



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
DEPARTAMENTO DE ENGENHARIA QUÍMICA

# **Detecção de Falhas em Processos Químicos Contínuos via Controle Estatístico de Processo Multivariado**

Disciplina: Laboratório de Operações e Processos

Alunos: Ana Brandão Belisário  
Bárbara de Sousa Pinto  
Gabriel Carvalho Zulato  
Rafael Francisco Ferraz Bicalho

Orientador: Professor Gustavo Matheus de Almeida

Co-orientador: Eng. Rodrigo César de Miranda / Optimus Engenharia

Belo Horizonte, 12 de Julho de 2017

# **Detecção de Falhas em Processos Químicos Contínuos via Controle Estatístico de Processo Multivariado**

Ana Brandão Belisário, Bárbara de Sousa Pinto, Gabriel Carvalho Zulato e  
Rafael Francisco Ferraz Bicalho

Relatório apresentado à disciplina Laboratório de Operações e Processos, do Programa de Graduação em Engenharia Química da Universidade Federal de Minas Gerais, como requisito parcial de obtenção do título de “Engenheiro Químico”.

Orientador: Professor Gustavo Matheus de Almeida

Co-orientador: Eng. Rodrigo César de Miranda / Optimus Engenharia

Belo Horizonte

2017

## **AGRADECIMENTOS**

Agradecemos ao nosso professor orientador Gustavo Matheus de Almeida por nos guiar durante os meses de trabalho e pelo apoio que nos foi dado na nossa formação como engenheiros.

Ao nosso co-orientador Rodrigo Miranda e à equipe da Optimus Engenharia, especialmente à Joana Bretz e ao Thiago Formiga, pelo interesse em contribuir com o nosso aprendizado e por toda a colaboração para o enriquecimento do nosso trabalho.

Ao engenheiro Simão Pedro de Oliveira, especialista da RHI Refratários do Brasil, e à professora Sílvia Maria Oliveira, do Departamento de Engenharia Sanitária e Ambiental da UFMG, pelo direcionamento e pela preocupação com o nosso aprendizado. À professora Miriam Cristina Santos Amaral e à doutoranda Aline Ribeiro Alkmim, do Departamento de Engenharia Sanitária e Ambiental da UFMG, que nos receberam atentamente, nos entregando a oportunidade de utilizar dados de sua pesquisa, e oferecendo suporte e informações essenciais para o desenvolvimento do nosso trabalho.

Ao engenheiro Leandro Couto Soares, que nos recebeu na fábrica da Brennand Cimentos em Sete Lagoas e nos apresentou todo o processo produtivo de cimento.

Por fim, agradecemos aos nossos pais e amigos pelo apoio e incentivo durante esse semestre.

## **RESUMO**

A quantidade de dados produzida na indústria química com os controladores de processos é enorme. Porém, os processos de tomada de decisão utilizando essas informações ainda são pouco explorados. Tendo em vista a crescente necessidade do uso de técnicas para a transformação das quantidades massivas de dados em informação útil, o presente trabalho apresenta a elaboração de um algoritmo para a técnica estatística multivariada denominada análise por componentes principais (PCA) e suas aplicações. Primeiro, o script foi validado a partir de seu uso no Benchmark Tennessee. Na sequência, dois processos químicos reais e contínuos foram utilizados como estudos de caso com o objetivo de detecção de falhas: um biorreator com membrana e um moinho vertical de uma fábrica de cimento. Os resultados mostram que o PCA pode ser utilizado com o propósito de monitoramento de processos químicos contínuos multivariados e com uma enorme quantidade de dados à disposição.

**PALAVRAS CHAVE:** Monitoramento de processos, Análise por componentes principais, PCA, Controle estatístico de processos, Detecção de falhas, Processos químicos contínuos.

## **ABSTRACT**

The data obtained in chemical processes in the industry is huge. In spite of that, the decision making process based on it is not well explored. Looking forward to the current necessity of data analysis, this article was written on the making of an algorithm for the statistical multivariate technique named principal component analysis (PCA) and its applications. First, the script was validated by using it on the Tennessee Eastman Process Benchmark. Then, two continuous chemical processes were analyzed for fault detection: the filterability of a membrane bioreactor and a vertical mill in a cement industry. The results show that the PCA technique can be used with the purpose of monitoring continuous multivariate chemical processes that have enormous quantities of data available.

**KEYWORDS:** Process monitoring, Principal component analysis, PCA, Statistical process control, Fault detection, Continuous chemical processes.

## Índice de Figuras

Figura 1 – Representação gráfica de processos sob controle e fora de controle. (Neto, 2003) .....	8
Figura 2 – Região de controle utilizando limites de controle independentes para x1 e x2 (Montgomery, 2004).....	9
Figura 3 – Gráfico da média para a característica Têmpera.....	10
Figura 4 – Gráfico da variabilidade para a característica Têmpera.....	10
Figura 5 – Gráfico da média para a característica Revenimento .....	11
Figura 6 – Gráfico da variabilidade para a característica Revenimento.....	11
Figura 7 – $T^2$ de Hotelling para tamanho de amostra variável para as características Têmpera e Revenimento.....	11
Figura 8 – Exemplos de cartas de controle multivariadas para os parâmetros $T^2$ e Q em Operação Normal (Silva R. H., 2015) .....	18
Figura 9 – Exemplos de Cartas de Controle Multivariadas para os parâmetros $T^2$ e Q em Operação com Falha (Silva R. H., 2015) .....	19
Figura 10 – Fluxograma esquemático do <i>Benchmark Tennessee</i> (Downs & Vogel, 1993)..	21
Figura 11 – Vetor de autovalores em ordem decrescente ( $\lambda_{(52 \times 1)}$ ).....	23
Figura 12 – Gráfico da variância explicada pelo número de componentes principais .....	24
Figura 13 – Cálculo dos limites de controle.....	25
Figura 14 – Cartas de controle para a falha 1 do processo <i>Tennessee Eastman</i> .....	25
Figura 15 – Representação esquemática do experimento .....	30
Figura 16 – Número de variáveis monitoradas em cada dia de monitoramento .....	31
Figura 17 – Teor de DQO (mg/L) em cada dia de monitoramento.....	33
Figura 18 – Correlação DQO e Fenol.....	34
Figura 19 – Correlação DQO e Óleos e Graxas .....	34
Figura 20 – Análise visual das regiões de medições de DQO .....	35
Figura 21 – Apresentação dos dados inseridos em meio ao conjunto de dados medidos de DQO .....	35
Figura 22 – Teor de amônia (mg/L) em cada dia de monitoramento .....	36
Figura 23 – Apresentação do conjunto de dados inseridos em meio ao conjunto de dados medidos de Amônia .....	36
Figura 24 – Teor de fenol (mg/L) em cada dia de monitoramento .....	37
Figura 25 – Correlação Fenol e DQO.....	37
Figura 26 – Correlação Fenol e Óleos e Graxas .....	38
Figura 27 – Correlação linear entre as variáveis Fenol e DQO .....	38
Figura 28 – Apresentação do conjunto de dados de inseridos em meio ao conjunto de dados medidos de fenol.....	39
Figura 29 – Teor de Óleos e Graxas (mg/L) em cada dia de monitoramento .....	39
Figura 30 – Correlação Óleos e Graxas e DQO .....	40
Figura 31 – Correlação Óleos e Graxas e Fenol .....	40

Figura 32 – Correlação linear entre as variáveis Óleos e Graxas e DQO.....	41
Figura 33 – Apresentação do conjunto de dados inseridos em meio ao conjunto de dados medidos de Óleos e Graxas.....	41
Figura 34 – Disposição das componentes principais, duas a duas. (Superior esquerdo: PC1 e PC2; superior direito: PC1 e PC3; inferior: PC2 e PC3) .....	43
Figura 35 – Representação da estatística $T^2$ para a operação normal .....	43
Figura 36 – Representação da estatística Q para a operação normal.....	44
Figura 37 – Representação da estatística $T^2$ para a operação teste .....	44
Figura 38 – Representação da estatística Q para a operação teste .....	45
Figura 39 – Comparação entre $T^2$ e Filtrabilidade .....	45
Figura 40 – Comparação entre Q e Filtrabilidade .....	46
Figura 41 – Relação entre as componentes principais (PC1 e PC2) .....	47
Figura 42 – Representação da estatística $T^2$ para a operação normal (não incluindo filtrabilidade) .....	47
Figura 43 – Representação da estatística Q para a operação normal (não incluindo filtrabilidade) .....	48
Figura 44 – Comparação entre a estatística $T^2$ e a variável filtrabilidade.....	48
Figura 45 – Comparação entre a estatística Q e a variável filtrabilidade .....	49
Figura 46 – Comparação entre os modelos 1 e 2.....	50
Figura 47 – Esquema do processo de produção do cimento (Rouge, 2012) .....	53
Figura 48 – Representação esquemática de um moinho vertical (ZKG, 2012) .....	55
Figura 49 – Representação de um permeabilímetro de Blaine (Júnior, 2013) .....	56
Figura 50 – Variância explicada em função das componentes principais.....	59
Figura 51 – Cartas de controle para o processo em operação normal .....	59
Figura 52 – Cartas de controle e Blaine para a primeira parte de fevereiro.....	60
Figura 53 – Cartas de controle e Blaine para a segunda parte de fevereiro .....	61
Figura 54 – Cartas de controle e Blaine para a terceira parte de fevereiro .....	62
Figura 55 – Cartas de controle e Blaine para a quarta parte de fevereiro.....	63
Figura 56 – Cartas de controle e Blaine para a quinta parte de fevereiro .....	64
Figura 57 – Cartas de controle e Blaine para a sexta parte de fevereiro .....	65
Figura 58 – Cartas de controle para a falha 1 .....	79
Figura 59 – Cartas de controle para a falha 2 .....	79
Figura 60 – Cartas de controle para a falha 3 .....	79
Figura 61 – Cartas de controle para a falha 4 .....	79
Figura 62 – Cartas de controle para a falha 5 .....	79
Figura 63 – Cartas de controle para a falha 6 .....	79
Figura 64 – Cartas de controle para a falha 7 .....	80
Figura 65 – Cartas de controle para a falha 8 .....	80
Figura 66 – Cartas de controle para a falha 9 .....	80
Figura 67 – Cartas de controle para a falha 10 .....	80
Figura 68 – Cartas de controle para a falha 11 .....	80
Figura 69 – Cartas de controle para a falha 12 .....	80

