



Análise de Falha de Sensores Através de  
Métodos Estatísticos Baseados em Componentes  
Principais – Aplicação em Planta *Offshore*

Felipe Telles Lesbaupin

Monografia de Final de Curso

Orientador

Prof. Carlos André Vaz Junior, D.Sc.

Junho de 2014

**ANÁLISE DE FALHA DE SENSORES ATRAVÉS DE MÉTODOS  
ESTATÍSTICOS BASEADOS EM COMPONENTES PRINCIPAIS – APLICAÇÃO  
EM PLANTA *OFFSHORE***

***Felipe Telles Lesbaupin***

Monografia de Final de Curso submetida ao Corpo Docente da Escola de Química como parte dos requisitos necessários à obtenção do Grau de Engenheiro Químico.

Aprovado por:

---

Andrea Valdman, D.Sc.

---

Florian Alain Yannick Pradelle, M. Sc.

---

Lara Arinelli, B.Sc.

Orientado por:

---

Carlos André Vaz Junior, D. Sc.

Rio de Janeiro, RJ – Brasil  
Junho de 2014

Lesbaupin, Felipe.

Análise de Falha de Sensores Através de Métodos Estatísticos Baseados em Componentes Principais – Aplicação em Planta *Offshore* / Felipe Lesbaupin. Rio de Janeiro: UFRJ/EQ, 2014.

xiv, 84p.; ii

(Monografia de Final de Curso) - Universidade Federal do Rio de Janeiro, Escola de Química,  
2014.

Orientador: Carlos André Vaz Junior.

1. Análise de Componentes Principais. 2. Falha de Sensores. 3. Segurança de Processos. 4. PCA. 5. Monografia de Final de Curso. (Graduação - UFRJ/EQ). 5. Carlos André Vaz Junior. I. Análise de Falha de Sensores Através de Métodos Estatísticos Baseados em Componentes Principais – Aplicação em Planta *Offshore*.

## Dedicatória

À Mari. Porque sem ela eu não sairia da faculdade.

Tardou, mas, enfim, chegou!

Resumo de Monografia de Final de Curso apresentada à Escola de Química como parte dos requisitos necessários para a obtenção do grau de Engenheiro Químico.

**Análise de Falha de Sensores Através de Métodos Estatísticos Baseados em Componentes Principais – Aplicação em Planta *Offshore***

Felipe Telles Lesbaupin

Junho de 2014

Orientador: Prof. Carlos André Vaz Junior, D.Sc

Este trabalho teve como objetivo testar a metodologia estatística de análise de componentes principais (PCA) para detecção de falhas de sensores em uma planta *offshore*. Os componentes principais apresentam vantagens quando se lida com sistemas multivariáveis de grande escala - exige menor esforço computacional - ao reduzir a dimensão dos dados processados. Dados reais de sensores submarinos que falharam foram submetidos ao método PCA, desenvolvido em Matlab. Os quantificadores SPE, Spca, Sdist e SF, além de análise gráfica, foram comparados na avaliação dos resultados obtidos. Diferentes parâmetros foram variados para se garantir uma análise ampla dos casos estudados. Utilizando dados em estado estacionário, foi possível efetuar a análise de diferentes tipos de falhas reais. Através de regimes transientes, verificou-se o comportamento do modelo frente a variações bruscas de sinal dos sensores. Ao final, pôde-se concluir que, apesar de bons resultados apresentados pelo índice SPE, a análise PCA isoladamente permanece insuficiente na detecção e análise de falhas em sensores reais.

# Sumário

<b>1</b>	<b>Introdução .....</b>	<b>11</b>
1.1	Contextualização.....	11
1.2	Objetivo.....	12
<b>2</b>	<b>Operação Offshore da Shell Brasil Petróleo .....</b>	<b>13</b>
2.1	Shell no Brasil.....	13
2.2	Indústria <i>offshore</i> .....	13
2.3	Esquemático do processo submarino.....	15
2.4	Transmissão de dados e controle remoto .....	17
2.5	Histórico de falhas.....	18
<b>3</b>	<b>Revisão Bibliográfica .....</b>	<b>20</b>
3.1	Falhas de sensores e suas consequências.....	20
3.2	Detecção e análise de falhas.....	22
3.3	Análise de Componentes Principais – PCA.....	23
3.4	Teoria do método PCA .....	24
3.5	Análise gráfica dos resultados.....	26
3.6	Quantificadores ou Indicadores.....	28
<b>4</b>	<b>Implementação .....</b>	<b>30</b>
4.1	<i>Softwares</i> utilizados.....	30
4.2	PI .....	30
4.3	Coleta dos dados.....	34
<b>5</b>	<b>Metodologia.....</b>	<b>35</b>
5.1	Estratégia de análise .....	35
5.2	Quantificadores ou Indicadores.....	37

5.2.1	<i>Q, T<sup>2</sup> de Hotelling e Erro Quadrático de Predição (SPE)</i> .....	37
5.2.2	<i>Fator de similaridade de angulação – Spca</i> .....	39
5.2.3	<i>Fator de similaridade de distância – Sdist</i> .....	40
5.2.4	<i>Fator de Similaridade - SF</i> .....	41
5.2.5	<i>Índice de Validade de Sensor - SVI</i> .....	41
5.3	<i>Análise de Sensibilidade</i> .....	42
5.3.1	<i>Grau de representatividade dos componentes principais</i> .....	42
5.3.2	<i>SPE e o efeito memória - <math>\lambda</math></i> .....	42
5.3.3	<i>Composição do SF</i> .....	42
5.3.4	<i>Instante da falha</i> .....	43
5.4	<i>Aplicação do método</i> .....	45
5.5	<i>Casos Estudados</i> .....	45
5.5.1	<i>Caso 1</i> .....	46
5.5.2	<i>Caso 2</i> .....	47
5.5.3	<i>Caso 3</i> .....	48
5.5.4	<i>Caso 4</i> .....	49
5.5.5	<i>Caso 5</i> .....	50
<b>6</b>	<b>Resultados</b> .....	<b>52</b>
6.1	<i>Resultados dos quantificadores</i> .....	53
6.1.1	<i>SPE</i> .....	53
6.1.2	<i>Índices de Similaridade</i> .....	54
6.2	<i>Comparação entre os casos</i> .....	56
6.3	<i>Resultados caso a caso</i> .....	57
6.3.1	<i>Caso 1</i> .....	57
6.3.2	<i>Caso 2</i> .....	58
6.3.3	<i>Caso 3</i> .....	59

6.3.4	Casos 4 e 5.....	60
6.4	Intervalo total e influência do instante da falha.....	60
6.5	Porcentagem de representação pelos PCs .....	61
<b>7</b>	<b>Discussão .....</b>	<b>62</b>
7.1	Desafios encontrados.....	62
7.2	Discussão dos Resultados .....	64
<b>8</b>	<b>Conclusão.....</b>	<b>66</b>
	<b>Bibliografia.....</b>	<b>67</b>
	<b>Anexos .....</b>	<b>70</b>
	Anexo A.....	70
	Anexo B .....	77



## Índice de Figuras

Figura 2-1: Exemplo do instante em que um sensor falha. ....	15
Figura 2-2: Esquemático simplificado do sistema submarino de produção. ....	16
Figura 2-3: Representação do sistema de transmissão de dados de plantas submarinas. (6) .....	18
Figura 3-1: Falha concomitante de múltiplos sensores devido a alta descarga elétrica. ....	20
Figura 3-2: Perda de sinal por parte de um sensor. Leitura desaparece, como se não existisse o equipamento em questão.....	21
Figura 3-3: Comparação gráfica entre Fase A (preto) e Fase B (vermelho). ....	27
Figura 3-4: Comparação gráfica entre Fase A (preto) e Fase B (vermelho). ....	27
Figura 3-5: Comparação gráfica em 3 dimensões entre Fase A (preto) e Fase B (vermelho).....	28
Figura 4-1: Resumo da regra de compressão por variação absoluta. Neste exemplo, apenas os pontos preenchidos são salvos. ....	31
Figura 4-2: Resumo da regra de compressão por variação de tendência.....	32
Figura 4-3: Sensor de pressão no momento da falha com a ferramenta citada ativa.....	33
Figura 4-4: Zoom do sensor da Figura 4.3 no momento da falha com a ferramenta citada ativa.	33
Figura 5-1: Exemplo gráfico do SPE, em preto.....	39
Figura 5-2: Representação esquemática da posição da falha em relação ao período analisado. .	43
Figura 5-3: Representação esquemática da análise feita para comparar diferentes durações do período com falha. ....	44
Figura 5-4: Tabela-modelo com resumo dos resultados obtidos cruzando todas as análises de sensibilidade. Os dados com fundo cinza são aqueles obtidos a partir de períodos estáveis para comparação direta. ....	44

Figura 5-5: Sensores analisados no Caso 1. ....	47
Figura 5-6: Sensores analisados no Caso 2. A Figura mostra o instante da primeira falha. ....	48
Figura 5-7: Caso 3: dois sensores perdem suas leituras e passam a indicar zero simultaneamente. .....	49
Figura 5-8: Caso 4: parada repentina de uma BCS devido a problemas na operação. ....	50
Figura 5-9: Aumento gradual da produção da BCS avaliado no Caso 5. ....	51
Figura 6-1: Porcentagem de falhas corretamente identificadas nos casos 1, 2 e 3 pelo SPE. ....	53
Figura 6-2: Porcentagem de alarmes falsos gerados pelo SPE nos casos 4 e 5 .....	54
Figura 6-3: Porcentagem acertos na detecção de falhas por parte dos índices de similaridade estudados. Valores de $\alpha$ entre parêntesis. ....	55
Figura 6-4: Porcentagem de alarmes falsos gerados pelos índices de similaridade estudados. Valores de $\alpha$ entre parêntesis.....	55
Figura 6-5: Dados modelados do Caso 1. Em preto é representada a Fase A e em vermelho, a Fase B. ....	58
Figura 6-6: Dados modelados do Caso 2. Em preto é representada a Fase A e em vermelho, a Fase B. ....	59
Figura 6-7: Influência do intervalo total da análise sobre a detecção de falhas.....	61
Figura 6-8: Influência da representação mínima acumulada pelos componentes principais sobre a detecção de falhas. ....	62
Figura 7-1: Momento em que dois sensores apresentam perda de sinal durante algumas horas. .....	63
<i>Figura B-1: Caso 2B, dados experimentais em preto e dados modelados em vermelho. Dois sensores durante a Fase A são representados nessa figura. O modelo mostra boa aderência. ....</i>	<i>77</i>

<i>Figura B-2: Caso 3A, dados modelados da Fase A, em preto, e da Fase B, em vermelho. Um sensor sadio e um sensor ruim foram plotados. ....</i>	<i>78</i>
<i>Figura B-3: Análise do SPE no Caso 2A A linha vermelha indica o limite aceitável. ....</i>	<i>78</i>
<i>Figura B-4: Análise do SPE no Caso 2B. A linha vermelha indica o limite aceitável. ....</i>	<i>79</i>
<i>Figura B-5: Análise do SPE no Caso 3A. A linha vermelha indica o limite aceitável. ....</i>	<i>79</i>
<i>Figura B-6: Análise do SPE no Caso 4. A linha vermelha indica o limite aceitável. ....</i>	<i>80</i>
<i>Figura B-7: Análise do SPE no Caso 5. A linha vermelha indica o limite aceitável. Nota-se que existem pequenos desvios já na Fase A devido a pequenas variações nos dados usados. ....</i>	<i>80</i>

# 1 Introdução

## 1.1 Contextualização

Problemas operacionais causados pela falha de sensores são muito comuns na indústria. Uma leitura incorreta de dados de vazão, pressão ou temperatura pode resultar até mesmo na parada total de uma planta, embora frequentemente sem necessidade real, gerando perdas enormes para a operação. Por esse motivo, estudos são feitos no sentido de identificar essas falhas antes que elas acarretem qualquer prejuízo (1). Isso permite que os operadores ignorem leituras de dispositivos em mau funcionamento.

