos que ultrapassaram o limite de controle não são pontos onde ocorreram distúrbios, mas sim pontos onde a carta identificou uma possível distúrbio e que nesse caso é um alarme falso, pois os dados utilizados nesta carta estão todos dentro da especificação da planta. Para a redução dos alarmes falsos um novo cálculo dos limites de controle deve ser realizado levando em conta o número de alarmes falsos (h) encontrados na primeira fase do cálculo dos limites de controle.

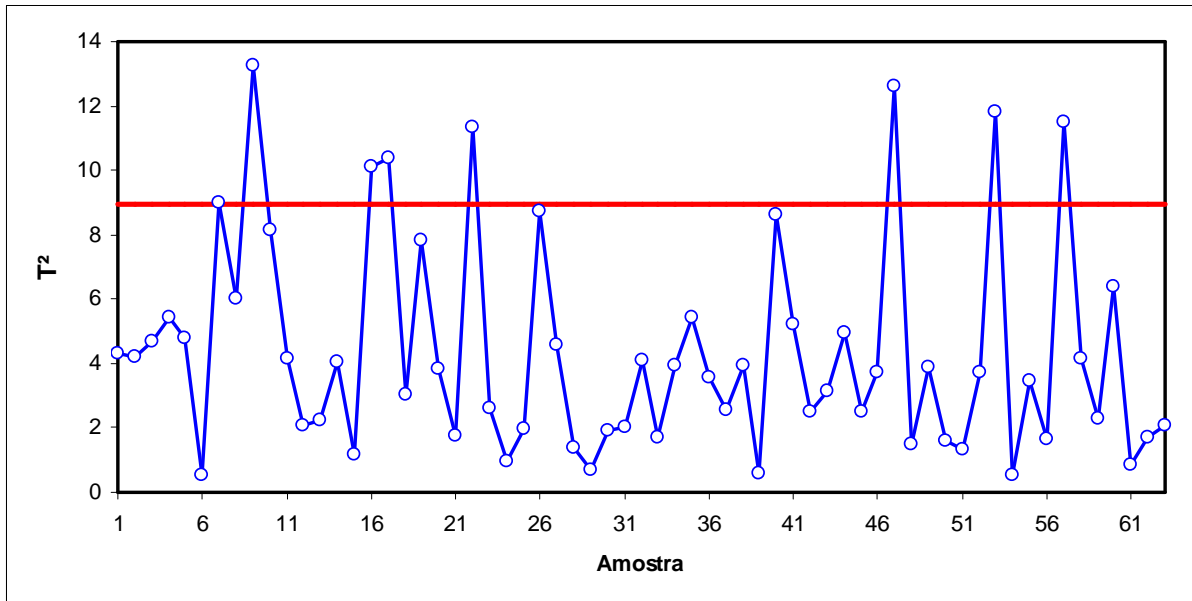


Figura 4.2: Carta de controle T^2 de Hotelling para primeira fase dos cálculos do limite de controle superior (LCS).

Através desta carta os dados fora do limite superior da primeira fase ($h = 8$) são excluídos para que um refino do cálculo do limite de controle superior da segunda fase seja feito. Desta forma é possível calcular o limite da segunda fase através da equação (2.16) com $m = (c - h) = 55$, e F com $\alpha = 10\%$, $v_1 = p = 5$ e $v_2 = (m - p) = 50$. O valor de F encontrado para a segunda fase foi 1,98 e o LCS da 2ª fase foi 10,96.

O LCS da 2ª fase será utilizado como limite para os dados futuros monitorados da coluna de separação de etilbenzeno operando. Para a verificação da eficácia da CCP desenvolvida, foram utilizados dados gerados em 20 horas de simulação em *Aspen Plus Dynamics* da coluna onde foram inseridos ruídos e distúrbios na vazão de entrada (V_e), na temperatura de entrada (T_e) e na pressão de topo (P_t) da coluna. A carta de controle obtida para os dados simulados esta representada na Figura 4.3, utilizando todas as variáveis monitoradas, descritas no capítulo 3, e aplicadas nos cinco PC selecionados anteriormente neste capítulo.

Para a análise dos pontos de distúrbio do sistema foi gerado uma lógica onde, é dado valor 1 quando a coluna esta fora de especificação, ou seja, composição de topo menor que 99,6% em etilbenzeno, e valor 0 quando os produtos de topo estão dentro da especificação. Essa lógica foi inserida para comparar a detecção do distúrbio pela carta de controle e a existência real deste distúrbio, verificando assim a eficácia na carta e os possíveis alarmes falsos indicados pela carta.