Figura 70 – Cartas de controle para a falha 13 .....	81
Figura 71 – Cartas de controle para a falha 14 .....	81
Figura 72 – Cartas de controle para a falha 15 .....	81
Figura 73 – Cartas de controle para a falha 16 .....	81
Figura 74 – Cartas de controle para a falha 17 .....	81
Figura 75 – Cartas de controle para a falha 18 .....	81
Figura 76 – Cartas de controle para a falha 19 .....	82
Figura 77 – Cartas de controle para a falha 20 .....	82
Figura 78 – Cartas de controle para a falha 21 .....	82

## **Índice de Tabelas**

Tabela 1 – Descrição das falhas no <i>Benchmark Tennessee</i> .....	22
Tabela 2 – Resumo dos resultados da aplicação do PCA no processo Tennessee .....	26
Tabela 3 – Quadro resumo comparativo do MDR do presente trabalho e do artigo de Russell et al. (2000).....	27
Tabela 4 – Quantidade de medições para cada variável de entrada .....	31
Tabela 5 – Quantidade de medições por variável de saída.....	32
Tabela 6 Coeficiente de correlação de Pearson entre os parâmetros da alimentação .....	32
Tabela 7 – Medianas associadas a cada período de observação do DQO .....	34
Tabela 8 – Conjunto de dados do moinho utilizados para aplicação do PCA .....	58

## Sumário

1.	Introdução.....	1
2.	Objetivos .....	5
3.	Revisão Bibliográfica .....	6
3.1.	Controle Estatístico de Processos (CEP).....	6
3.1.1.	Cartas de Controle.....	6
3.1.2.	CEP Univariado .....	8
3.1.3.	CEP Multivariado .....	9
4.	Metodologia .....	14
4.1.	Conceitos.....	14
4.2.	Etapas .....	16
4.2.1.	Construção do Modelo PCA.....	16
4.2.2.	Determinação dos limites de controle .....	17
4.2.3.	Detecção de Falhas .....	18
5.	Estudo de Caso 1: Benchmark Tennessee .....	20
5.1.	Apresentação.....	20
5.2.	Metodologia .....	23
5.2.1.	Construção do modelo PCA.....	23
5.2.2.	Definição dos limites de controle.....	24
5.2.3.	Detecção de falhas .....	25
5.3.	Resultados e Discussão.....	26
6.	Estudo de Caso 2: Biorreator com Membrana .....	28
6.1.	Apresentação.....	28
6.2.	Pré-Tratamento dos Dados – Variáveis de Entrada .....	30
6.2.1.	Inserção de Dados Faltantes ( <i>Missing data imputation</i> ).....	32
6.3.	Resultados e discussão .....	42
6.3.1.	Modelo PCA 1: Filtrabilidade incluída como variável de entrada para o modelo .....	42
6.3.2.	Modelo PCA 2: Filtrabilidade não incluída como variável de entrada para o modelo.....	46
6.3.3.	Comparação entre os modelos PCA 1 e PCA 2.....	49
6.4.	Conclusão.....	50
6.5.	Considerações finais e sugestões para trabalhos futuros .....	51

7.	Estudo de Caso 3: Moinho Vertical de Rolos na Produção de Cimento .....	52
7.1.	Produção de Cimento .....	52
7.1.1.	Moinho Vertical .....	53
7.2.	Resultados e Discussão.....	57
7.2.1.	Tratamento dos dados .....	57
7.2.2.	Construção do modelo PCA e definição dos limites de controle .....	58
7.2.3.	Detecção de falhas .....	59
7.3.	Conclusão.....	66
7.4.	Considerações finais e sugestões para trabalhos futuros .....	66
8.	Considerações Finais .....	67
9.	Referências Bibliográficas.....	68
	Anexo A – Script da Análise por Componentes Principais no Scilab .....	73
	Anexo B – Script da Análise por Componentes Principais no MATLAB .....	77
	Anexo C – Resultados da aplicação do PCA ao Benchmark Tennessee .....	79

## **1. Introdução**

Mudanças aceleradas movimentam e perturbam esse século XXI, pós-surgimento, desenvolvimento e expansão da Internet e de tecnologias de coleta e armazenamento de dados. A capacidade de assimilação e implementação das novas ferramentas e técnicas, em geral, está incrivelmente defasada em relação ao aumento da velocidade de desenvolvimento de novas tecnologias. Ainda assim, muitas inovações próprias desse tempo já estão sendo implementadas e desenvolvidas por alguns setores industriais. Desse fato, surge um novo momento, uma nova abordagem: a Indústria 4.0.

A Indústria 4.0 é caracterizada pela aplicação das Tecnologias de Informação, Comunicação e Localização (TICLs) nos processos industriais já conhecidos. Isso implica na evolução dos sistemas de produção, passando pelos mais diversos níveis, incluindo compartilhamento de dados e informações em tempo real, envolvendo homens, máquinas e robôs. Neste modelo da Indústria 4.0, os meios de produção e seus agentes estão conectados digitalmente, as cadeias de abastecimento e os canais de distribuição estão interligados. As implicações mais imediatas disso são as possibilidades de previsão de falhas, de adaptação da produção conforme o cenário do momento e de integração do processo produtivo global. (Amaral, 2016)

A tendência de expansão de empresas ligadas intrinsecamente à tecnologia, movimenta uma mudança de cenário para organizações mais antigas, gerando uma necessidade urgente de adaptação ao uso de ferramentas digitais e à implementação de modelos de negócio e de processos mais inovadores. Não é simples mudar comportamentos e rotinas de pessoas e, pensando em unidades de produção em larga escala, o desafio é ainda maior. Entretanto, como a familiarização e a implementação de recursos digitais impactam diretamente na competitividade, as mudanças precisam acontecer e de forma acelerada (Afonso, 2016). A implementação de novos métodos de operação demanda uma quantidade de esforços, passando pela integração de tecnologias físicas e mecânicas aos meios digitais de coleta e processamento de dados. O resultado é uma quantidade de informações apropriadas para (re)engenharia e controle da produção, bem como para as etapas de pesquisa e desenvolvimento de produtos, de processos e de novos modelos de negócios (Gonçalves, et al., 2016).

Nesse meio de tantas conexões, sensores e integrações entre início e fim de processos, as medições para obtenção de dados sobre o processo desempenham um papel fundamental para análise, controle e otimização do mesmo. Sistemas de medições e de instrumentação industrial começaram a se difundir intensamente entre as décadas de 1950 e 1970, com o avanço dos computadores e dos circuitos integrados. Implicação disso é o fato de que a grande maioria das indústrias reúne grandes conjuntos de dados de medições de seus processos, e já possui estruturas de controle e sistemas controladores bastante desenvolvidos (Wallén, 2008). Entretanto, o histórico volume de dados gerado pelos sistemas de medições e de controle ainda é pouco explorado. Em outras palavras, a

capacidade de interpretação desses dados para o monitoramento e o controle mais eficaz do processo, bem como para a tomada de decisões imediatas, emergenciais ou futuras, ainda é subutilizada. Nessa direção já existem metodologias e ferramentas de análise e interpretação de grandes quantidades de dados; e o desenvolvimento delas continua em foco, sendo suportado pela capacidade dos computadores em processar bancos de dados cada vez maiores (Wang, 1999).

Um dos vieses importante da análise e interpretação de grandes quantidades de dados está relacionado à identificação de falhas em processos. O monitoramento de processos é importante para a melhoria da eficiência operacional, o que significa realizar as tarefas com o máximo de rendimento e o mínimo de desperdício. Isso passa pela busca da eficácia em cada operação, ou seja, realizar as coisas certas no momento adequado. A eficácia em um processo industrial é especialmente importante em casos de falhas, seja de equipamentos, sensores, controladores, quando a correção deve ser adequada, precisa e imediata. Assim torna-se necessário um sistema capaz de monitorar, detectar e diagnosticar falhas, em concomitância ao monitoramento do processo. Falhas podem ser caracterizadas como condições de perturbações ou comportamentos anormais (Das, Maiti, & Banerjee, 2012)

Definição de monitoramento, por R. Insermann e P. Balle (1997):

*Monitoring: A continuous real-time task of determining the conditions of a physical system, by recording information, recognising and indicating anomalies in the behaviour.*

O monitoramento de processos idealmente ocorre em tempo real, com o objetivo de reagir previamente diante de qualquer comportamento inesperado das variáveis que estão sendo coletadas e avaliadas (Dunn, 2016). Uma parte importante do monitoramento está na escolha das variáveis a serem monitoradas; elas precisam representar muito bem todos os pontos em que erros poderiam acontecer. É bastante comum o monitoramento de um grande número de variáveis. E, por isso, é importante o desenvolvimento de modelos capazes de avaliar bem e em alta velocidade grandes volumes de dados, oferecendo respostas em tempo real.

Nessa direção, a atividade de monitoramento de processos pode ser dividida em três categorias: baseado em modelos analíticos (fundamentado em princípios básicos da física e da química, que relacionam as variáveis por meio de equações), baseado em expertise (tem como fundamento o conhecimento sobre o comportamento do processo advindo da experiência dos operadores da planta) e baseado em dados (advém das informações carregadas pelos parâmetros do processo que são coletados e monitorados). O monitoramento baseado em dados se torna cada vez mais popular ao longo dos anos devido à complexidade dos processos industriais. Grandes quantidades de dados são coletadas e usadas para modelagem, monitoramento e controle (Ge, Song, & Gao, 2013).

A Análise por Componentes Principais (PCA) é um exemplo de técnica de monitoramento baseado em dados, e que pode ser considerada uma ferramenta básica para a análise de dados multivariada (L. Erikson, 2013), sendo uma das metodologias mais aplicadas para a

detecção de falhas. Um dos principais motivos para a expansão dessa técnica está na capacidade de redução dimensional da matriz de variáveis original para uma nova matriz que contém a representação dos dados originais pela maximização da variância explicada deles (Severson, Chaiwatanodom, & Braatz, 2016).

Para exemplificar Rato, Schmitt & Reis (2011) comparam métodos de monitoramento baseados em Análise por Componentes Principais (*Principal Component Analysis*, PCA). Já Detroja, Gudi & Patwardhan (2006) testam o PCA e o DPCA (PCA dinâmico). Yin (2012), por sua vez, compara o PCA com o método dos mínimos quadrados (PLS ou *Partial Least Squares*), análise por componentes independentes (ICA), análise de discriminante de Fisher (FDA), entre outros.