Muitos dos sensores instalados são atrelados a *set points* de parada parcial ou total da produção. Ou seja, caso sua leitura ultrapasse os limites pré-determinados, o sistema de segurança irá atuar para interromper aquela situação de operação inaceitável. Nesse sentido, a detecção e análise de falhas apresentada neste trabalho vem em prol da otimização do *uptime* da planta, prevenindo contra paradas desnecessárias (1).

Com o avanço tecnológico, a quantidade de sensores utilizada atualmente é muito maior do que décadas atrás. Esse aumento do número de sensores trouxe consigo um aumento na quantidade de dados gerados. A ocorrência de falhas, considerando todo o conjunto de sensores, também cresceu consideravelmente. Afinal, ainda que os novos equipamentos sejam mais confiáveis, uma planta com 1000 sensores está mais sujeita a sofrer falhas em sensores do que uma com apenas 50.

Diversos métodos de detecção de falhas de sensores já foram estudados e muitos mostraram ótimos resultados a partir de dados simulados. A grande maioria destes métodos se baseia em análises estatísticas, necessárias para analisar a grande quantidade de dados gerados pelas plantas. Esses métodos são capazes de definir padrões de comportamento dos dados durante fases estáveis de produção. Quando ocorre uma falha de sensor, o método detecta o desvio do comportamento esperado e então alerta os operadores (1).

Modelos autorregressivos, redes neuronais e análise de componentes principais (PCA) são alguns dos métodos estatísticos mais citados na literatura (2). No presente trabalho, a técnica de análise de componentes principais (*Principal Components Analysis*, ou PCA, em inglês) foi escolhida para analisar as falhas reais de sensores instalados em duas plantas de produção de petróleo *offshore*. Ambas

as plantas são operadas pela Shell Brasil Petróleo Ltda. e estão atualmente em atividade no Brasil, na Bacia de Campos. Os dados foram cedidos pela empresa para fins acadêmicos com a restrição do detalhamento do processo. Todos os casos estudados neste trabalho são reais e os dados foram obtidos diretamente do sistema da empresa. Todos os exemplos apresentados também são baseados nesses dados, exceto quando explicitamente retirados de referência literária.

## **1.2 Objetivo**

A ideia central deste trabalho é aplicar a metodologia PCA (Principal Component Analysis) em dados reais de operação para analisar a viabilidade da utilização de tal técnica para auxílio aos operadores na detecção de falhas. Devido ao uso de dados reais, é esperado que o método esteja exposto a imperfeições dos sistemas de transmissão e coleta de dados. Desde o equipamento que faz as medições – o sensor em si – até o software de interface – PI –, passando por quilômetros de fibra ótica e cabos elétricos, a amplificação dos ruídos e da imprecisão dos dados é um fator que certamente dificulta a análise proposta. Apesar de inúmeros artigos apresentarem bons resultados de técnicas baseadas na análise de componentes principais para detecção de falhas, sua aplicação direta em dados industriais brutos pode se mostrar difícil e restrita.

## 2 Operação Offshore da Shell Brasil Petróleo

### 2.1 Shell no Brasil

A Shell Brasil possui atualmente dois ativos em produção no Brasil, ambos na Bacia de Campos: Parque das Conchas e Bijupirá-Salema. O último, mais antigo, entrou em operação em 2003 e teve recentemente sua vida útil estendida para 2022. Já o Parque das Conchas começou a produzir em 2009 e ainda está em expansão, com novos poços sendo perfurados e novos reservatórios sendo estudados.

A entrada da Shell no mercado de óleo e gás no Brasil completou 100 anos em 2013. Em 1913, a empresa iniciou suas atividades vendendo querosene de aviação para os primeiros voos no país. Com o tempo, ampliou sua atuação para outros combustíveis e chegou ao mercado de distribuição, que atualmente é gerido pela Raízen, *Joint-Venture* da Shell com a Cosan (3) (4).

Após a abertura do mercado brasileiro de produção e exploração de petróleo e gás para a iniciativa privada, em 1997, diversas empresas passaram a operar na costa brasileira em parceria com a Petrobras. Em 2001, a Shell comprou a empresa Enterprise, que possuía os direitos de exploração do campo de Bijupirá-Salema, criando a Shell Brasil. O campo viria a produzir seu primeiro poço em Setembro de 2003 (4).

A partir de 2005, a Shell Brasil adquiriu outros blocos em leilões públicos, dos quais até o momento apenas um entrou em operação: o Parque das Conchas.

Mais recentemente, a empresa elevou seus investimentos nos recursos nacionais. Isto culminou na compra, em 2013, de 20% do campo de Libra, uma das maiores reservas de petróleo do pré-sal, em parceria com a Petrobrás e outras 3 empresas privadas.

A Shell Brasil atualmente possui cerca de 800 funcionários locais e estrangeiros trabalhando no Rio de Janeiro, São Paulo e Espírito Santo (4).

### 2.2 Indústria offshore

A indústria *offshore* de produção de petróleo e gás natural se diferencia das demais devido ao acesso remoto a todos os dispositivos e máquinas submersos, muitas vezes, a quilômetros de

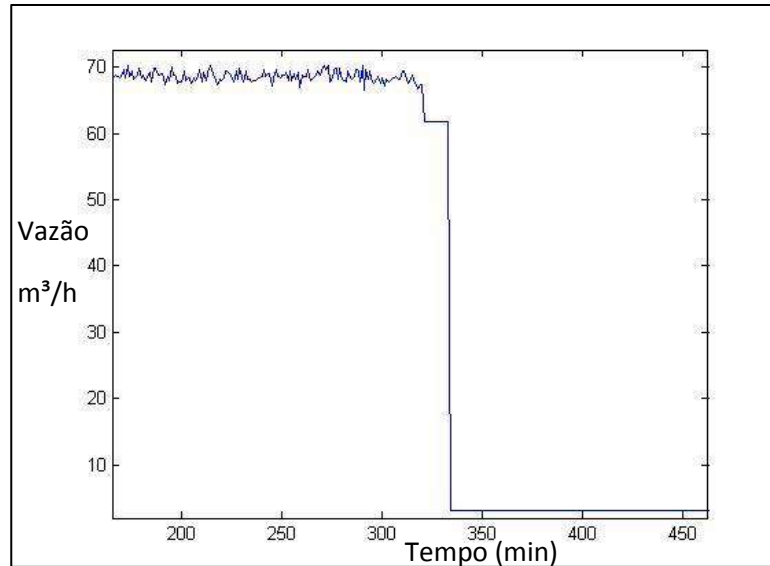
profundidade. As características particulares desses equipamentos submarinos faz com que o preço dos mesmos seja muito elevado se comparado a plantas tradicionais, em terra. Por consequência, a quantidade de sensores existente costuma ser muito reduzida, e a possibilidade de troca ou reparo de equipamentos nem sempre existe no curto ou médio prazo (5).

Conforme mencionado, um sensor que atinge um valor limite de operação – *High High* ou *Low Low* – pode ativar uma sequência de parada do processo. Supõe-se assim que o sensor esteja refletindo uma grave falha real do processo. Caso o sistema de controle, através da análise estatística aqui proposta, identifique que este sensor encontra-se na verdade em falha, esse encadeamento pode ser inibido e reduzido a um alarme para os operadores.

Um exemplo ilustrativo: um sensor de pressão em uma tubulação está programado para desencadear uma parada total da produção de um determinado poço caso sua leitura atinja 350 bar. Este valor equivale ao limite de resistência do material da tubulação. Esse parâmetro (pressão) está intimamente ligado à temperatura do fluido e pressão da linha a montante e a jusante, além da vazão. É, portanto, inconcebível que apenas a leitura desse sensor específico (pressão) apresente repentinamente valores acima de 350 bar sem que os outros sensores o acompanhem com semelhante variação. Assim, o método de detecção de falhas em sensores seria capaz de identificar a falha do dispositivo e inibir a sequência de parada gerada por esse sensor. Alarmes seriam emitidos para que os operadores pudessem, então, tomar uma decisão sobre o que fazer.

Além de interromper a produção imediatamente, um erro do sensor poderia ainda mobilizar a equipe responsável da empresa no sentido de compreender se aqueles dados são reais ou se houve falha. Usualmente esses dispositivos são programados para, em caso de falha, emitir um valor nitidamente irreal. Em geral este valor é igual a zero ou igual ao máximo (ou mínimo) para o qual o sensor é calibrado. Note que zero muitas vezes pode ser um valor factível, então esta escolha de *default* para sinal de erro pode não ser adequada.

De uma forma ou de outra, uma maneira mais garantida de diagnosticar que um sensor falhou é analisar se há qualquer variação em torno do valor medido. Sensores de pressão, temperatura e vazão sempre apresentam pequenas e constantes oscilações, os ruídos, que frequentemente deixam de existir quando o sensor falha. Vê-se então apenas uma linha horizontal. No exemplo da Figura 2-1, um tubo de Venturi lendo valores próximos a 69 m<sup>3</sup>/h cai para 3,2 m<sup>3</sup>/h em instantes. Os ruídos do sinal cessam.



*Figura 2-1: Exemplo do instante em que um sensor falha.*

### 2.3 Esquemático do processo submarino

A Figura 2-2 foi elaborada a partir dos diagramas do campo Parque das Conchas (6). Ela mostra de forma simplificada uma das plantas submarinas de produção de petróleo da Shell Brasil, desde o momento em que os hidrocarbonetos deixam o poço até sua chegada ao navio-plataforma FPSO (*Floating, Production, Storage and Offloading*), onde ocorre o processamento primário do petróleo – a separação óleo-água-gás. Nesse esquema é possível ver as principais válvulas e sensores. Observa-se ainda um item crucial para viabilizar a produção desse reservatório: a bomba centrífuga submersível (BCS), indicada pela letra B na Figura. Note que em um processo real existem mais sensores e válvulas do que o mostrado na Figura 2-2, porém o processo pode ser compreendido em linhas gerais apenas pelo que está indicado.

A válvula mais próxima ao poço (representada em vermelho) é a válvula de segurança, projetada para fechar e conter o fluxo em caso de problemas ou emergências. Ela é precedida por 2 sensores, um de pressão e um de temperatura, que monitoram o fundo do poço, próximo ao reservatório. A sequência de sensores e a segunda válvula, uma válvula-agulha, resumem as funcionalidades presentes na árvore de natal (equipamento típico na cabeça de poços de petróleo): monitoramento do poço e controle de vazão.