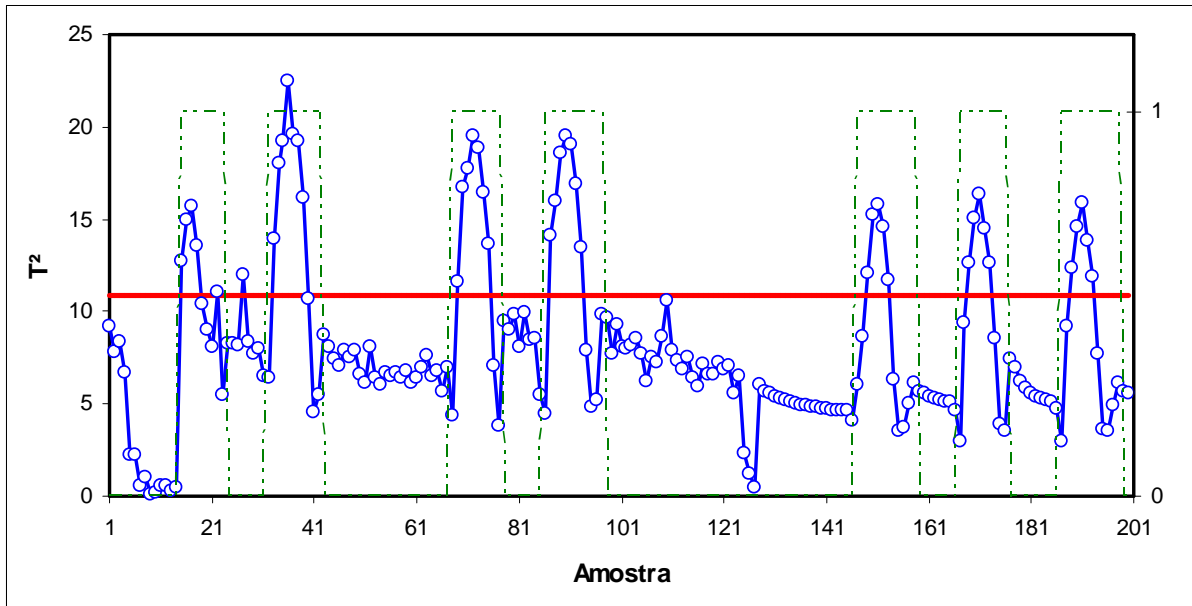


Figura 4.3: Carta de controle multivariada T^2 de Hotelling baseada em PCA para os dados simulados da coluna de separação de etilbenzeno.

Como se pode verificar nessas 20 horas de simulação, onde foram inseridos ruídos e distúrbios nas variáveis já mencionadas, a carta de controle multivariada baseada em componentes principais foi muito eficaz sendo capaz de detectar 100% dos distúrbios inseridas no sistema e gerando apenas 2 alarmes falsos. Para a redução dos alarmes falsos gerados pela carta, deve-se trabalhar com a probabilidade α de alarmes falsos para o cálculo do limite de controle. Porém cuidados devem ser tomados neste procedimento, pois uma probabilidade α de alarmes falsos muito elevada pode levar a não detecção de alguns pontos de distúrbio que se distanciam pouco do limite de controle.

Além disso, um atraso na detecção dos distúrbios pode ser verificado em alguns casos. Este atraso pode ser explicado pelo fato que tanto os limites de controle estatístico quanto as componentes principais foram calculados utilizando uma amostragem com intervalos muito maiores que o intervalo da amostragem dos dados da simulação gerando assim este atraso na detecção do distúrbio. Para corrigir este problema um banco de dados maior da coluna operando dentro das especificações e estável seria necessário para o cálculo dos limites e das componentes principais.

5 CONCLUSÃO

Neste trabalho foi analisada a carta de controle multivariada T^2 de Hotelling baseada na análise de componentes principais para a detecção de distúrbios na operação da coluna de separação de etilbenzeno pertencente a Innova-PETROBRAS. A carta de controle multivariada se mostrou muito eficaz, detectando 100% dos distúrbios inseridas nas simulações e gerando apenas 2 alarmes falsos.

As vantagens da utilização desta técnica são:

- A facilidade de implementação, não sendo necessárias equações fenomenológicas do processo nem mesmo um conhecimento aprofundado do seu funcionamento. São necessário somente bancos de dados para a obtenção das componentes principais e do limite de controle estatístico da carta de controle.
- Pode ser implantada como um dispositivo de alarme online do processo, sendo alimentada com dados atualizados e gerando as estatísticas instantaneamente, possibilitando a visualização de tendências de distúrbios do sistema e a atuação do operador antes que o distúrbio ocorra.
- A utilização da análise de componentes principais reduz significativamente o número de variáveis envolvidas no cálculo das estatísticas e reduzindo a memória de cálculo necessária.

A principal desvantagem encontrada foi a dificuldade de descobrir qual variável, ou quais variáveis, está fora do controle estatístico, gerando instabilidade do processo e tirando a coluna da especificação.