Apesar de cada grupo de autores retirar suas conclusões sobre a robustez e a eficiência dos métodos, algumas ideias convergem e se destacam nas análises sobre as tendências para campos promissores de pesquisa (Ge, Song, & Gao, 2013):

- Monitoramento de processos dinâmicos complexos: requer métodos que consigam acompanhar as mudanças do processo ao longo do tempo e seu comportamento complexo.
- Monitoramento de processos “*plantwide*” (larga escala): requer métodos que sejam eficientes para analisar a planta industrial como um todo e as causas raiz de falhas, ao invés de analisar as operações separadamente.
- Combinação de modelos: Métodos que, quando agrupados, ajam em sinergia e se tornem mais robustos e eficientes.
- Monitoramento de processos adaptativos: demanda métodos que consigam acompanhar os processos que mudam de condição operacional ao longo do tempo.

Em todos os casos, a detecção de falhas é uma etapa essencial do monitoramento de processos. Cabe ressaltar que a técnica de Análise por Componentes Principais (PCA) será detalhada no Capítulo 4 deste trabalho.

O acontecimento de uma falha precisa ser detectado para que as providências necessárias sejam tomadas. Segundo Insermann e Balle (1997), ‘detecção de falhas’ pode ser definido como a determinação das falhas presentes em um sistema e do tempo em que foram detectadas. A informação da existência de uma falha detectada leva à necessidade de se isolarem as variáveis de forma a determinar o tipo, a localização e o tempo em que ela ocorreu (*fault isolation*). Com isso, torna-se possível identificar a falha, ou seja, precisar a sua proporção e o seu comportamento temporal. Por fim, resta apresentar o diagnóstico da falha, que é o compilado de todas as informações obtidas pelas etapas anteriores: determinação do tipo, do tamanho, do tempo, das proporções e do comportamento ao longo do tempo. Com este diagnóstico, é possível agir de forma eficaz, propondo e executando as correções apropriadas e no tempo certo, a fim de que o processo retorne ao ponto de operação normal. Interessante é também a oportunidade do desenvolvimento de sistemas de proteção a falhas, o que envolve conhecer o processo a ponto de, diante de qualquer

comportamento estranho, conseguir agir preventivamente, contornando a situação e evitando que o comportamento estranho se desenvolva (Insermann & Balle, 1997).

O foco do presente trabalho está na tarefa de detecção de falhas tendo como base o modelo PCA.

## **2. Objetivos**

O objetivo geral do presente trabalho consistiu na construção de um sistema de detecção de falhas para processos químicos contínuos a partir da técnica multivariada denominada Análise por Componentes Principais (PCA, *Principal Component Analysis*).

Para tanto, quatro objetivos específicos foram propostos:

1. Implementação do código do modelo PCA em linguagem Scilab
2. Validação do modelo PCA utilizando dados de comportamento conhecido (Estudo de caso 1: Benchmark Tennessee)
3. Aplicação da técnica PCA para um experimento em planta piloto de tratamento de efluentes (Estudo de caso 2: Biorreator com membranas)
4. Aplicação da técnica PCA para uma indústria de cimentos (Estudo de caso 3: Moinho vertical de rolos).

### **3. Revisão Bibliográfica**

Descreve-se a seguir, de modo resumido, sobre controle estatístico de processos (CEP) e sua principal ferramenta: a carta de controle. Na sequência, é descrito o CEP univariado e a sua limitação para o monitoramento de processos multivariados. Desse modo, fecha-se essa seção com uma descrição sobre o CEP multivariado, onde se situa a técnica de Análise de Componentes Principais (PCA).

#### **3.1. Controle Estatístico de Processos (CEP)**

O crescimento econômico e o acirramento da competição estão desencadeando uma busca na melhoria da qualidade e produtividade dos processos produtivos e, consequentemente, aperfeiçoamentos nos produtos finais. Atualmente, a sobrevivência do setor industrial está substancialmente baseada na melhoria contínua de seus processos e produtos. Os esforços para alcançar melhorias da qualidade devem focar em aprimoramentos contínuos, atitudes que, promovidas continuamente, permitam reconhecer problemas, priorizar ações corretivas, implantá-las e dar sequência a postura proativa, agindo corretamente. (Silva L. , 1999)

O Controle Estatístico de Processo (CEP) é um poderoso conjunto de ferramentas de monitoramento on-line da qualidade. Com tais ferramentas, consegue-se uma descrição detalhada do comportamento do processo. Assim, é possível identificar a variabilidade do processo e possibilitar seu controle ao longo do tempo. Através da coleta contínua de dados e da análise e bloqueio de possíveis causas especiais, responsáveis pelas instabilidades do processo, assegura-se uma melhoria contínua da qualidade e da produtividade, conjuntamente (MOREIRA, 2004; WHEELER, 1992). O CEP abrange a coleta, a análise e a interpretação de dados com a finalidade de resolver um problema particular (Paranthaman, 1990). A ideia principal é melhorar os processos de produção com menos variabilidade proporcionando níveis melhores de qualidade nos resultados da produção. Assim, permite, também, uma avaliação adequada da qualidade. Agir no processo é, antes de tudo, evitar defeitos, independentemente de onde eles possam manifestar-se. Este é o princípio do CEP que usa a Estatística como instrumento básico para a organização, tratamento e a análise das informações do processo, operando preventivamente (Paladini, 1990).

##### **3.1.1. Cartas de Controle**

Na aplicação do CEP, diversas metodologias são usadas para entender, modelar e reduzir a variabilidade de um processo industrial durante todo seu período. A estratégia do CEP é controlar o processo, medindo as variáveis de interesse em pontos espaçados no tempo e registrando os resultados em cartas de controle. Essas cartas são ferramentas fundamentais, que têm como objetivo detectar desvios de parâmetros representativos do processo, reduzindo, assim, a quantidade de produtos fora de especificações e os custos de produção.

Os gráficos de controle são elementos visuais para o monitoramento de processos, pois conseguem detectar rapidamente alterações desproporcionais e causas especiais, sinalizando a necessidade de intervir no processo. De acordo com Montgomery (2004) há dois tipos de causas: especiais e comuns. Causas especiais são eventos que ocorrem ao acaso, em períodos irregulares, e que devem ser eliminados e reduzidos. As causas comuns atuam continuamente sobre o processo, sendo de natureza essencialmente aleatória e de difícil controle. As causa comuns são inerentes ao processo e são praticamente inevitáveis. Atuar sobre causas comuns normalmente requer investimentos na melhoria de equipamentos, troca de matérias primas ou treinamento de colaboradores. Uma vez mantidas em níveis razoáveis, as causas comuns não afetam de maneira nociva a qualidade dos itens manufaturados. A utilização das cartas de controle pressupõe que o processo seja estatisticamente estável, ou seja, com ausência de causas especiais de variação. As sucessivas amostragens devem representar um conjunto de valores independentes ou não correlacionados. Este pressuposto é dificilmente atendido e muitas vezes as cartas de controle são usadas com limites inadequados e com a frequente ocorrência de alarmes (pontos fora ou próximos aos limites da carta) sem que, de fato, tenha ocorrido uma causa especial (Montgomery, 2004).

Os gráficos de controle representam uma das técnicas estatísticas que servem de apoio ao controle da qualidade de um processo, fornecendo evidências de suas variações tanto de caráter aleatório quanto de caráter determinável. Eles permitem que se possa atuar no processo de forma preventiva, corrigindo possíveis desvios de qualidade, em tempo real, no momento em que eles estão ocorrendo. Dessa forma, impedem que possíveis inconformidades perdurem até a reprovação do lote final. É importante destacar que um gráfico de controle não identifica quais são as causas especiais de variação que estão atuando em um processo fora de controle estatístico, mas ele processa e dispõe informações que podem ser utilizadas na identificação destas causas (Werkema, 1995).

O Dr. Walter A. Shewhart, do Bell Telephone Laboratories, foi o primeiro a formalizar a distinção entre variação controlada e não controlada, que corresponde às causas comuns e causas especiais, respectivamente. Ele desenvolveu, então, a carta de controle, que é uma ferramenta simples, mas poderosa, capaz de separar esses dois tipos de causas. O modelo de carta estabelecido por Shewhart apresenta uma linha central, um limite superior de controle (LSC) e um limite inferior de controle (LIC). Os dados de entrada são medições realizadas de uma determinada característica de qualidade ou parâmetro de processo que influencie na qualidade do produto. Se os valores marcados no gráfico estiverem dentro dos limites de controle, com disposição aleatória dos pontos dentro dos limites e sem qualquer tendência particular, o processo é considerado sob controle. Entretanto, se os pontos incidirem fora dos limites de controle ou apresentarem uma disposição atípica, o processo é julgado fora de controle (Kume, 1993). A Figura 1 representa em um mesmo gráfico, o mesmo processo sob controle e fora de controle.

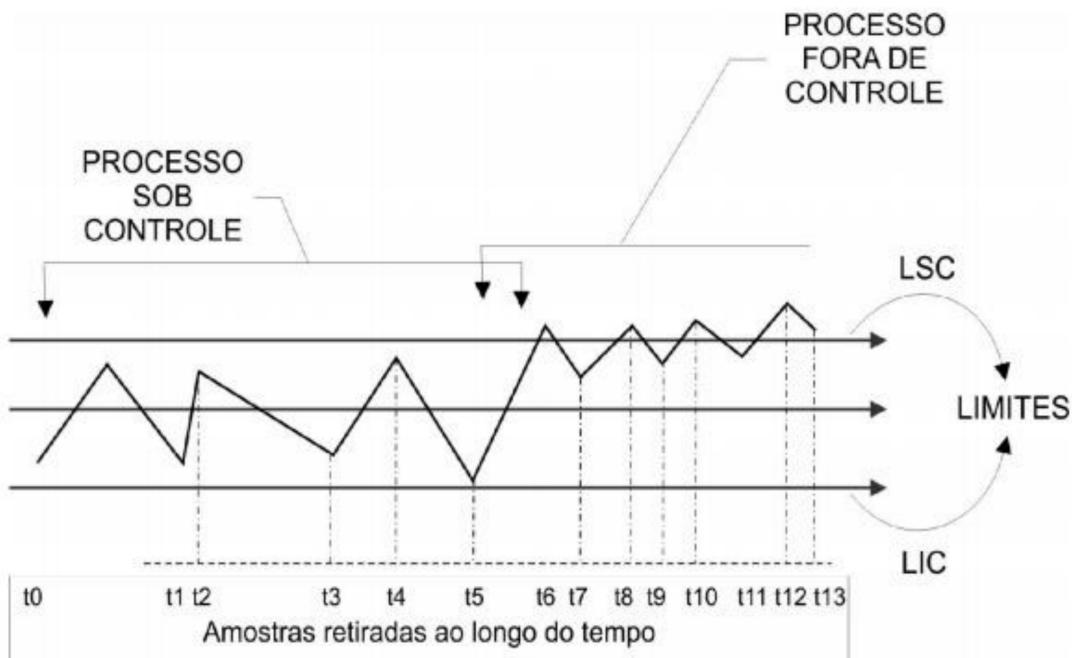


Figura 1 – Representação gráfica de processos sob controle e fora de controle. (Neto, 2003)

### 3.1.2. CEP Univariado

As características de qualidade são mensuráveis, sendo variáveis cujo monitoramento deve considerar tanto o valor médio da característica como sua variabilidade. Dessa forma, gráficos de controle são amplamente utilizados, baseando-se na distribuição contínua das medições, que podem, de certa forma, medir o grau de aceitabilidade ou a não aceitabilidade (Alves, 2003).