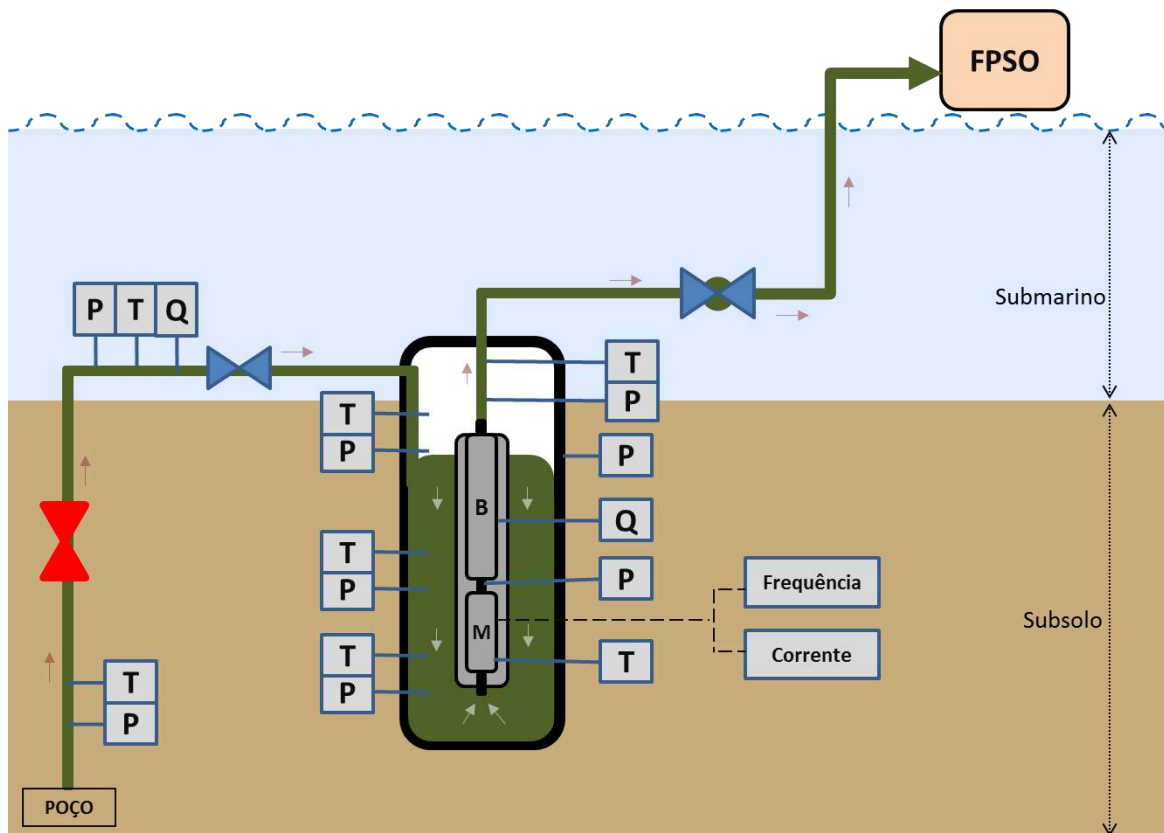


Figura 2-2: Esquemático simplificado do sistema submarino de produção.

Devido à grande profundidade do solo marinho fez-se necessário o uso de um método de elevação artificial. Neste caso, foi escolhido utilizar bombas (BCS) para garantir que os hidrocarbonetos consigam vencer a barreira de pressão hidrostática. Outra técnica existente é a elevação por gás, presente na outra planta submarina da Shell Brasil, esta última estando em solo menos profundo.

Explicando com mais detalhes o esquema apresentado: o fluido deixa o poço impulsionado pela alta pressão do reservatório e segue por tubulações até o módulo de bombeio, que consiste na estrutura cilíndrica chamada de *caisson* e na BCS, localizada no centro deste. O *caisson* é um buraco cavado no solo marinho, de dezenas de metros de profundidade, revestido com uma estrutura metálica para abrigar o fluido e a BCS. O fluido despejado dentro do *caisson* preenche-o até certo nível e entra na bomba pelo ponto de sucção, localizado no fundo. Antes de entrar na bomba centrífuga em si o fluido passa em torno do motor da BCS, resfriando-o. A bomba promove então o ganho de carga necessário para que o fluido chegue ao FPSO.

Nesse esquema são mostrados alguns dos sensores de Pressão, Temperatura, e Vazão presentes ao longo das tubulações. Sensores de corrente e frequência são indicadores do status da BCS. Essas duas grandezas são de extrema importância já que a bomba não pode operar aquém de uma frequência mínima nem além de uma amperagem máxima (limites operacionais do modelo) (3).

O sistema de controle da bomba, a fim de manter a pressão de sucção constante, regula o nível de líquido dentro do *caisson*. A bomba é acelerada ou desacelerada de acordo com a variação do mesmo. Por isso a existência de 3 sensores de pressão dentro do *caisson*: eles determinam o nível de líquido baseado nas diferentes profundidades de cada um. Um sensor de nível comum não pode ser utilizado nesse caso devido à frequente formação de espuma dentro do *caisson*, que dificulta a identificação da interface gás-líquido. Além disso, os sensores de pressão também conseguem identificar indiretamente a densidade do fluido, parâmetro importante para o controle da bomba e impossível de ser definido com sensores de nível. Esse controle faz com que variações no escoamento vindo de cada poço possam ser amortecidas pelo aumento ou diminuição do nível, sem que a bomba seja afetada diretamente, de forma brusca. A manutenção da estabilidade das BCS apesar das flutuações dos escoamentos dos poços é essencial e um dos maiores desafios nessa operação (3).

Vale também ressaltar que cada BCS pode receber fluidos de diversos poços diferentes desde que não ultrapasse o limite máximo de vazão da mesma.

## **2.4 Transmissão de dados e controle remoto**

Todos esses sensores e válvulas mostrados são controlados ou monitorados remotamente, diretamente do FPSO. Existem Módulos de Controle Submarinos (MCS) que promovem a transmissão dos dados dos sensores para os computadores no navio (Figura 2-3). Também são responsáveis pelo sentido inverso, traduzindo os comandos em atuações de válvulas. O MCS é uma espécie de computador submarino; cada unidade é capaz de controlar diversos sensores e válvulas. É, portanto, um item crítico para o processo e por isso costuma ter redundância interna (3).

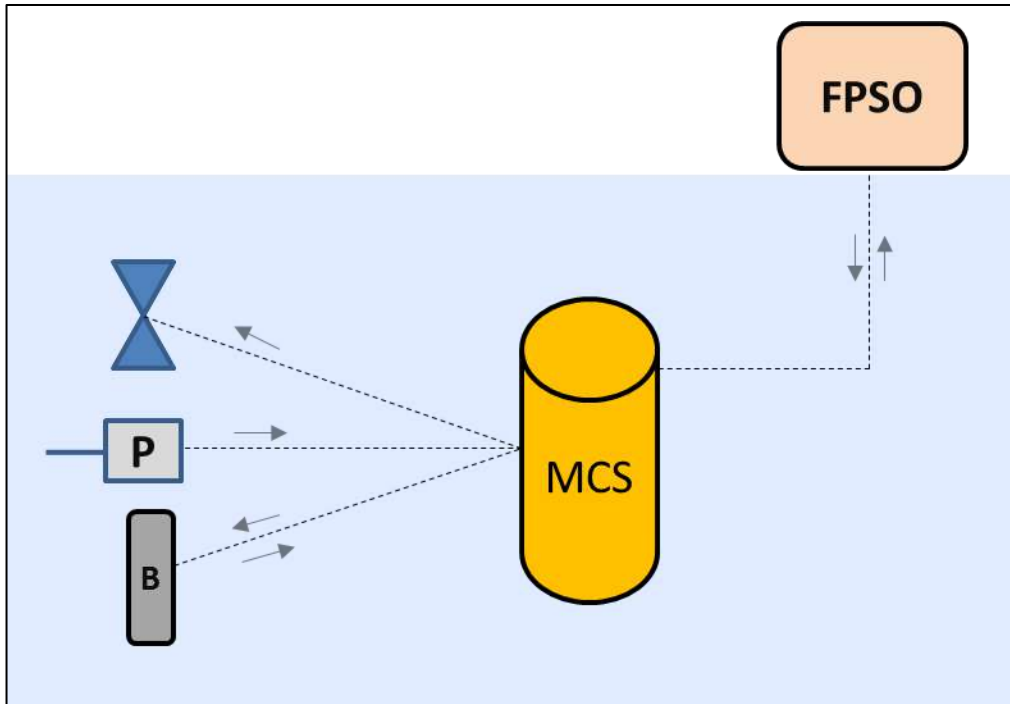


Figura 2-3: Representação do sistema de transmissão de dados de plantas submarinas. (6)

Os dados recebidos pelos MCS chegam ao computador central da plataforma e são transmitidos ao servidor do PI, software de monitoramento usado pela empresa. Os funcionários têm acesso a todas as informações em tempo real através da interface do PI.

## 2.5 Histórico de falhas

Em ambos os campos *offshore* que a Shell opera no país, problemas com sensores submarinos não fazem parte das suas principais preocupações, devido a uma taxa de falha relativamente baixa. A empresa realizou recentemente um levantamento dos sensores submarinos que apresentaram problemas em todos os campos em produção *offshore* da Shell nas Américas. O resultado demonstrou que o desempenho geral da Shell Brasil é bom.

Com o alto custo de cada sensor, projetado para as condições de alta pressão do fundo do mar, não existe viabilidade de dispor de uma grande redundância. Sendo assim, quando um sensor falha, em

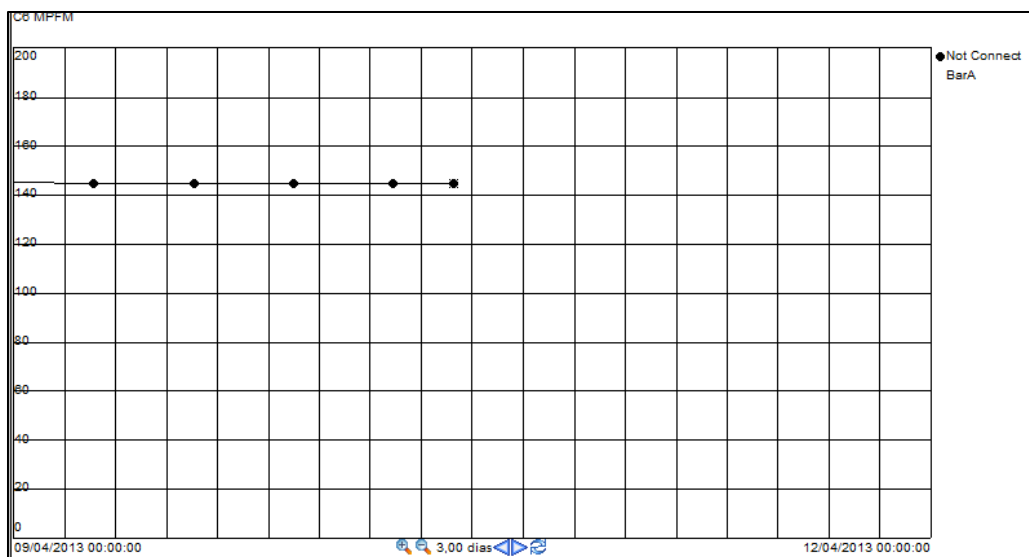
geral, uma informação insubstituível é perdida. Deste modo, apesar de não haver um número excessivo de falhas, toda perda é preocupante.