Em trabalhos futuros esta carta será refinada, com mais dados da operação da coluna, visando diminuir o atraso na detecção dos distúrbios, e também será realizada a verificação da carta para dados reais da coluna, ao invés de dados simulados, para a posterior implementação da carta para o monitoramento online da coluna.

REFERÊNCIAS

- DEAN, A. & VOSS, D. *Design and Analysis of experiments*. New York, Springer, 1999.
- DUNCAN, A.J. *Quality Control And Industrial Statistics*. 5ª Edição. Richard D. Irwin, Inc, 1986.
- HAIR, JR., J.F., ANDERSON, R.E., TATHAM, R.L. & BLACK, W.C. *Multivariate Data Analysis with Readings*. 4ª Edição. Prentice Hall, 1995.
- HOTELLING, H. *Multivariate Quality Control. Techniques of Statistical Analysis*. New York: McGraw-Hill, 1947.
- JOLLIFFE, IAN. *Principal Component Analysis*. 2ª Edição. Springer, 2002.
- KOURTI, T. & MCGREGOR, J. F. *Multivariate SPC Methods for Process and Product Monitoring*. J. Quality Technology, 1996.
- KONRATH, A. C. *Decomposição da estatística do Gráfico de Controle Multivariado T^2 de Hotelling por Meio de um Algoritmo Computacional*. Dissertação de Mestrado, 2002.
- MINGOTI, SUELI APARECIDA. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG, 2005.
- MONTGOMERY, D.C. *Introduction to Statistical Quality Control*. 3ª Edição. New York: John Wiley, 1996.
- PAVONI, P. A. *Proposição de um modelo de inferência baseado em Just In Time para coluna de Etilbenzeno*. Trabalho de Conclusão de Curso, 2011.
- ROSA, A. F. P. *Cartas de Controle Multivariadas Baseadas em Componentes Principais*. UFRGS – Porto Alegre – RS, 1999.
- SIQUEIRA, L. G. P. *Controle Estatístico de Processos*. Equipe Grifo. São Paulo: Pioneira, Série Qualidade Brasil. Livro nº 9, 1997.
- WARNE, K., G PRASAD, N. H. SIDDIQUE & L. P. MAGUIRE. *Development of a hybrid pca-anfis measurement system for monitoring product quality in the coating industry*. IEEE International Conference on Systems, Man and Cybernetics, 2004

APENDICE A

```
1 function [PC, PCn, VA, VAn, R, Rn] = ccpca(X);
2
3 - C = Cov(X);           % calcula a matriz de covariância C.
4
5 - [PC,VA] = eig(C);     % calcula os autovetores PC e os autovalores VA
6                       % associados a matriz de covariância C.
7
8 - VA = diag(VA);       % extração dos autovalores VA.
9
10 - R=100*VA/sum(VA);    % calculo da porcentagem da variancia
11                       % representada por cada PC obtido.
12
13
14 %Quando a normaização dos dados se faz necessaria, utiliza-se a matriz de correlação Cn.
15
16 - Cn = Corrcoef(X);    % calcula a matriz de correlação Cn.
17
18 - [PCn,VAn] = eig(Cr); % calculo dos autovetores PC e os autovalores VA
19                       % associados a matriz de correlação Cn.
20
21 - VAn = diag(VAn);    % extração dos autovalores VA.
22
23 - Rn=100*VAn/sum(VAn); % calculo da porcentagem da variancia
24                       % representada por cada PC obtido.
25
```

Figura A.1: Comando em MatLab para a obtenção dos PCs.

```

1 function [T2, Z, Xn, Xm, dx] = t2pca(X,PCs,VAs);
2
3     % PCs = PCs selecionados para representar o sistema
4     % VAs = Variância dos PCs selecionados
5 % caso os PCs forem obtidos através de dados normalizados, a aplicação dos
6 % PCs também devem ser feitas sobre dados normalizados, segue o
7 % procedimento de normalização dos dados.
8
9 - [h,j] = size(X);           % determinando as dimensões de X.
10
11 - Xm = mean(X);            % cálculo da média de X.
12
13 - dx = sqrt(VAR(X));      % cálculo do desvio padrão de X.
14
15 - Xn = (X - Xm(ones(h,1),:))./dx(ones(h,1),:); % normalização dos dados X.
16
17
18 % procedimento para o cálculo da estatística T².
19
20 - Z = Xn*PCs';           % calcula o escore Z do conjunto de variáveis normalizadas Xn
21                          % aplicadas os PCs selecionados anteriormente.
22
23 - VA = 1./VAs;           % calcula o inverso da variância para cada PC.
24
25 - T2 = Z.*Z*VA;          % calcula a estatística T² para dados normalizados da matriz X.
26
27

```

Figura A.2: Comandos em MatLab para o cálculo da estatística T².