O controle da média do processo ou é usualmente feito através do gráfico de controle para médias ou através do gráfico de  $\bar{x}$ . A variabilidade do processo pode ser monitorada tanto através do gráfico de controle para o desvio-padrão, chamado gráfico S, como do gráfico de amplitude, chamado gráfico R. A diferença básica entre os gráficos S e R é que o primeiro se aplica a casos onde o número de amostras é moderadamente grande. Exemplificando melhor a função de cada gráfico, observa-se que através do gráfico de  $\bar{x}$ , monitora-se a variabilidade entre amostras e, através do gráfico de S, monitora-se a variabilidade dentro da amostra (Control, 1986). A média  $\mu$  e o desvio padrão  $\sigma$  são usados para construir a carta de controle. Os parâmetros para o gráfico  $\bar{x}$  são:

$$LSC = \mu + A \cdot \sigma$$

$$LC = \mu$$

$$LIC = \mu - A \cdot \sigma$$

E os parâmetros para o gráfico R serão:

$$LSC = D_2 \cdot \sigma$$

$$LC = d_2 \cdot \mu$$

$$LIC = D_1 \cdot \sigma$$

Em que  $A, D_1$  e  $D_2$  são constantes dependentes de  $n$  e de  $d_2$  a média da distribuição da amplitude relativa. (Montgomery, 2004)

No controle estatístico de processo clássico (ou univariado), cada variável ou característica de qualidade é plotada em gráficos de controle separados, não levando em consideração a relação entre as variáveis. Desse modo, assume-se que as variáveis do processo são independentes. Segundo Montgomery (2004), quando existem múltiplos parâmetros do processo que são correlacionados entre eles, os gráficos univariados não devem mais ser utilizados. Nesse caso, o desempenho destes é prejudicado ao se ter uma falta de controle do processo. A solução para estes casos é o uso de cartas de controle multivariadas.

### 3.1.3. CEP Multivariado

Quando existem múltiplos parâmetros do processo que precisam ser monitorados simultaneamente, devido ao fato de serem correlacionados entre si, o controle estatístico univariado não deve mais ser utilizado. A Figura 2 ilustra uma situação em que o monitoramento independente de duas variáveis (isto é, a partir de duas cartas de controle univariadas) de um processo não é capaz de detectar uma falha, que se torna bem evidente quando as variáveis são monitoradas simultaneamente (isto é, a partir de uma única carta de controle multivariada).

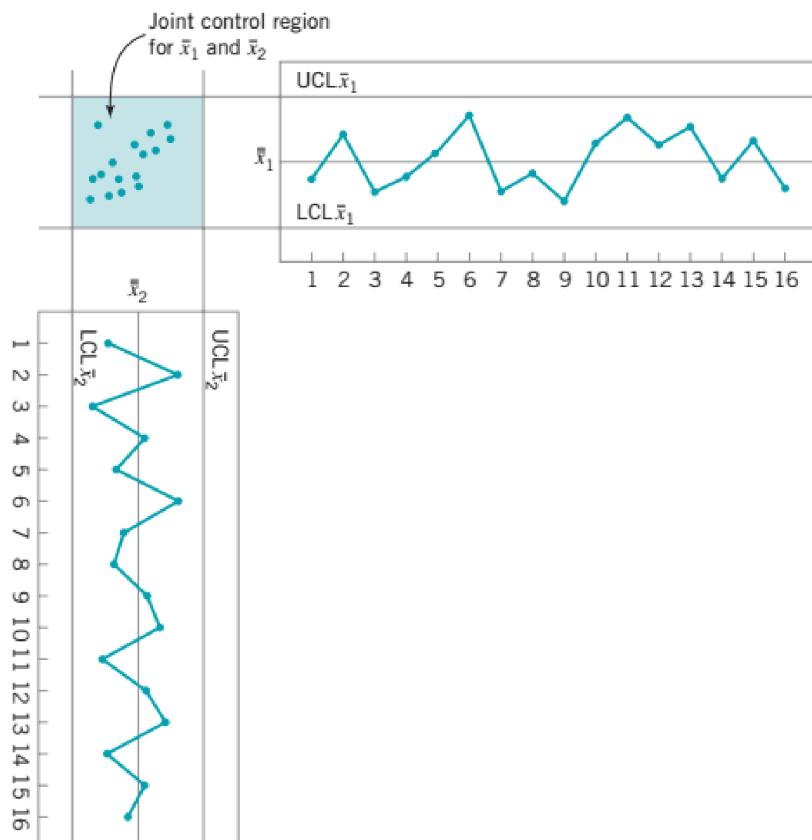


Figura 2 – Região de controle utilizando limites de controle independentes para  $x_1$  e  $x_2$  (Montgomery, 2004).

Para ilustrar a diferença entre o CEP univariado e multivariado pode-se avaliar o seguinte trabalho em que ambos foram aplicadas ao processo de produção de hastes metálicas em empresas da região de São João del Rei – MG. Atualmente, o monitoramento de características desse processo tem sido realizado com gráficos de controle de qualidade univariados. No entanto, esses processos analisam diversas variáveis, que usualmente estão correlacionadas, e a literatura mostra que analisar separadamente não é adequado.

Os gráficos de controle mais empregados para monitorar uma característica da qualidade numérica são os gráficos Xbarra (monitorando o valor médio) e o gráfico S (monitorando a variabilidade). O gráfico de controle multivariado  $T^2$  de Hotelling é um gráfico análogo ao gráfico Xbarra de Shewhart univariado e é utilizado no monitoramento simultâneo de p características da qualidade. Além disso, apresenta mais sensibilidade do que os gráficos univariados, permitindo ao operador detectar mais rapidamente os possíveis problemas existentes no processo e, com isso, corrigi-los com mais agilidade. Para a melhoria da qualidade das hastes metálicas, duas características foram analisadas: têmpera e revenimento. Os resultados são mostrados nas Figuras 3 a 7.