Analisando historicamente as falhas ocorridas nos dois campos nacionais citados, foi possível verificar as seguintes tendências (7):

- O maior índice de falhas recai sobre os sensores mais profundos, de temperatura e pressão, que medem dados do poço na interface com o reservatório. Localizados no fundo da coluna de produção, são instalados comumente durante a completação dos poços. Muitos desses equipamentos apresentaram falhas logo no início da operação – nunca funcionaram –, denotando um problema de fabricação ou de instalação. Vale ressaltar que esses sensores de fundo de poço são instalados em pares para garantir uma redundância. Porém, ambos costumam falhar concomitantemente, perdendo-se de uma vez só toda a informação do reservatório próximo àquele poço. No total, oito poços perderam esses pares de sensores, dentre um total de 40 poços operados pela Shell Brasil.
- O segundo maior índice de falhas pertence aos sensores localizados nas árvores de natal, que ficam nas cabeças dos poços. Incluem-se aqui sensores de pressão e temperatura principalmente.
- Poucos sensores localizados nas tubulações, bombas e estruturas submarinas de distribuição de fluxo apresentaram falhas.

É importante frisar que, para a análise de componentes principais aqui proposta, os casos onde houve falha logo no início da operação não podem ser usados. Isso se deve ao fato de não haver quantidade suficiente de dados sadios daquele sensor. Dados sadios são necessários para treinar o método, que irá utilizá-los como base para comparação.





*Figura 3-2: Perda de sinal por parte de um sensor. Leitura desaparece, como se não existisse o equipamento em questão.*

Olhando por outra perspectiva, um exemplo conhecido na indústria *offshore* relacionado a problemas com sensores é o caso da BP (*British Petroleum*) no poço Macondo, em 2010 (8). Uma série de erros em cascata, todos durante a etapa final de completação do poço, acarretou na perda de contenção do mesmo. Com isto, houve ingresso descontrolado de hidrocarbonetos na sonda (plataforma de perfuração), causando a explosão da mesma e a consequente morte de 11 pessoas, além de um vazamento de grande escala no mar do Golfo do México (8).

O primeiro problema ocorreu quando o poço já havia sido perfurado e a sonda já estava se preparando para fechar o poço e sair do local. O processo de vedação do poço com concreto apresentou problemas, porém estes não foram identificados. Ao realizarem o teste de pressão para verificar se o poço estava seguro, o sensor acusou vazamento, mostrando um valor inesperadamente elevado de pressão. Os operadores, confiantes de que o trabalho havia sido bem feito, assumiram que o sensor estaria com defeito e descartaram o resultado do teste. Declararam então que o sensor não estava mostrando valores reais e seguiram adiante, sem repetir o teste. Ao final, com a continuação da série de falhas na operação, o poço fluiu para dentro da sonda, ocasionando o grave acidente.

Esse incidente foi seguido do vazamento de milhares de barris de petróleo para o mar que durou 3 meses.

Isso mostra o quão crucial uma leitura de sensor pode ser durante operações *offshore*. Nessa indústria, muitas vezes a única informação que se tem sobre um poço é um único sensor de pressão, e através dele diferencia-se a normalidade da irregularidade. No caso de Macondo o sensor estava certo e foi desconsiderado. Porém o oposto também pode ocorrer.

### **3.2 Detecção e análise de falhas**

A detecção e análise de falhas consistem no monitoramento dos sensores com o fim de identificar desvios inesperados de comportamento de forma precoce e acurada. Esses desvios podem ocorrer devido a uma mudança de regime na operação ou a uma falha de medição causada por um sensor defeituoso, entre outros.

Os modelos matemáticos que monitoram o processo a fim de alarmar os operadores para mudanças podem ser “físicos” ou “estatísticos”. Independentemente de sua natureza, um método de detecção e análise de falha segue a seguinte metodologia: à medida que novos dados são gerados eles são comparados ao padrão de comportamento normal da planta. Se existe uma divergência muito grande, o modelo identifica essa alteração inesperada, acusando a irregularidade (9).

Os modelos físicos costumam ser muito complexos e exigem alta capacidade de processamento. São conjuntos de equações que descrevem o processo levando em consideração reações químicas, transferência de massa e calor, entre outros fenômenos. Esses modelos fenomenológicos precisam descrever todas as características da operação, de cada máquina e cada corrente, a fim de simular a variação esperada das medições e compará-las com as mudanças reais.

Os modelos estatísticos são aqueles que “ignoram” a fenomenologia dos dados analisados e trabalham apenas com a relação numérica entre eles. Se resumem à busca de correlações numéricas a fim de diagnosticar variações fora do comum no processo.

Em plantas modernas, com dados sendo gerados a cada instante, é muito difícil haver um tratamento eficiente de toda a informação vinda dos sensores. Tratar os dados brutos e extrair informação relevante é fundamental para a segurança de processos. Com essa finalidade o método de análise de componentes principais (PCA) pode ser utilizado. O PCA reduz enormemente o volume de dados ao reduzir a dimensão dos mesmos. Em seguida utiliza-se metodologia para avaliar o cenário e detectar falhas.

Diversas outras técnicas estatísticas, diferentes ou complementares do PCA, todas com resultados promissores, já foram propostas para análise de falhas (10) (1) (11) (12) (13) (14). Estas são capazes de realizar detecção de vazamentos, controle de qualidade e detecção de falha de sensor. Dependendo da sensibilidade do método, é possível detectar desde equipamentos que param de funcionar até sinais que divergem levemente do valor esperado.

Métodos como redes neuronais, modelos autorregressivos, classificadores hierárquicos, entre outros, são algumas das abordagens que já foram utilizadas para detecção de falhas ou vazamentos (2) (15). Muitas técnicas foram também propostas para se aprimorar os resultados obtidos a partir dos modelos PCA puros: mineração de dados (16), clusters (17), análise qualitativa de comportamento (10) e grafos orientados (11). Entretanto, visto que a análise proposta no presente trabalho é pouco usual na literatura – uso de dados reais de processos *offshore* submetidos à análise PCA – e não se tem certeza dos resultados que serão encontrados, optou-se por focar na análise somente por componentes principais. Aprimoramentos e estudos de novos métodos estatísticos, em substituição ou em complemento ao PCA, são assuntos para futuros trabalhos.

Análises estatísticas envolvendo PCA para melhoria de processos *offshore* já foram propostas, porém usando dados simulados (17) (15) (1).

### **3.3 Análise de Componentes Principais – PCA**

A análise de componentes principais (PCA) é uma técnica estatística amplamente utilizada que se baseia em sistemas multivariáveis. Ela apresenta vantagens em relação ao custo computacional e à facilidade de implementação. Possui ainda grande versatilidade: podem-se aliar ao PCA diversos outros métodos estatísticos. A análise PCA pode ser aplicada a diversos tipos de dados, incluindo séries temporais.

Séries temporais são definidas por conjuntos de dados distribuídos cronologicamente e separados por intervalos de tempo. Esses intervalos podem ser definidos de acordo com a especificidade de cada caso, podendo variar de segundos a meses (9).

Sistemas multivariáveis são aqueles que apresentam dimensão superior ou igual a 2, onde cada dimensão denota uma variável do sistema (2). Costumam caracterizar-se pela grande quantidade de dados e pela dificuldade de equacionamento.



Modelos PCA são generalistas e podem ser usados em diferentes áreas. Por exemplo, para se entender a ocorrência de turistas em certa região, assim como seus padrões de consumo. Através de dados com gênero, idade, estado civil, nacionalidade e renda média, o modelo PCA é capaz de gerar um padrão que auxilia as autoridades a entenderem qual o foco que deve ser tomado nas propagandas e nos investimentos em infraestrutura (18) (19).

O PCA pode ainda ser usado no processamento de imagens ou monitoramento de indicadores de qualidade de produtos (12).

Assim, a análise de componentes principais pode ser usada para pré-tratamento dos dados e redução da dimensão dos dados tratados (10), como também para executar a análise em si.

Pesquisadores e indústrias do mundo inteiro já tentaram – e conseguiram – implementar metodologias das mais diversas em sistemas de controle muitas vezes complexos (20). Grande parte dessas técnicas se baseia, de algum modo, na análise dos componentes principais. Isso deve muito à simplicidade desse método. Entretanto, usualmente técnicas adicionais (complementares) são utilizadas para obter resultados melhores e mais confiáveis do que aqueles obtidos através do método PCA apenas (10) (18) (19) (13) (15) (11) (17) (21).

### **3.4 Teoria do método PCA**

Esse método puramente estatístico utiliza a matriz covariância dos dados para determinar correlações existentes entre eles (22). Essa correlação pode ser física, química, ou simplesmente uma relação de causa e consequência (aumento dos preços e número de consumidores, por exemplo). Embora os fenômenos estejam refletidos nos dados e correlações, não é necessária uma modelagem fenomenológica para a utilização desse método. Basta que haja uma relação ou padrão entre os dados analisados.

A base matemática deste método é a matriz covariância, que quantifica, par a par, as correlações dentro do conjunto de dados analisado. Essa matriz, aliada aos conceitos de autovetor e autovalor, promove uma seleção dos componentes principais (*principal components*, ou PCs, na sigla em inglês). A matriz de componentes principais é usada, por exemplo, para verificar a presença de falhas.

Os PCs são aqueles que representam sinteticamente todo o conjunto de dados. Em termos numéricos, trata-se dos autovetores de maior autovalor associado. A escolha dos PCs promove uma redução da dimensão total dos dados sem perda significativa de informação útil.

O método PCA foi escolhido para analisar as falhas reais de sensores da Shell Brasil. O método foi implementado em Matlab (vide Anexo A).

Um conjunto de dados com  $s$  sensores (colunas) e  $2n$  instantes de tempo (linhas) é gerado na planta. Em seguida divide-se esse conjunto em duas matrizes de dimensões iguais  $[n, s]$ , adjacentes cronologicamente e de forma que a falha, caso exista, ocorra apenas na segunda matriz. Trata-se da Fase A (primeiro momento) sem falha, e da Fase B (segundo momento) com falha ou não.

Como pré-tratamento dos dados foi escolhido subtrair de todos os sensores a média, coluna a coluna, e dividir cada valor pelo desvio padrão daquela série (coluna). O desvio padrão utilizado nessa divisão é aquele dos dados sem falha (Fase A) (22) (9).

Calcula-se então a matriz covariância dos dados dos sensores. Os componentes principais são os autovetores da matriz covariância que apresentam maiores autovalores. Ou seja, os autovetores de maior autovalor associado são mantidos na matriz de componentes principais até que a soma dos autovalores represente a porcentagem escolhida do total. Escolhe-se, portanto  $k$  autovetores para desempenharem o papel de componentes principais. Foram avaliados diferentes graus de representatividade por parte desses componentes: 90%, 05% e 99% (23) (9) (20) (24).

Os autovetores restantes, por possuírem baixa representatividade, são descartados. Obtém-se então, ao fim, uma matriz de autovetores com reduzido número de colunas em relação à original. Esta matriz de dimensão reduzida  $[s, k]$  é chamada de MC. Ela é então multiplicada por sua transposta (Equação 1), obtendo-se uma matriz quadrada (matriz C) de dimensão igual ao número de sensores estudados  $[s, s]$ .

$$C = MC * MC^T \quad \text{Equação 1}$$

Multiplicando-se a matriz de dados original (após seu pré-tratamento) pela matriz C do modelo PCA (obtida conforme explicado no parágrafo anterior), é obtida a matriz de dados modelados, conforme a Equação 2.

$$T_{\text{mod}} = C * T_{\text{exp}}^T \quad \text{Equação 2}$$

### 3.5 Análise gráfica dos resultados

A análise gráfica é um método muito didático, porém pouco versátil, de se avaliar os resultados do método.