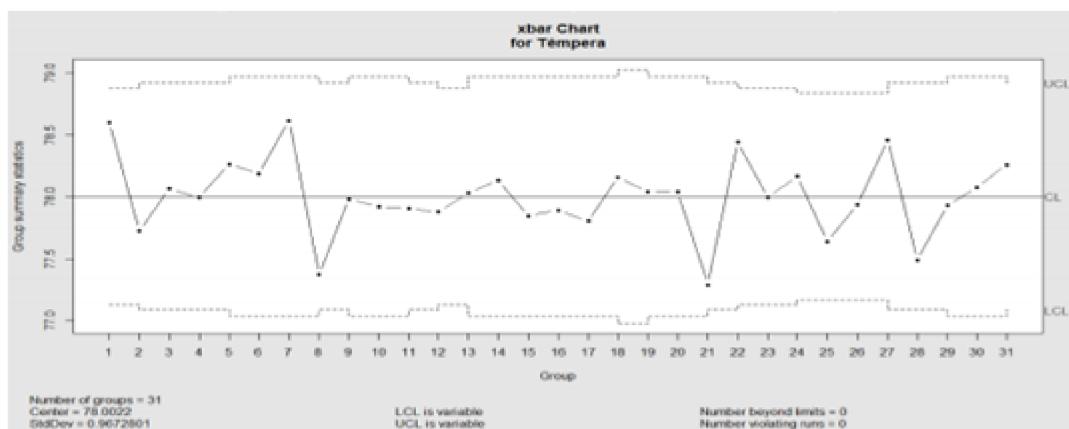


Figura 3 – Gráfico da média para a característica Têmpera

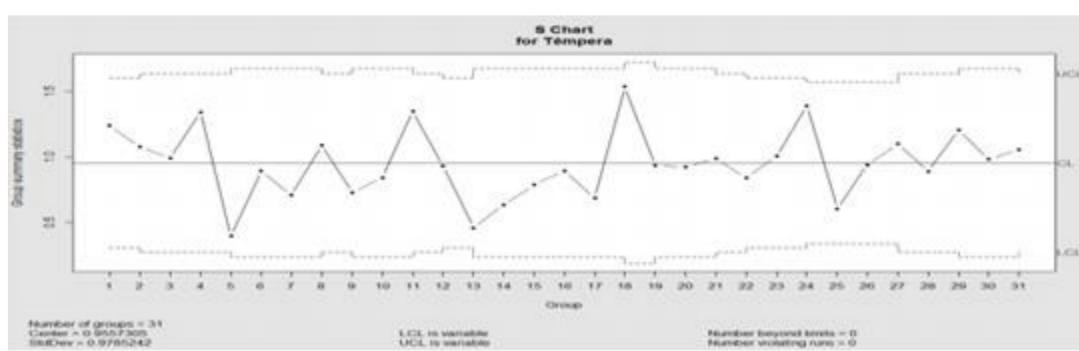


Figura 4 – Gráfico da variabilidade para a característica Têmpera

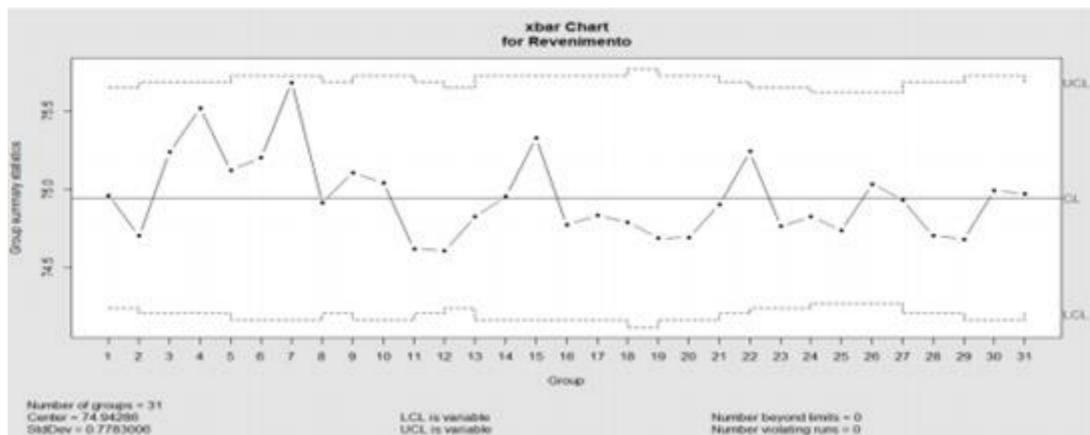


Figura 5 – Gráfico da média para a característica Revenimento

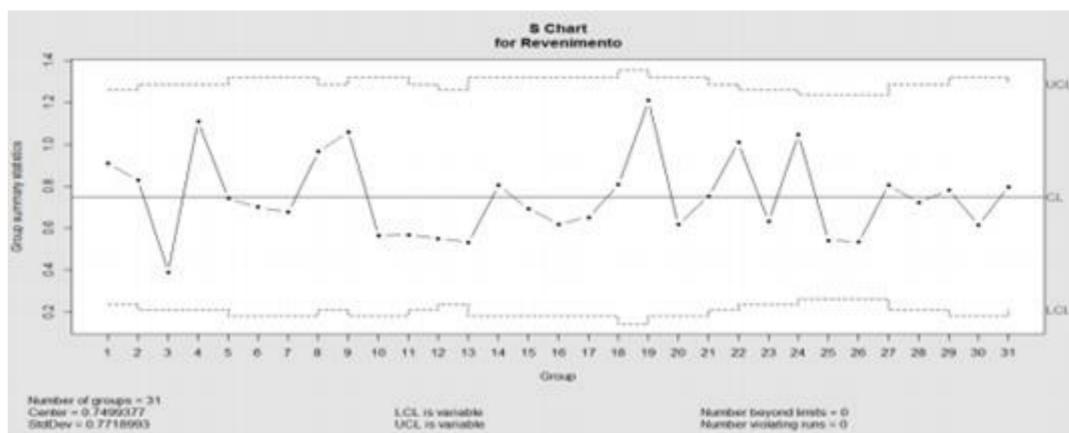


Figura 6 – Gráfico da variabilidade para a característica Revenimento

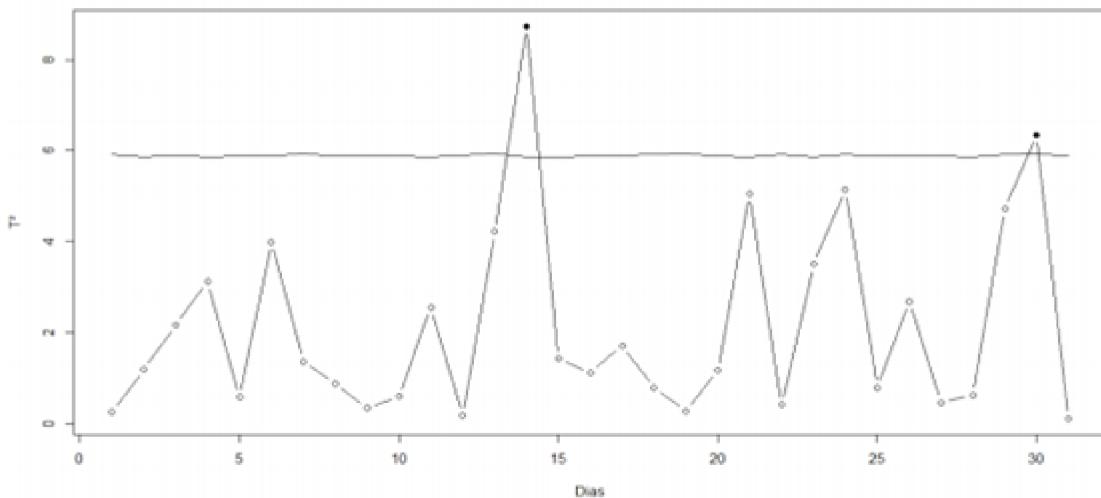


Figura 7 – T<sup>2</sup> de Hotelling para tamanho de amostra variável para as características Têmpera e Revenimento

Dessa forma, percebe-se a importância do uso do gráfico multivariado, pois as cartas univariadas não foram eficazes para o monitoramento da produção, uma vez que não detectaram nenhuma causa especial no processo. Já o gráfico T<sup>2</sup> de Hotelling detectou dois pontos fora de controle.

Nos gráficos de controle multivariado, a estabilidade do processo é obtida quando o vetor das médias, variâncias e covariâncias das variáveis do processo permanecem estáveis ao longo das observações. Entre outras técnicas estatísticas multivariadas, o PCA é o mais usual para o monitoramento de processos químicos.

O PCA é uma técnica estatística multivariada criada por Hotelling em 1933 e que se dedica à explicação da estrutura de variância-covariância existente em um conjunto de dados, utilizando-se combinações lineares das variáveis originais. Essa técnica apresenta dois objetivos principais: a redução de dimensionalidade e a interpretação de dados.

Embora  $p$  componentes sejam necessários para se reproduzir a variabilidade total de um sistema de interesse, geralmente, a maior parte desta variabilidade pode ser representada por um pequeno número  $k$  de componentes principais. Isto quer dizer que existe quase tanta informação em  $k$  componentes principais quanto nas  $p$  variáveis originais. A ideia geral do PCA é, portanto, que  $k$  componentes principais possam substituir, sem perda considerável de informação, as  $p$  variáveis originais. O conjunto original de dados, consistindo de  $n$  medições das  $p$  variáveis, é reduzido para um conjunto posterior formado por  $n$  medições de  $k$  componentes principais. Dessa forma, o PCA é capaz de revelar relacionamentos que não seriam previamente identificados com o conjunto original, o que resulta em uma interpretação mais abrangente do fenômeno em estudo.

A utilidade do PCA está, basicamente, na redução de dimensionalidade de vetores de entradas ou de saídas em determinados equacionamentos. Algumas das várias aplicações dessa técnica estão exemplificadas a seguir.

Lopes (2001) desenvolveu sua tese de doutorado através da criação de uma metodologia multivariada para determinar a confiabilidade e o tempo médio de falha de um produto, cujos componentes tenham seus tempos de vida correlacionados. Para isto, utilizou-se da análise de componentes principais. Numa aplicação apresentada em sua tese, a adoção do PCA permitiu a redução de um grande número de variáveis (80 peças originais) para um grupo de apenas três variáveis (componentes principais), optando-se, posteriormente, pela utilização de apenas um componente. Observou-se que a diferença entre a confiabilidade do primeiro componente principal e a média das confiabilidades das peças originais para três percentis foi de 0,2%.

Yacoub e MacGregor (2004) utilizaram o algoritmo de PCA para otimizar um processo químico com dez variáveis de resposta altamente correlacionadas e sete variáveis de entrada. Após a aplicação do método multivariado, apenas 3 componentes principais foram suficientes para representar o conjunto original.

Tino (2005) utilizou o PCA para desenvolver um método de regulagem de máquinas de injeção plástica, visando a redução de rejeitos. Os ajustes paramétricos foram propostos para serem feitos durante o processo de regulagem, utilizando a metodologia de Análise por Componentes Principais, com base nos dados amostrados por um sistema de monitoração em tempo real. Os parâmetros amostrados durante o período de regulagem de uma determinada máquina injetora, molde e matéria-prima, foram armazenados para que, em

seguida, com o PCA fosse possível a definição de grupos de parâmetros com maior representatividade.

A formulação matemática do PCA será descrita a seguir, em conjunto com a apresentação das etapas da metodologia para a construção e aplicação dessa técnica.

## 4. Metodologia

Apresenta-se, inicialmente, a formulação matemática da técnica PCA, e na sequência, o conjunto de etapas para a construção e um modelo PCA e sua aplicação.

### 4.1. Conceitos

A seguir serão apresentadas algumas definições importantes, tendo como referência Mingoti (2005):

#### Def. 1    *Vetor Aleatório*

Vetor em que cada componente é uma variável aleatória, ou seja, uma variável cujos valores dependem dos resultados de um experimento aleatório.

$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ , em que cada  $X_i$  é uma variável aleatória (por exemplo, uma temperatura).

#### Def. 2    *Vetor de médias*

Para um vetor aleatório  $X = [X_1 \ X_2 \ \dots \ X_p]'$ , o vetor de médias de X é dado por:

$$\mu = E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Em que cada média  $\mu_i$  indica a tendência central da distribuição de valores de  $X_i$ .

#### Def. 3    *Variância*

Indica a dispersão da variável  $X_i$  em relação à sua média, é uma medida da proximidade entre os valores de  $X_i$  e  $\mu_i$ .

$$Var(X_i) = \sigma_i^2 = \sigma_{ii}$$

O desvio padrão é denotado por  $\sigma_i$ . Valores grandes de  $\sigma_i$  indicam que os valores de  $X_i$  são mais dispersos em relação à média de sua distribuição.

#### Def. 4    *Covariância*

É a medida do grau de relacionamento linear entre duas variáveis aleatórias, ou seja, indica como variáveis variam de forma conjunta, sendo calculada por:

$$Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ij}$$

Não é simples julgar se a relação entre as variáveis é forte ou não, uma vez que depende da escala de medida das duas variáveis. A padronização da covariância é chamada de *correlação*:

$$\rho_{ij} = corr(X_i, X_j) = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Os valores de correlação são sempre entre -1 e 1, sendo que quanto mais próximo de 1 há forte indicação de relação linear positiva entre as variáveis  $X_i$  e  $X_j$ . E quanto mais próximo de -1, maior a indicação de relacionamento linear negativo.

#### Def. 5 Matriz de Covariâncias

É a matriz de variâncias e covariâncias do vetor aleatório  $X$ .

$$Cov(X) = Var(X) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

A matriz de covariâncias é simétrica:  $\sigma_{ij} = \sigma_{ji} \forall i \neq j$ , não negativa definida  $\Sigma \geq 0 \forall a_{px1} \in \mathbb{R}$ , cuja implicação é autovalores  $(\lambda_1, \lambda_2, \dots, \lambda_p)$  não negativos:  $\lambda_i \geq 0$ .

De modo semelhante, é possível definir uma *matriz de correlações*, composta uma medida de correlação entre cada par de variáveis:

$$\rho_{pxp} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}$$

#### Def. 6 Variância total

A variância total de um vetor aleatório  $X$  é a soma de todas as variâncias de suas variáveis:

$$tracô(\Sigma) = tr(\Sigma) = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$$

#### Def. 7 Combinações lineares

Seja  $X$  um vetor aleatório composto por  $p$  variáveis, com vetor de médias  $\mu$  e matriz de covariâncias  $\Sigma$ , e seja um vetor de constantes conhecidas:  $W_{pxp} \in \mathbb{R}^p$ , então  $PC$ , uma variável aleatória, é uma combinação linear das variáveis aleatórias do vetor  $X$  se:

$$PC = X_1 W_1 + X_2 W_2 + \cdots + X_p W_p$$

Média de PC:

$$\mu_{PC} = E(PC) = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_p \mu_p$$

Variância de PC:

$$\sigma_{PC}^2 = \Sigma = \sum_{i=1}^p a_i^2 \sigma_i^2 + \sum_{i \neq j} a_i a_j \sigma_{ij}$$

Para  $p$  combinações lineares do vetor  $X$ , e  $n$  observações, é possível expressar o conjunto de combinações num vetor aleatório da forma:

$$PC_{[n,p]} = X_{[n,p]} W_{[p,p]}$$

Em que  $W$  é a matriz de transformação linear composta por  $p$  autovetores. Neste caso, o vetor de médias e a matriz de covariâncias de  $PC$  são dados por:

$$\mu_{PC} = W'\mu \quad e \quad \Sigma = W'\Sigma W$$

No caso de PCA, cada combinação linear  $PC_i, i = 1, 2, \dots, p$  é denominada de Componente Principal, daí a denominação da técnica: Análise por Componentes Principais.

## 4.2. Etapas

O software utilizado para a implementação do modelo de Análise por Componentes Principais foi o Scilab. O motivo da escolha por esse software de computação numérica decorre principalmente do fato de ser código aberto e ter uma interface bastante amigável.

### 4.2.1. Construção do Modelo PCA

A primeira etapa para a implementação do PCA está na construção de um modelo matemático característico de “operações normais”. Os dados estão armazenados em uma matriz  $X_{n \times p}$ , em que  $p$  é o número de variáveis originais e  $n$  é o número de observações. Como existe a possibilidade de variáveis com ordens de grandeza muito díspares, é realizado um pré-tratamento dessas para que seja obtida uma matriz com dados padronizados. Para isso, é necessário calcular a média ( $\mu_i^2$ ) e o desvio padrão ( $\sigma_i$ ) de cada variável, sendo a matriz de dados padronizada ( $X_{n \times p}$ ) dada por:

$$X_{i,j} = \frac{Y_{i,j} - \mu_i^2}{\sigma_i}$$

em que  $X_{ij}$  é a j-ésima observação da i-ésima variável.

De posse dessa matriz padronizada, calcula-se a matriz de covariâncias  $E_{p \times p}$ .

$$E_{i,j} = Cov(X_i, X_j) = [(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ij}$$

Essa matriz dá origem ao vetor de autovalores ( $\lambda$  (px1)) e à matriz de autovetores associados ( $W_{[p \times p]}$ ) que descrevem a estrutura das variáveis originais. Os autovalores precisam estar organizados em ordem decrescente, uma vez que os maiores autovalores carregam maior percentual de explicação da variância total do sistema. É verificado que grande parte da variância total dos dados originais pode ser explicada por um número  $k < p$  de componentes principais. Esse percentual da variância explicada ( $\gamma$ ) pode ser calculada conforme a seguir:

$$\gamma = \frac{\sum_{i=1}^k \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i}$$

Isso significa que a matriz de autovetores ( $W_{p \times p}$ ) pode ter sua dimensão reduzida para uma quantidade  $k$  de colunas. A matriz reduzida de autovetores  $W_{p \times k}$  constitui o modelo PCA, característico de operações normais, dada pelo conjunto de dados originais,  $X$ . Esse modelo utilizado como um sistema de detecção de falhas, conforme o objetivo proposto neste trabalho. O programa completo implementado no Scilab se encontra no Anexo A.

#### 4.2.2. Determinação dos limites de controle

Para evitar a existência de dados “viciados”, um novo conjunto de dados ( $X_{2(nxp)}$ ) de operação normal é empregado no cálculo dos limites de controle. Os valores de média e desvio-padrão, bem como o modelo PCA obtido anteriormente (resumido em  $(\lambda, w)$ ), serão utilizados nesta etapa.

Novamente é feita a padronização dos dados da matriz  $X_2$ , utilizando a média e o desvio padrão da matriz X anterior.

$$X_{(i,j)} = \frac{X_{ij}^2 - \mu_i^2}{\sigma_i}$$

A partir da matriz  $X_2$  padronizada as Componentes Principais (PC) podem ser calculadas conforme a seguir, sendo  $k$  o número de componentes principais retido no modelo PCA:

$$PC_{(nxk)} = X_{(nxp)} \cdot W_{(nxk)}$$

A seguir, com base nas  $k$  componentes principais, calcula-se a estatística  $T^2$  de Hotelling, para cada observação:

$$T^2 = \frac{\sum_{i=1}^k PC_i^2}{\lambda_i}$$

Além dessa etapa de monitoramento, calcula-se também a estatística Q:

$$X_{(nxp)}^* = PC_{(nxk)} \cdot W_{(kxp)}$$

$$R = (X_{(nxp)} - X_{(nxp)}^*)$$

Para cada observação tem-se um valor de  $T^2$  e de Q, que podem ser utilizados para o cálculo dos limites de controle. Para ambos os parâmetros se escolheu uma taxa de alarmes falsos equivalente a 1% dos dados de operação normal. Deste modo, os limites das cartas de controle de  $T^2$  e Q são dados pelo percentil 99.

O modelo PCA obtido em 4.2.1, em conjunto com os limites de controle obtidos em 4.2.2 são usados para a construção das cartas de controle multivariadas. Exemplos dessas cartas são mostradas na Figura 8. Essas cartas são usadas na etapa de detecção de falhas, conforme descrição a seguir.

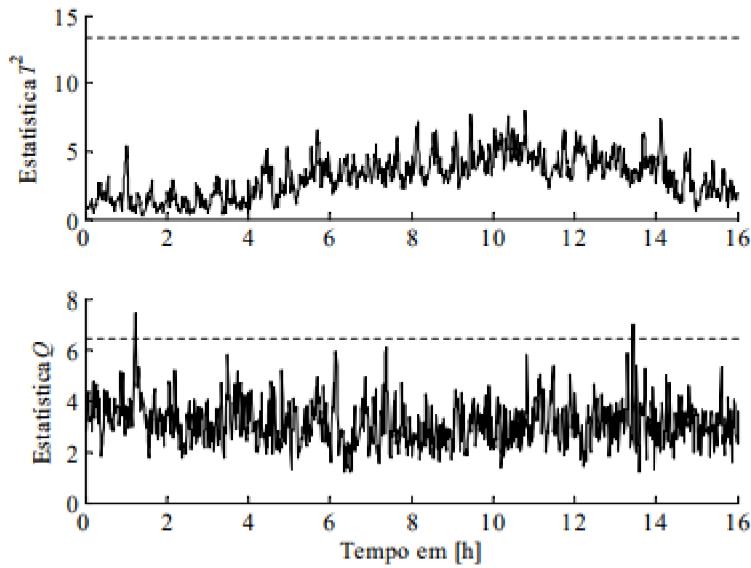


Figura 8 – Exemplos de cartas de controle multivariadas para os parâmetros  $T^2$  e  $Q$  em Operação Normal (Silva R. H., 2015)

#### 4.2.3. Detecção de Falhas

Com o modelo PCA ( $W$ ) e com os limites de controle de operação normal, um novo conjunto de dados em que é conhecida a existência de uma falha passa pela avaliação do modelo a fim de se verificar a sua capacidade de detecção.

Mais uma vez, os dados de falha são padronizados, baseando-se na mesma média e no mesmo desvio padrão da matriz original (dados de operação normal). Na sequência calculam-se os scores, conforme a seguir:

$$PC_{(nxk)} = X_{(nkp)} * W_{(pxk)}$$

E então  $T^2$  e  $Q$  que são apresentados em um gráfico (cartas de controle multivariada) de  $T^2$  vs observação e um de  $Q$  vs observação.

A existência da falha é dada pelos pontos acima do limite de controle de  $T^2$  e  $Q$  (calculados na etapa anterior). Exemplos de cartas de controle em operação com falha são mostrados na Figura 9.