Plota-se na mesma figura os dados modelados (obtidos pela Equação 2) antes e depois da falha, todos projetados no subespaço vetorial de dimensão dois, cuja base vetorial é composta pelos dois componentes principais mais importantes (9) (25). Ou seja, independentemente da dimensão dos dados originais ou da quantidade de PCs escolhidos, esta análise leva em conta apenas a representação por dois componentes principais. No caso de falha espera-se que os dados da Fase A projetados sobre esses vetores vão gerar um *cluster* e os dados da Fase B, outro. Observa-se então graficamente, no domínio  $R^2$ , se há algum desvio nítido de comportamento dos dados, denotado pela separação desses *clusters* (9).

Na Figura 3-3 é apresentado um exemplo de comparação entre Fase A e Fase B através da análise gráfica. Como se pode ver, o *cluster* da Fase B, que apresenta falha, aparece separado do *cluster* da Fase A.

Na Figura 3-4 é mostrado outro exemplo de falha de sensor corretamente identificada pela análise gráfica. Note que no período de falha – Fase B, em vermelho – dois *clusters* podem ser observados, ambos distantes do *cluster* da Fase A.

É possível ampliar essa análise para a terceira dimensão, acrescentando-se o terceiro componente de maior relevância na análise (Figura 3-5).

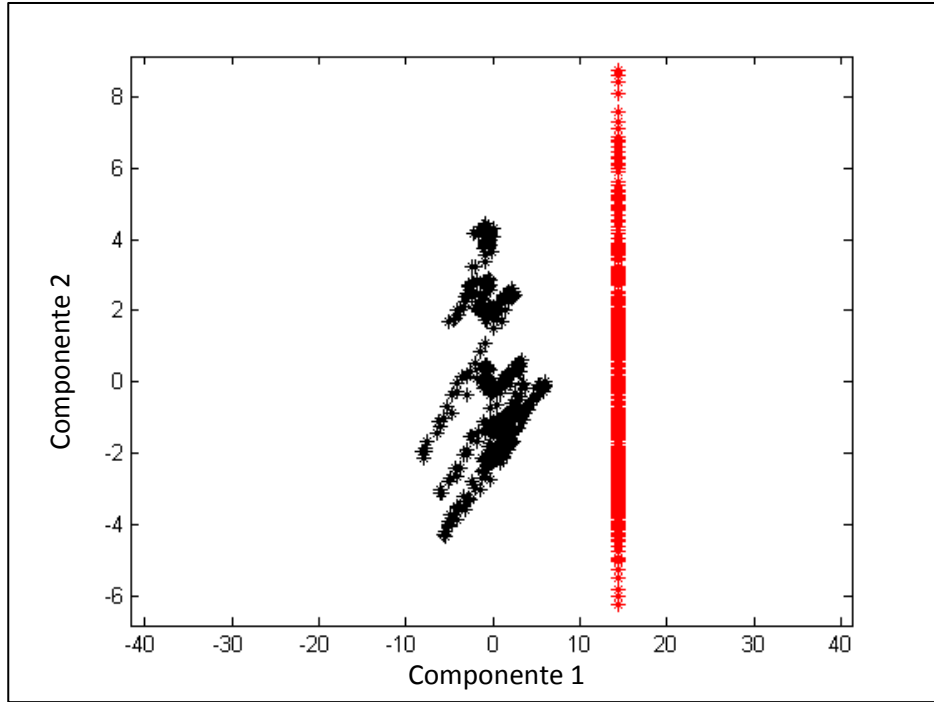


Figura 3-3: Comparação gráfica entre Fase A (preto) e Fase B (vermelho).

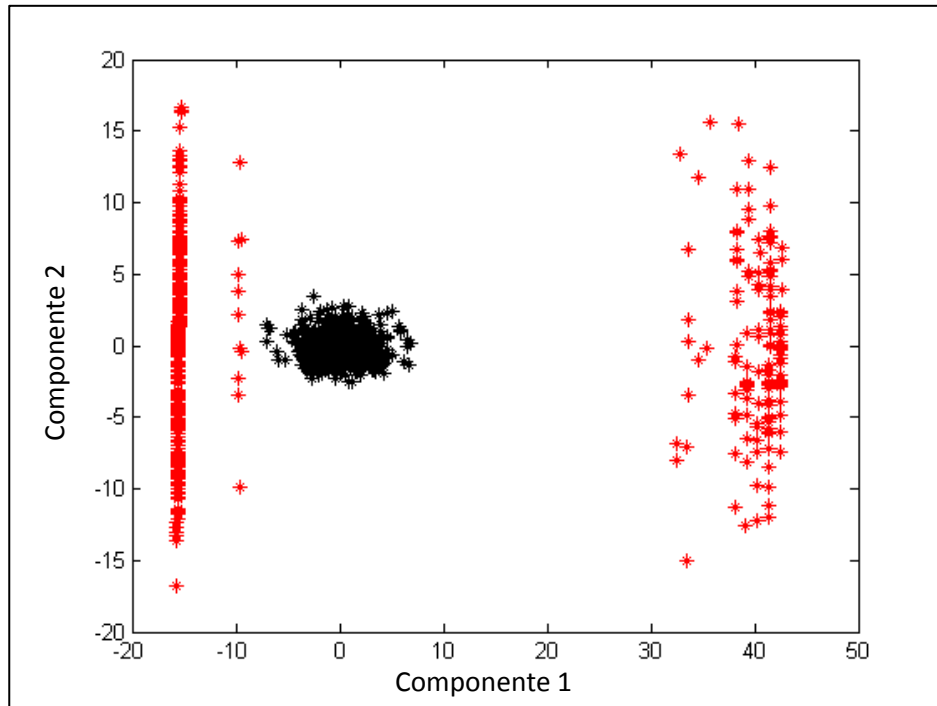
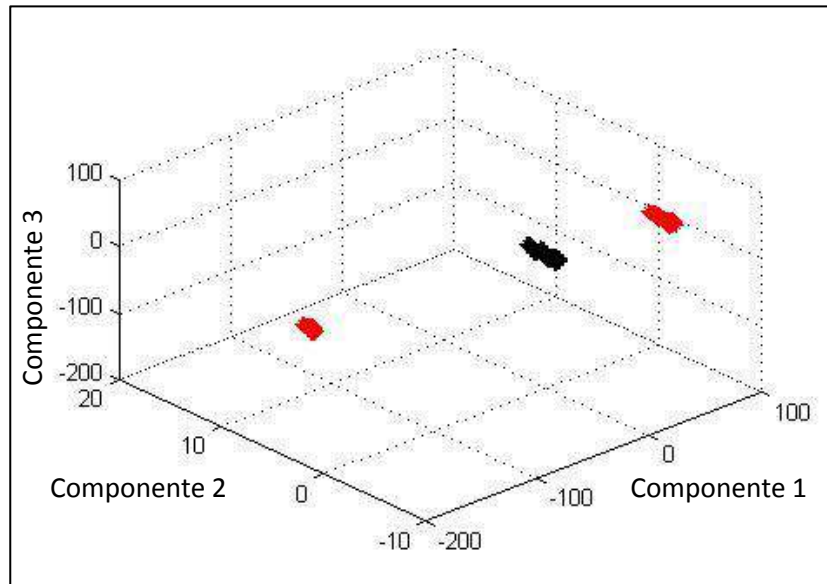


Figura 3-4: Comparação gráfica entre Fase A (preto) e Fase B (vermelho).



*Figura 3-5: Comparação gráfica em 3 dimensões entre Fase A (preto) e Fase B (vermelho).*

Uma grande desvantagem do método gráfico é sua limitação a três dimensões. Caso se esteja trabalhando com um grande número de sensores, é provável que um maior número de componentes principais seja necessário para representá-los fielmente. Dessa forma, como em geral mais de três componentes principais são necessários para garantir a representação mínima dos dados pelo modelo, o uso de “quantificadores” é necessário.

### 3.6 Quantificadores ou Indicadores

Quantificadores são indicadores estatísticos criados para se avaliar os modelos em relação à aderência ao conjunto de dados analisado. São grandezas numéricas para determinar o grau de confiança que se pode ter naquele modelo, naquele caso. Alguns são índices estatísticos amplamente usados, enquanto outros foram criados especificamente para medir a acurácia dos modelos PCA. Diversos quantificadores são discutidos na literatura. Alguns destes são baseados na identificação do erro (diferença) entre os dados modelados e os dados experimentais (24). Trata-se do Q,  $T^2$  de Hotelling e SPE (*Square Prediction Error*, ou erro quadrático de predição). Estes três índices estatísticos calculam o erro associado ao modelo comparando os dados gerados por ele aos dados originais.

Outros quantificadores medem o grau de similaridade entre dois modelos PCA diferentes (20) (24) (23) (14). Estes, chamados de Spca, Sdist e SF, comparam modelos PCA obtidos a partir de conjuntos de dados diferentes.

Há ainda o SVI (*Sensor Validity Index*, ou índice de validade de sensor), que além de acusar a falha indica sua origem: qual sensor falhou.

Neste estudo foram aplicados diferentes quantificadores e métodos de avaliação dos resultados devido à grande divergência, na literatura, sobre qual é o melhor índice. Cada um dos artigos publicados apresentou melhores resultados com um quantificador específico. Sendo assim, somente depois de avaliados os resultados obtidos que foi feita a discussão sobre qual indicador seria mais confiável para o presente caso.

## 4 Implementação

### 4.1 Softwares utilizados

Para aplicação dos métodos estatísticos aos dados reais, dois *softwares* foram empregados: Matlab (versão 7.12, MathWorks) e PI (versão 3.2, OSI Soft). O primeiro foi usado para programação do código matemático, cálculos e análise dos resultados (26) (27). O segundo foi usado para pesquisa e coleta dos dados a serem analisados. Foi necessário ainda o uso do Microsoft Excel 2010 para importação dos dados obtidos no PI e conversão dos mesmos para um formato aceito pelo Matlab.

### 4.2 PI

PI é uma plataforma de monitoramento de dados de processos. É o software utilizado mundialmente pela Shell para monitorar suas plantas. Através dele o usuário acompanha em tempo real os dados de todos os sensores e ainda tem acesso ao histórico de dados. É usado tanto *in situ*, nas plantas, pelos operadores, quanto à distância, pela equipe de suporte com base nos escritórios da empresa.