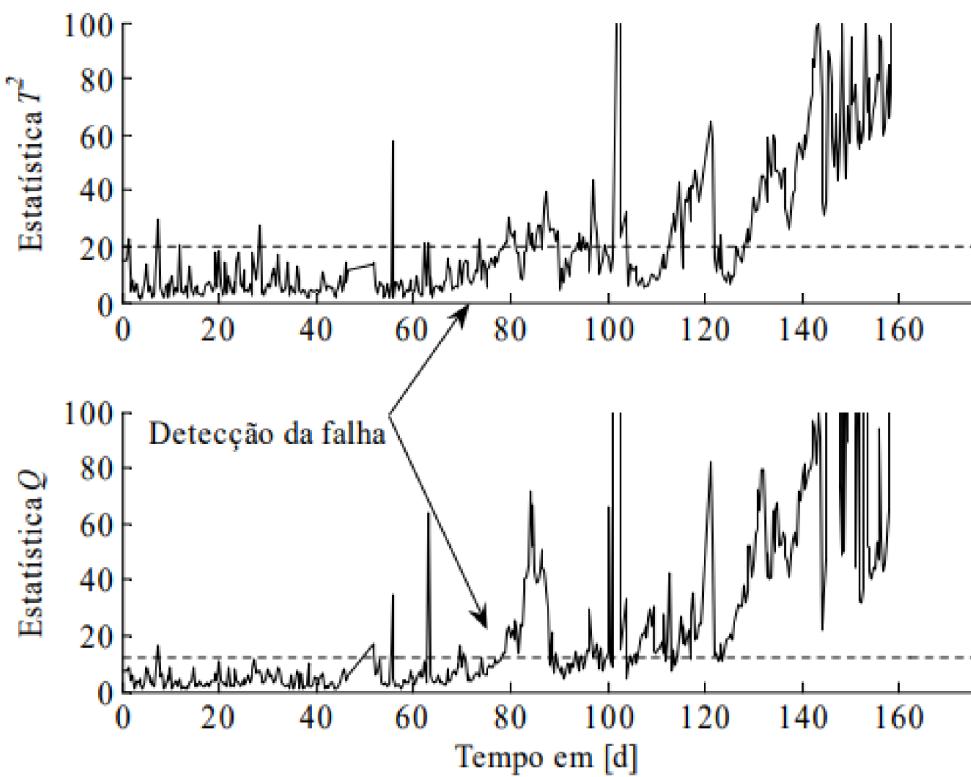


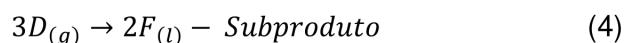
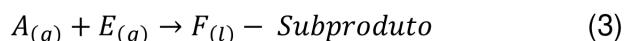
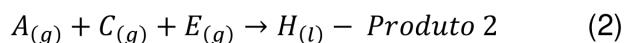
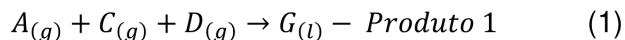
Figura 9 – Exemplos de Cartas de Controle Multivariadas para os parâmetros  $T^2$  e  $Q$  em Operação com Falha  
(Silva R. H., 2015)

## 5. Estudo de Caso 1: Benchmark Tennessee

### 5.1. Apresentação

O Benchmark Tennessee foi criado pela *Tennessee Eastman Chemical Company* para fornecer um processo industrial realista para avaliação de estratégias de controle e monitoramento (Downs & Vogel, 1993). Esse simulador tem sido largamente utilizado pela comunidade de monitoramento de processos como fonte de dados para comparação de várias abordagens. Russell et al. (2000) aplica o PCA e outras duas técnicas de análise multivariada no Benchmark e fornece resultados que serão utilizados neste trabalho para comparação e validação do script construído para a técnica de análise por componentes principais.

O processo do Tennessee é descrito pelas quatro seguintes reações químicas irreversíveis, exotérmicas e com cinéticas que seguem a Lei de *Arrhenius* de primeira ordem de acordo com as concentrações dos reagentes:



São quatro reagentes (A, C, D, E), um inerte (B), dois produtos de interesse (G e H) e um subproduto (F). Conforme visto no diagrama da Figura 10, o processo apresenta cinco unidades principais de operação: um reator, um condensador, um compressor, um separador vapor-líquido e uma coluna de dessorção ou *stripping*.

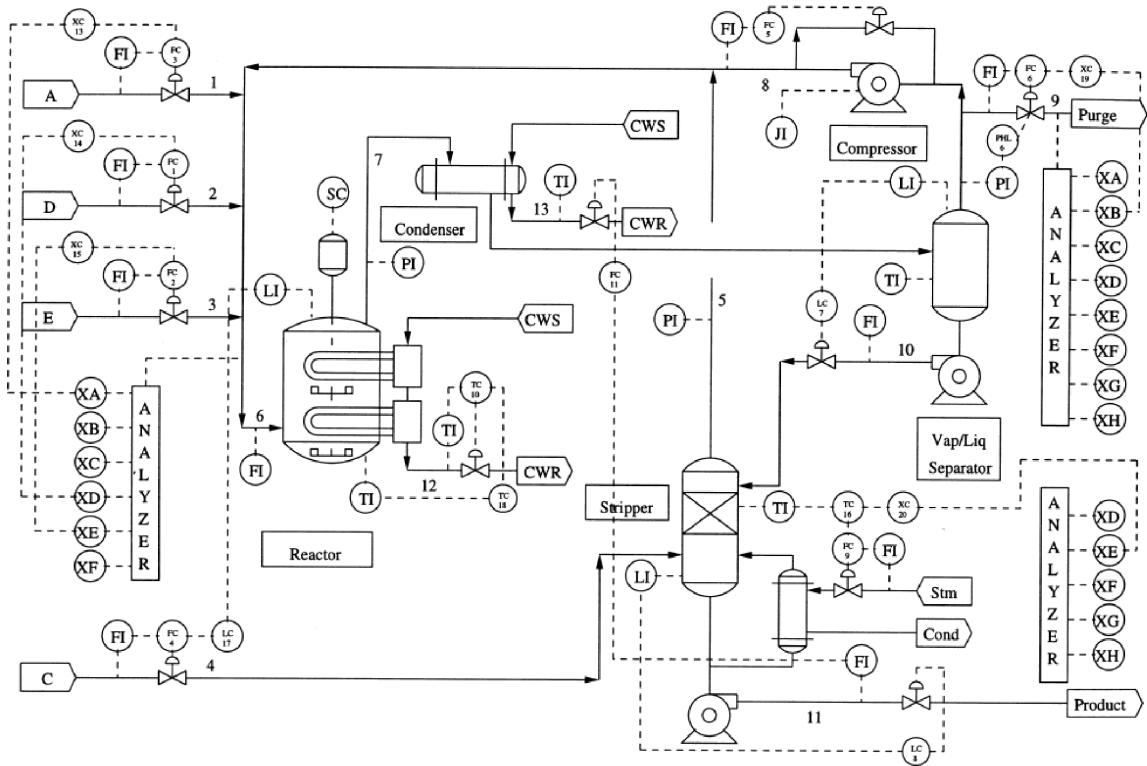


Figura 10 – Fluxograma esquemático do *Benchmark Tennessee* (Downs & Vogel, 1993)

Sobre o banco de dados, há 41 medições das variáveis de processo mais 12 variáveis manipuladas, totalizando 53 variáveis de entrada. O artigo de Downs & Vogel fornece 22 conjuntos de dados duplicados para serem testados, sendo um conjunto referente à operação normal da planta e 21 referente a operações com falhas, descritas na Tabela 1. Cada uma delas ilustra um aspecto diferente do funcionamento do processo. Todas as 21 falhas foram investigadas neste trabalho.

Tabela 1 – Descrição das falhas no *Benchmark Tennessee*

Variável	Descrição	Tipo
IDV (1)	Razão de alimentação A/C, B composição constante	Degrau
IDV (2)	Composição B, razão A/C constante	Degrau
IDV (3)	Temperatura de alimentação de D (linha 2)	Degrau
IDV (4)	Temperatura água de resfriamento entrada reator	Degrau
IDV (5)	Temperatura agua entrada condensador	Degrau
IDV (6)	Perda de alimentação A (linha 1)	Degrau
IDV (7)	Queda de pressão de C - diminuição de disponibilidade	Degrau
IDV (8)	Composições alimentação A, B e C (linha 4)	Variação aleatória
IDV (9)	Temperatura alimentação D (linha 2)	Variação aleatória
IDV (10)	Temperatura alimentação C (linha 4)	Variação aleatória
IDV (11)	Temperatura água de resfriamento entrada Reator	Variação aleatória
IDV (12)	Temperatura água entrada condensador	Variação aleatória
IDV (13)	Cinética do reator	Alteração lenta
IDV (14)	Válvula da água de resfriamento	Emperramento
IDV (15)	Válvula da água do condensador	Emperramento
IDV (16)	Desconhecido	
IDV (17)	Desconhecido	
IDV (18)	Desconhecido	
IDV (19)	Desconhecido	
IDV (20)	Desconhecido	
IDV (21)	Válvula linha 4 fixa em estado estacionário	Posição fixa

Fonte: Russell, Chiang, & Braatz (2000)

## 5.2. Metodologia

### 5.2.1. Construção do modelo PCA

Utilizou-se o primeiro conjunto de dados referente a uma operação normal para a construção do modelo PCA para o processo Tennessee. O vetor de autovalores, em ordem decrescente, é mostrado na Figura 11.

14,3431	1,8381	0,8656
8,7695	1,7986	0,8183
5,4465	1,7657	0,692
4,1424	1,7245	0,5774
4,0334	1,655	0,4599
3,8321	1,6062	0,3203
3,7549	1,5289	0,1028
3,2324	1,5057	0,0771
2,9758	1,4574	0,0275
2,6455	1,4141	0,0144
2,435	1,343	0,0103
2,3568	1,3371	0,0063
2,2751	1,1862	0,0037
2,1923	1,1706	0,0002
2,0961	1,1342	0
2,0086	1,0591	0
1,9951	1,0378	
1,9495	0,978	

Figura 11 – Vetor de autovalores em ordem decrescente ( $\lambda_{(52 \times 1)}$ )

A matriz  $W_{(52 \times 52)}$  contém os 52 autovetores com 52 elementos cada. Já cada valor  $\lambda$  é um autovalor correspondente a um autovetor. A variância explicada por cada autovalor permite encontrar um número significativo de componentes principais ( $k$ ) para o processo, conforme a Figura 12. Estabeleceu-se um critério de variância acumulada mínima de 55% e então reteve-se apenas 12 das 52 componentes principais, que explicam 56,6% da variância dos dados do processo em operação normal.