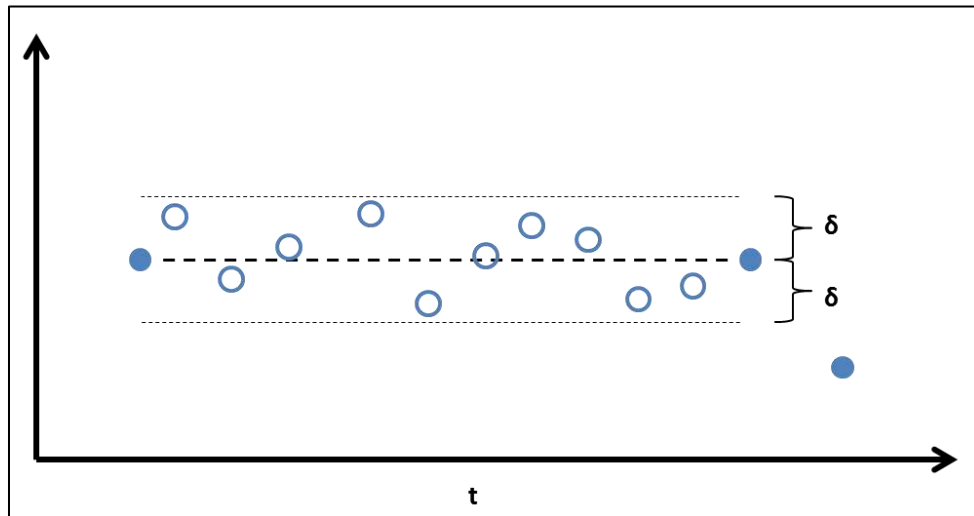
Entre diversas ferramentas disponíveis para o usuário do PI, duas se mostraram importantes no desenvolvimento deste trabalho. A interface do usuário com o sistema de coleta de dados ocorre após a aplicação automática de dois filtros. Os dados enviados pelos sensores chegam ao servidor do PI e são testados segundo dois critérios. Se o dado for aprovado em pelo menos um dos dois critérios, ele é descartado por não possuir grande relevância e, com isso, poupar memória do servidor. Somente se o dado recém recebido não passar nesses critérios que ele é registrado e fica salvo no servidor.

Esses critérios são a compressão por variação absoluta e a compressão por variação de tendência.

#### A. Compressão por variação absoluta

A primeira regra visa a descartar dados que não mostram alteração significativa em relação ao último dado registrado. Especifica-se uma variação  $\delta$  que o servidor irá usar para aplicar o filtro. Se um

dado novo estiver dentro do intervalo  $\pm\delta$  do último dado, descarta-se o novo e utiliza-se o último dado salvo como dado atual. Esse descarte pode ser feito por até 8 horas. Isso significa que se um sensor transmite valores basicamente idênticos durante 1 dia, o servidor do PI irá mostrar apenas 3 pontos iguais com linhas de interpolação entre eles. A Figura 4-1 representa a regra descrita (28).

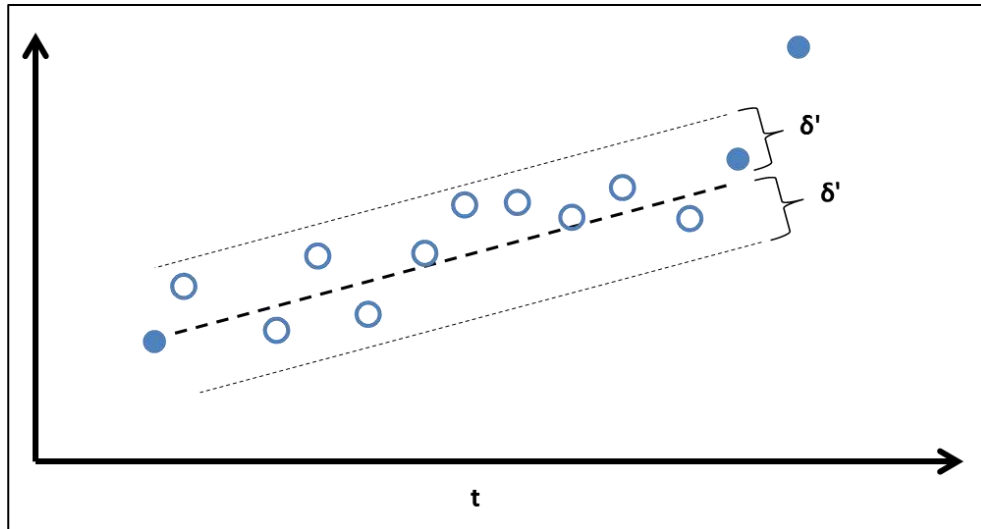


*Figura 4-1: Resumo da regra de compressão por variação absoluta. Neste exemplo, apenas os pontos preenchidos são salvos.*

#### B. Compressão por variação de tendência

Esta segunda regra é semelhante à primeira, porém amplia essa compressão de dados à verificação de tendências lineares. Ao se estipular que uma sequência de dados está seguindo uma tendência linear com variação de  $\pm\delta'$  (também pré-definida no servidor), o servidor guarda apenas o primeiro e último ponto. Em seguida, promove a interpolação desses pontos, otimizando o uso do espaço disponível. A Figura 4-2 representa graficamente a regra descrita neste parágrafo.





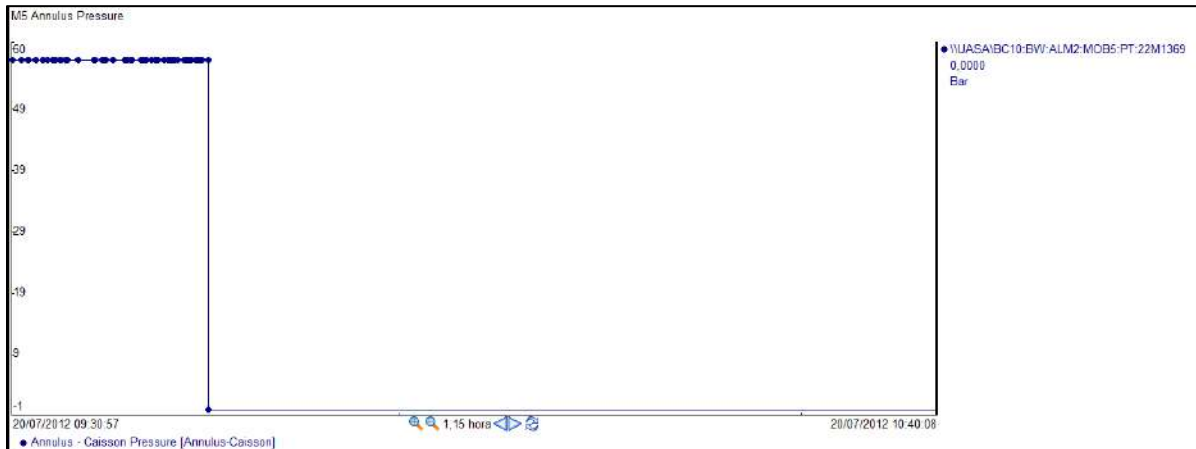
*Figura 4-2: Resumo da regra de compressão por variação de tendência.*

Finalmente, se um dado recém-chegado não se enquadra em nenhum desses dois filtros, ele é armazenado.

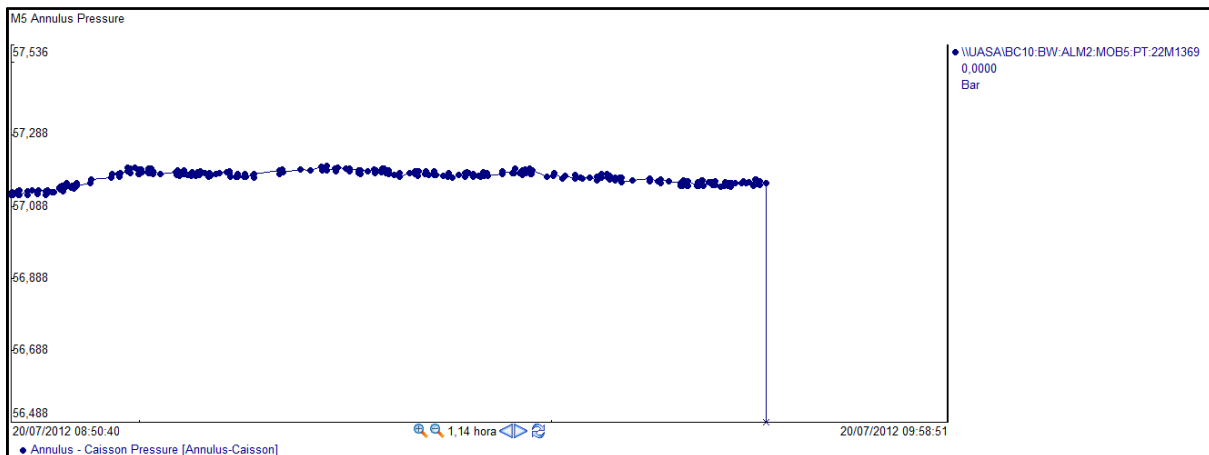
A importância de saber como funciona esse armazenamento e descarte de informação se deve ao fato de que isso influencia o comportamento aparente de um sensor. Muitas vezes um sensor transmite durante períodos longos um sinal constante, que é representado por uma linha horizontal reta, sem qualquer ruído. Na maioria das vezes isso é indicação de que o sensor não está funcionando, já que ruídos são esperados. Porém, é possível que aquele sensor esteja sofrendo uma filtragem por parte do servidor que não permite que o usuário veja suas pequenas variações. Os valores de  $\delta$  ou  $\delta'$  imputados descartam todos os dados que chegam. Nesse caso, trata-se de um sensor sadio que pode ser erroneamente interpretado como ruim. Seus parâmetros de compressão devem ser modificados.

Uma maneira de saber com mais certeza se o sensor está funcionando é obrigar o software a mostrar ao usuário todos os dados registrados pelo servidor. Ao se usar essa ferramenta, o PI explicita no gráfico, através de marcadores circulares, o que é dado real, salvo no servidor (Figura 4-3 e Figura 4-4). Assim, ao invés de ver apenas uma linha, o usuário vê também círculos indicando todos os pontos coletados e registrados. As únicas situações que justificariam a ausência de marcadores seriam a compressão dos dados pelos filtros citados ou a perda daquele sinal, possivelmente devido a uma falha.

É importante destacar que os dados descartados pelas regras de compressão não são salvos no servidor, então não irão aparecer. O uso dessa ferramenta torna o programa um pouco mais lento, pois amplia a quantidade de informação na tela, por isso não é uma configuração padrão.



*Figura 4-3: Sensor de pressão no momento da falha com a ferramenta citada ativa.*



*Figura 4-4: Zoom do sensor da Figura 4.3 no momento da falha com a ferramenta citada ativa.*

### 4.3 Coleta dos dados

A pesquisa por dados de sensores com falha para este estudo foi facilitada devido à existência de documentos internos da Shell Brasil com a relação de todos os sensores que apresentam ou apresentaram mau funcionamento. Basta procurar no sistema a data da falha e exportar os dados de todos os sensores relevantes para o Excel. Essa exportação é feita através de uma extensão do PI chamada PI Datalink, desenvolvida especialmente para esse fim.

No Excel são selecionados o período total desejado, o intervalo entre cada ponto (no caso foi utilizado 1 minuto (9)), e as *tags* de todos os sensores escolhidos. Em seguida, os dados obtidos são convertidos para um formato numérico aceito pelo Matlab (troca de vírgulas por pontos) para, então, neste último software, as matrizes de dados experimentais serem criadas e analisadas.

Utilizou-se o Matlab para implementar a metodologia discutida no próximo capítulo. O código foi elaborado de forma genérica para conseguir manipular automaticamente qualquer matriz de dados, independentemente de suas dimensões (Anexo A). Não foram utilizadas as ferramentas de PCA nativas do pacote do Matlab.

## 5 Metodologia

### 5.1 Estratégia de análise

Existem diferentes maneiras de utilizar um modelo estatístico para detecção de falhas. O modelo pode ser usado para comparar as tendências dos dados analisados com tendências conhecidas (previamente imputadas) ou ainda para avaliar mudanças de comportamento entre os dados novos e os dados recentes.

Uma opção então consiste em criar uma base de dados com leituras de sensores sadios durante os diferentes tipos de regime de operação normal que ocorrem durante um processo. Essa base de dados é criada a partir de dados antigos, quando se sabia exatamente o que estava acontecendo na planta. Os novos dados são então comparados com essa base de dados para se classificar o regime atual. Um erro nessa classificação – ou seja, os dados atuais não se assemelham a nenhum regime conhecido – indica uma anormalidade na operação (24) (25).

Outra abordagem é comparar os dados atuais com dados recentes, onde o intervalo entre os dois conjuntos de dados pode ser tanto fixo quanto móvel. Isso significa especificar o momento atual como “Fase B” e o que o antecede como “Fase A”. Uma disparidade entre esses dois grupos indica uma possível anormalidade ou mudança de regime (20). Se não houve alteração no processo, a conclusão que se tem é que os novos dados contêm erros. Logo, é provável que algum sensor tenha falhado (25) (29). Como mudanças de regime costumam ocorrer após ações por parte dos operadores, eles saberão ignorar ou inibir possíveis alarmes do método enquanto estiverem atuando em válvulas ou modificando condições do processo.

Em resumo, o modelo pode ser treinado a partir de dados conhecidos, antigos, para reconhecer fases de operação, ou pode ser treinado a partir dos dados que antecedem imediatamente o conjunto que está sendo analisado.

Antes de se discutir vantagens e desvantagens de cada abordagem, vale notar aqui três pontos importantes e relevantes para este trabalho:

1. A evolução das condições naturais de um reservatório de petróleo e gás implica que o processo de escoar esse fluido até a plataforma esteja em constante mudança. Queda

da pressão e aumento da fração de água com o tempo são os dois principais efeitos que limitam a vida útil de um reservatório. A escala de tempo para tais transições é bastante grande. Entretanto, essas características de processo tornem a padronização de regimes de operação seja muito complicada. Apesar do uso de válvulas-agulha, bombas submarinas ou elevação a gás para manter o escoamento estável, comparar dados distantes temporalmente dificilmente trará bons resultados.

2. Como consequência das mudanças nos fluidos do reservatório, no suporte de pressão dos poços ou no comportamento das bombas, os responsáveis pela planta avaliam semanalmente a possibilidade de realizar mudanças de processo que possam trazer benefícios à operação. Essa busca constante por otimização da produção leva a alterações frequentes no processo.
3. Os regimes de operação que podem ser analisados pelo método estatístico em uma planta para análise de falhas são estado estacionário e aumento/redução controlados da produção. Qualquer cenário que inclua paradas de emergência, desligamento de bombas, mudança no trajeto do escoamento, ou seja, qualquer situação em que válvulas se fechem ou se abram bruscamente, geram variações demasiado repentinas para serem padronizadas (1). Sendo assim, nestas situações esse método irá gerar falsos positivos, indicando falha de sensor quando o mesmo estiver funcionando normalmente. Durante um aumento controlado da produção (ou, similarmente, uma redução controlada), os operadores estão com toda a atenção voltada para os sensores relacionados à operação. Não é tão importante, portanto, que nesse momento o modelo PCA busque falhas nesses sensores. Caso haja alguma, ela seria notada pelos próprios operadores.

Com isso em mente, deve-se refletir sobre o melhor uso da metodologia PCA para evitar a criação de um método que gere mais alarmes falsos do que verdadeiros. No caso presente, concluiu-se que seria inviável criar um banco de dados com todos os regimes de operação possíveis, principalmente porque novos regimes de operação são frequentemente propostos e colocados em prática. Decidiu-se

então usar a segunda abordagem, na qual o modelo avalia os dados atuais comparando-os com dados recentes.

Foi decidido também avaliar apenas períodos de produção em estado estacionário. Como essa suposição não é embasada em resultados da literatura, mas sim na interpretação do método e de suas limitações, se decidiu testa-la. Assim, dados de regimes instáveis de escoamento, porém sem falha de sensores, foram analisados pela mesma metodologia.

## 5.2 Quantificadores ou Indicadores

Segue descrição dos quantificadores utilizados.

### 5.2.1 Q, T<sup>2</sup> de Hotelling e Erro Quadrático de Predição (SPE)

Para verificar a aderência do modelo aos dados, basta comparar os dados modelados aos dados originais. Essa comparação pode ser feita de forma quantitativa através de três abordagens: Q, T<sup>2</sup> de Hotelling e SPE. Os três métodos são bastante semelhantes e se baseiam no desvio definido na Equação 3. Trata-se da diferença entre o dado experimental e o dado modelado pelo modelo PCA obtido.

$$e_k = (I - C)x_{expk} \quad \text{Equação 3}$$

onde,  $e_k$  é o erro associado ao modelo,  $I$  é a matriz identidade de dimensão apropriada,  $C$  é a matriz quadrada do modelo PCA e  $x_{expk}$  é a matriz de dados experimentais após pré-tratamento.

Q e T<sup>2</sup> de Hotelling são muito semelhantes ao SPE, porém mostram resultados menos promissores na literatura. Deste modo, o SPE foi escolhido para ser avaliado no presente trabalho. Ele é capaz de levar em consideração os valores de SPE calculados para os intervalos passados, de forma a ponderar na medida mais recente os erros já observados nas medidas passadas (24) (9). Seu cálculo ocorre a partir das equações 4 e 5:

$$e_{sqrk} = e_k^2 \quad \text{Equação 4}$$

$$SPE_k = SPE_{k-1}(1 - \lambda) + e_{sqrk}(\lambda) \quad \text{Equação 5}$$

onde  $e_{sqrk}$  é o erro quadrático e tende para  $SPE_k$  quando  $\lambda$  tende para 1.

O parâmetro  $\lambda$  determina o grau de influência do passado na leitura atual. Um valor próximo de 1 minimiza o efeito memória do índice. Diferentes valores de  $\lambda$  foram avaliados para se entender o efeito dos dados passados. Espera-se que maiores efeitos memória façam com que o índice sofra uma espécie de “efeito tampão”, diminuindo sua sensibilidade a falhas instantâneas, passageiras e de baixa amplitude. Ao mesmo tempo porém, a quantidade de falsos positivos tenderia a diminuir.

O SPE é avaliado da seguinte forma: estipula-se um valor máximo de desvio a partir da média desse índice; toda vez que o valor de SPE ultrapassar esse limite, considera-se que o processo falhou. Esse limite de operação normal geralmente é relacionado ao desvio padrão dos dados experimentais (9). Entretanto, nos diferentes casos analisados neste trabalho a ordem de grandeza do desvio padrão variou muito. Além disso, a própria amplitude do índice SPE não apresentou comportamento regular entre os casos. Sendo assim, decidiu-se por uma abordagem alternativa: o máximo valor de SPE durante a Fase A é aquele que limita o índice. A Figura 5-1 mostra o índice SPE durante as duas fases e o limite (em vermelho) sendo quebrado. A linha verde apenas indica o dobro do limite máximo para se visualizar melhor a disparidade dos novos valores em relação aos antigos. A falha é então detectada, conforme previsto, no início da Fase B.

Devido à possibilidade de ruídos nos valores de SPE durante a Fase A, decidiu-se implementar um pequeno filtro: utilizou-se o segundo maior valor de SPE na Fase A como limite, ao invés do primeiro. Em alguns casos observou-se que um pico logo no início dessa Fase tornava o método muito pouco sensível pois o mesmo utilizava um limite muito elevado. Usualmente o primeiro ponto da série, na Fase A, apresentava um valor mais elevado.

Os valores de SPE durante a Fase B são finalmente avaliados em relação a esse limite: quando ultrapassarem-no, são classificados como falha. Em seguida a quantidade de falhas detectadas é ponderada em relação ao total de pontos.

A quantidade de pontos em falha é avaliada em relação ao tamanho total do vetor  $SPE_k$  para que picos de falha instantâneos sejam descartados do método. Por exemplo, se o vetor  $SPE_k$  possui 1000 pontos e apenas 2 deles ultrapassaram o limite, considera-se que não foi detectada falha e que esses pontos são ruídos.

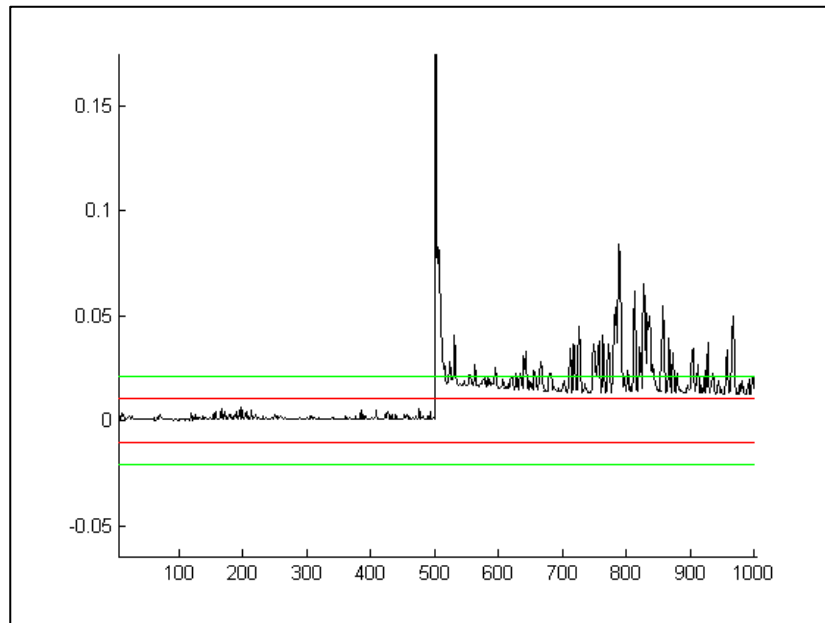


Figura 5-1: Exemplo gráfico do SPE, em preto.

### 5.2.2 Fator de similaridade de angulação – Spca

Outra forma de detectar falhas usando PCA é através dos indicadores de similaridade. Estes não comparam os dados modelados com os originais. Ao invés disso, eles comparam dois modelos PCA distintos obtidos a partir de conjuntos de dados diferentes. O mais conhecido é o *Similarity PCA* (Spca) (29).

Escolhem-se dois conjuntos de dados distintos, A e B, de mesmas dimensões. Aplica-se a cada um deles a metodologia para chegar às matrizes PCA (MC), uma para A e outra para B. O Spca calcula a diferença de angulação de cada vetor de cada matriz PCA. Matematicamente, pode-se calcular tal diferença através da Equação 6.

$$Spca = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k \cos^2 \theta_{ij} \quad \text{Equação 6}$$



onde  $k$  é o número de componentes principais escolhido e  $\theta_{ij}$  o ângulo entre esses vetores par a par.

Ou, na forma matricial, conforme mostrado na Equação 7:

$$Spca = \frac{\text{trace}(MC_A^T MC_B MC_B^T MC_A)}{k} \quad \text{Equação 7}$$

onde novamente  $k$  é o número de componentes principais, e  $MC$  são as matrizes do modelo PCA. A função *trace* calcula a soma dos componentes da diagonal principal.

Se o conjunto A for semelhante ao conjunto B, as orientações dos vetores das matrizes dos modelos PCA gerados –  $MC_A$  e  $MC_B$  – também serão semelhantes. Assim, o  $Spca$  terá um valor próximo da unidade. Caso contrário, o valor de  $Spca$  será mais próximo de zero.