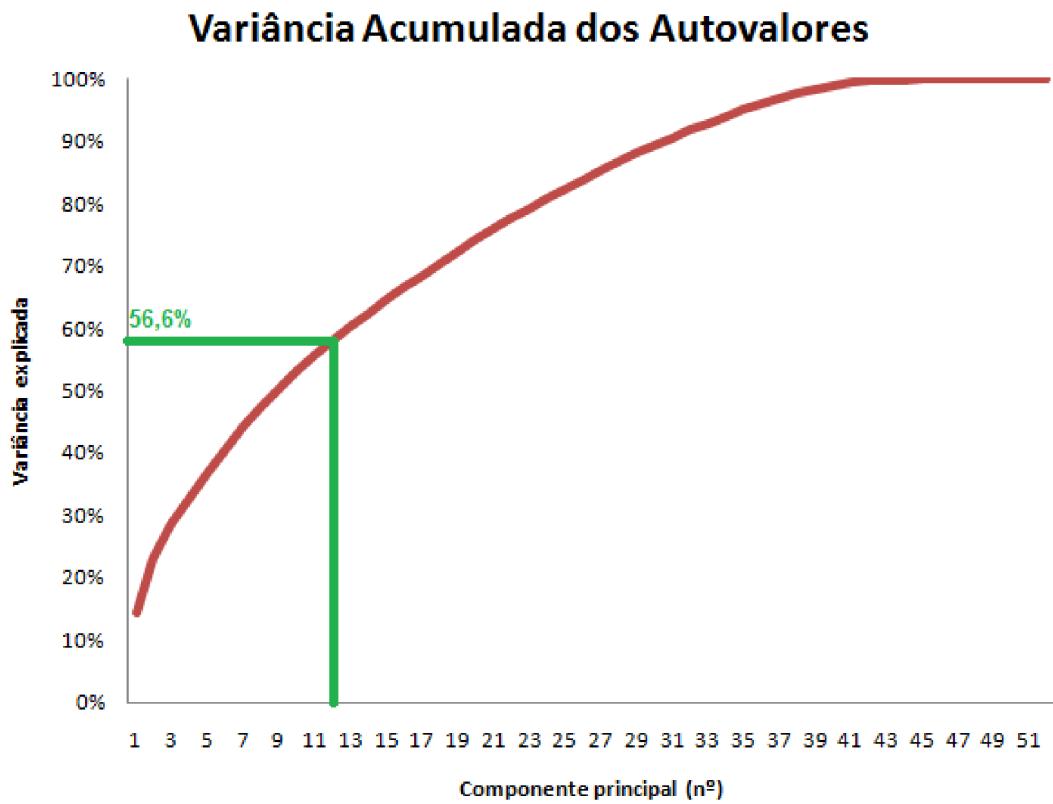


Figura 12 – Gráfico da variância explicada pelo número de componentes principais

### 5.2.2. Definição dos limites de controle

Nesta etapa, utiliza-se a matriz  $W_{(52 \times 12)}$  obtida no primeiro passo e o segundo conjunto de dados referentes a uma operação normal do processo. As estatísticas  $T^2$  de Hotelling e  $Q$  são calculadas e o limite superior de controle (LSC) é estabelecido no percentil 99 (dada taxa de alarmes falsos de 1%), conforme a Figura 13, sendo  $T^2_{LSC} = 31$  e  $Q_{LSC} = 49$ . Já o limite inferior de controle (LIC) é zero para ambas as métricas.

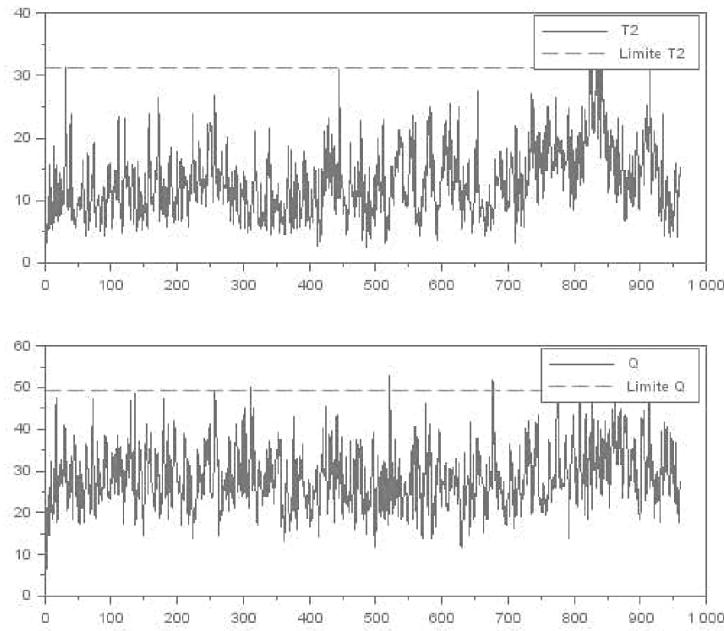


Figura 13 – Cálculo dos limites de controle

### 5.2.3. Detecção de falhas

De posse do modelo PCA e dos limites de controle, os conjuntos de dados para as 21 falhas do processo são investigadas. Um exemplo é apresentado na Figura 14, em que ambas as cartas de controle acusaram ocorrência da falha número 1, uma vez que os valores de  $T^2$  e  $Q$  extrapolam o limite superior de controle após o tempo decorrido de 8 horas, característico de operação normal. As cartas de controle de  $T^2$  e  $Q$ , para todas as falhas, encontram-se no Anexo C.

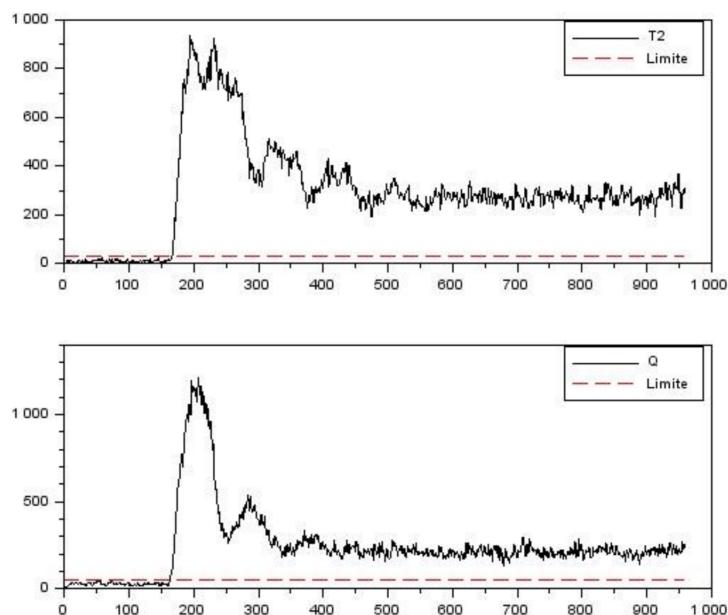


Figura 14 – Cartas de controle para a falha 1 do processo *Tennessee Eastman*

### 5.3. Resultados e Discussão

Pela análise visual das cartas de controle para cada conjunto de dados de falha, pode-se observar se o método PCA foi eficaz na detecção da falha. Em alguns casos, ambos os gráficos evidenciam a ocorrência de falha. Em outros, apenas uma carta acusa a falha. Pode-se usar ainda uma métrica de eficácia para o método, que é o tempo decorrido do processo em que a falha é evidenciada pelo PCA, sendo que o ideal é que isso ocorra no instante de ocorrência da falha (em  $t = 8$  horas). A Tabela 2 mostra o resumo dos resultados para este estudo de caso. Os casos em que o PCA falhou, seja por não detectar a falha ou por detectar com atrasos, foram evidenciados em negrito.

Tabela 2 – Resumo dos resultados da aplicação do PCA no processo Tennessee

Falha nº	Falha evidenciada pela carta $T^2$ ?	Falha evidenciada pela carta Q?	Identificação da falha (horas)
1	Sim	Sim	8
2	Sim	Sim	8
3	<b>Não</b>	<b>Não</b>	-
4	<b>Não</b>	Sim	8
5	Sim	Sim	8
6	Sim	Sim	8
7	Sim	Sim	8
8	Sim	Sim	<b>9</b>
9	<b>Não</b>	<b>Não</b>	-
10	Sim	Sim	<b>10</b>
11	Sim	Sim	8
12	Sim	Sim	8
13	Sim	Sim	<b>9</b>
14	Sim	Sim	8
15	<b>Não</b>	<b>Não</b>	-
16	Sim	Sim	<b>17</b>
17	Sim	Sim	<b>9</b>
18	Sim	Sim	<b>12</b>
19	<b>Não</b>	<b>Não</b>	-
20	Sim	Sim	<b>12</b>
21	Sim	Sim	<b>22</b>

Das 21 falhas do processo, 17 foram detectadas pelo modelo PCA, ou seja, em 81% dos casos. Porém, em quase metade deles (8 em 17), o método atrasou para acusar a falha, ou seja, em 47% dos casos.

Outra métrica referente à eficácia do modelo é o MDR (*Missed Detection Rate*) ou taxa de detecções perdidas, que foi calculada para cada conjunto de dados, conforme a Tabela 3. A taxa de detecção perdida é calculada da seguinte forma: a partir do momento em que a

falha ocorreu – um dado conhecido do Benchmark (observação de número 160) – contam-se quantas observações estiveram na região de operação normal (abaixo do limite de controle) em relação ao total de observações existentes a partir da observação 160. A Tabela 3 ainda compara o resultado com o valor encontrado por Russell et al. (2000). Os valores encontrados foram bem próximos daqueles do artigo de Russell, diferindo na maioria das vezes apenas na terceira casa decimal. Isso é um resultado importante para validar o script do PCA confeccionado pelos membros do presente trabalho, script que será usado nos próximos estudos de caso que envolvem dados reais da indústria.

Tabela 3 – Quadro resumo comparativo do MDR do presente trabalho e do artigo de Russell et al. (2000)

Falha nº	T <sup>2</sup> do presente trabalho	T <sup>2</sup> de Russell et al.	Q do presente trabalho	Q de Russell et al.
1	0,0088	0,008	0,0038	0,003
2	0,0213	0,020	0,0150	0,014
3	1,0000	0,998	0,9913	0,991
4	0,9625	0,956	0,0375	0,038
5	0,7750	0,775	0,7463	0,746
6	0,0125	0,011	0,0013	0
7	0,0788	0,085	0,0013	0
8	0,0338	0,034	0,0263	0,024
9	0,9988	0,994	0,9825	0,981
10	0,6513	0,666	0,6525	0,659
11	0,7888	0,794	0,3450	0,356
12	0,0225	0,029	0,0263	0,025
13	0,0613	0,060	0,0463	0,045
14	0,1238	0,158	0,0013	0
15	0,9800	0,988	0,9713	0,973
16	0,8313	0,834	0,7538	0,755
17	0,2538	0,259	0,1038	0,108
18	0,1138	0,113	0,1013	0,101
19	1,0000	0,996	0,8638	0,873
20	0,7025	0,701	0,5513	0,550
21	0,6950	0,736	0,5788	0,570