### 5.2.3 Fator de similaridade de distância – Sdist

Da mesma forma que se pode analisar a diferença entre os ângulos, foi criado um indicador que calcula a distância entre os centros de cada uma dessas matrizes (24). O  $Sdist$  usa a distância de Mahalanobis ( $\phi$ ). Numericamente, este indicador se aproxima de 1 quando os dois conjuntos são semelhantes e tende a zero quando eles se afastam. A Equação 8 mostra como esse índice é calculado:

$$Sdist = 2 \left[ 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\phi} e^{-\frac{z^2}{2}} dz \right] \quad \text{Equação 8}$$

onde  $\phi$  é a “distância de Mahalanobis”, calculada pela Equação 9.

$$\phi = \sqrt{(\bar{x}_A - \bar{x}_B) \sum_s^{*-1} (\bar{x}_A - \bar{x}_B)^T} \quad \text{Equação 9}$$

sendo  $\Sigma$  a pseudo-inversa da matriz de covariância dos dados da Fase B.

#### 5.2.4 Fator de Similaridade - SF

Finalmente, existe um indicador que alia os dois indicadores de similaridade citados (24). O *similarity factor*, SF, ou fator de similaridade, é uma média ponderada do Spca com o Sdist (Equação 10).

$$SF = \alpha Spca + (1 - \alpha) Sdist \quad \text{Equação 10}$$

O valor de  $\alpha$  pode variar entre 0 e 1. Usualmente,  $\alpha = 2/3$  apresenta bons resultados (24) (20).

#### 5.2.5 Índice de Validade de Sensor - SVI

Existe ainda um indicador de validade de sensor, SVI, que é usado para identificar qual sensor apresentou falha. O SVI é capaz de apontar ao usuário qual o sensor que deve ser trocado, ou ao menos ignorado. Este indicador parte do princípio que dois sensores não podem falhar concomitantemente (14). Essa é uma hipótese improvável, não sendo sempre verdadeira. Os sensores estudados podem ser instalados em blocos; ou muitas vezes compartilham um mesmo módulo de controle submarino para transmitir seus dados; ou ainda possuem alimentação elétrica por vias comuns. Uma falha qualquer que afete um desses equipamentos compartilhados por mais de um sensor gera imediatamente a “falha” de múltiplos sensores. Alguns exemplos de tais falhas são rompimentos de cabos elétricos, falhas nos módulos de controle submarinos, flutuações no fornecimento de energia ou falhas mecânicas nos blocos contendo mais de um sensor. Ainda que o dispositivo do sensor em si não tenha parado de funcionar, o que aparece na interface para o usuário é uma falha múltipla e simultânea. Vale ressaltar que isto foi observado no conjunto de dados das plantas estudadas, sendo relativamente comum de ocorrer.

Como consequência, a única maneira de empregar tal índice no presente caso seria através de uma pré-seleção de dados. Ou seja, seria necessário garantir que os dados analisados não apresentaram falhas múltiplas nos períodos estudados. Isto não seria razoável para a finalidade deste estudo.

### **5.3 Análise de Sensibilidade**

Devido à grande incerteza relacionada ao estudo proposto, onde dados reais brutos são utilizados, foram feitas análises de sensibilidade com diversos parâmetros da metodologia. O objetivo com essa análise foi buscar o melhor resultado para os casos estudados.

#### **5.3.1 Grau de representatividade dos componentes principais**

O grau de representatividade total acumulada dos componentes principais não é unanimidade na literatura (2) (30) (24) (23). Por isso, foi escolhido trabalhar com 3 valores diferentes: 90%, 95% e 99%. Quanto maior essa porcentagem, menor a perda de informação em relação aos dados originais, porém maior a quantidade de componentes principais empregados.

#### **5.3.2 SPE e o efeito memória - $\lambda$**

O parâmetro  $\lambda$ , que afeta o efeito memória do indicador SPE, também foi analisado com 3 valores distintos: 0.5, 0.75 e 1.0.

Em princípio, as falhas de sensores nas plantas estudadas costumam ser de grande magnitude. Além disso, os dados estudados possuem normalmente ruídos elevados, especialmente quando regimes de escoamento pouco estáveis se estabelecem. Com isso, espera-se que o aumento do efeito memória – redução do valor de  $\lambda$  – traga benefícios ao método: pequenos e rápidos desvios da média (erros pontuais grandes) serão minimizados quando os dados precedentes, estáveis, tiverem um maior peso sobre o quantificador.

#### **5.3.3 Composição do SF**

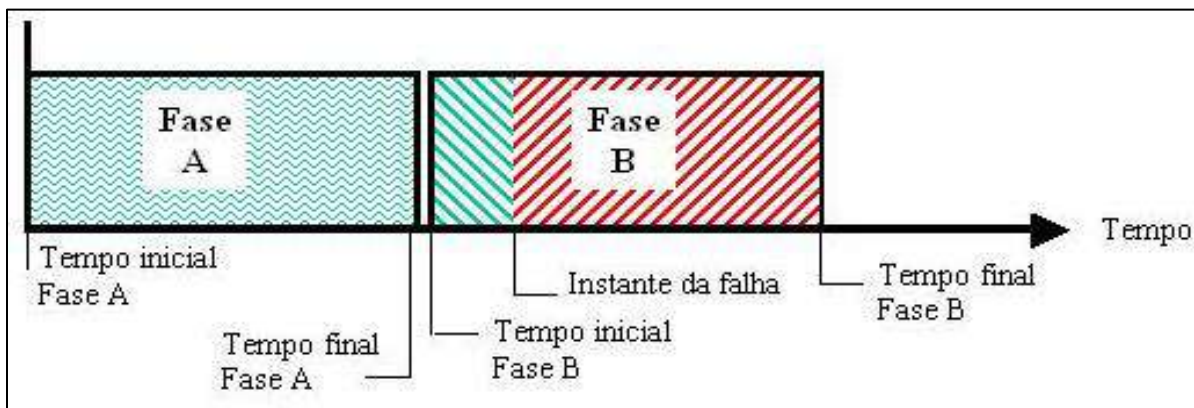
O fator  $\alpha$ , que distribui o peso do Spca e do Sdist sobre o valor final do SF também foi submetido à análise: foram avaliados  $\alpha = 1/2$  e  $\alpha = 2/3$ .

### 5.3.4 Instante da falha

Sabe-se que existe uma influência do instante da falha, conforme Figura 5-2, sobre o resultado final do método de detecção.

O peso do período com falha sobre a extensão da Fase B pode alterar a resposta do modelo. Por esse motivo, decidiu-se aplicar dois períodos diferentes, cada um com um instante relativo da falha diferente. O instante da falha foi fixado no minuto 501 e a duração total da análise variou: avaliaram-se períodos de 1000 e 600 minutos de duração (esquema na Figura 5-3). No primeiro caso, a falha ocorre logo no início da Fase B. No segundo caso, a falha ocorre apenas no final da Fase B.

A Fase A corresponde sempre à primeira metade do período total avaliado e a Fase B, à segunda metade. Desta forma, a Fase A e a Fase B possuem sempre durações iguais (500 minutos cada ou 300 minutos cada, dependendo da duração total). Porém, o período com falha (dentro da Fase B) pode tornar-se mais ou menos significativo dentro do intervalo total. Na Figura 5-3, do item I para o item II verifica-se que o peso do período com falha sobre o período total diminui de 50% para 16,7%.



*Figura 5-2: Representação esquemática da posição da falha em relação ao período analisado.*

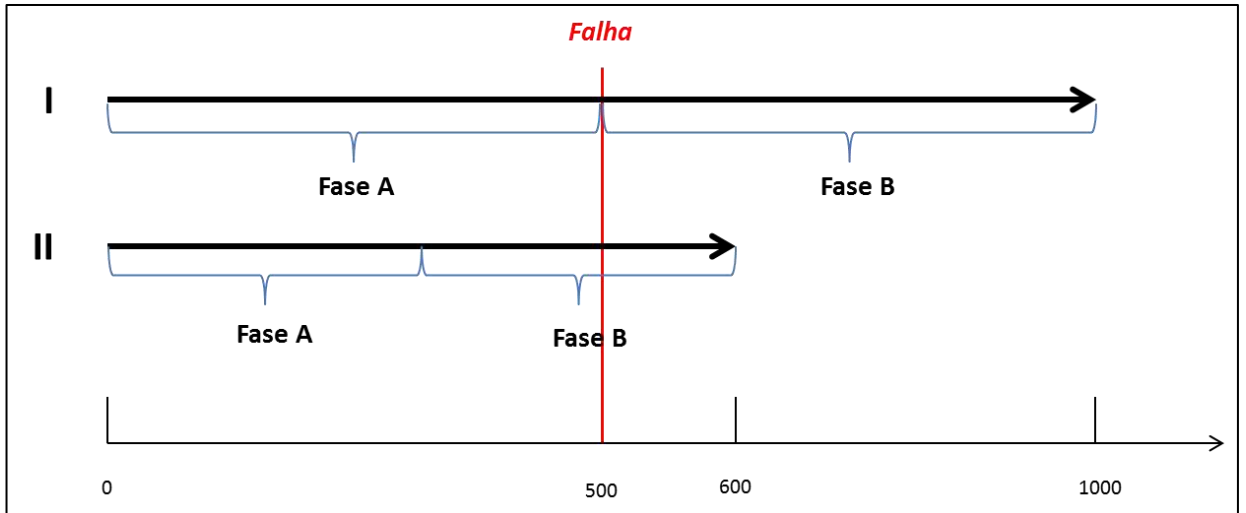


Figura 5-3: Representação esquemática da análise feita para comparar diferentes durações do período com falha.

Em suma, cada caso avaliado apresentará diferentes resultados, todos resumidos em duas tabelas, uma para a duração de 1000 minutos e outra idêntica para resultados de 600 minutos. A Figura 5-4 exemplifica o modelo de tabela de resultados parciais. Devido à grande quantidade de resultados gerados, não foram apresentadas tais tabelas com os resultados brutos, mas sim as estatísticas provenientes da avaliação do desempenho das abordagens propostas.

% Mínima acumulada	% Representada pelos PCs	k Quantidade de PCs	SPE			Spca	Sdist	SF	
			$\lambda = 0.5$	$\lambda = 0.75$	$\lambda = 1.0$			$\alpha = 1/2$	$\alpha = 2/3$
90%									
95%									
99%									

Figura 5-4: Tabela-modelo com resumo dos resultados obtidos cruzando todas as análises de sensibilidade. Os dados com fundo cinza são aqueles obtidos a partir de períodos estáveis para comparação direta.

## 5.4 Aplicação do método

Todos os casos avaliados foram submetidos a um mesmo critério de seleção e a uma mesma sequência de testes. Tais escolhas são explicadas a seguir:

- A. O período da falha foi separado em duas fases de igual tamanho, A e B, sendo a Fase A imediatamente anterior à falha e a Fase B aquela que contém a falha.
- B. Além do intervalo da análise, foi também realizado um teste “branco”, analisando um período de igual duração, mas sem incidentes, quando todos os sensores estavam funcionando. Este período “branco” é anterior à falha do sensor, e quando o processo encontrava-se estável.

Os resultados dos testes “brancos” permitem a análise dos quantificadores de forma mais clara. Não se conhece *a priori* qual o valor limite de cada indicador que determina se foi ou não detectada falha nos dados analisados. Assim, foram utilizados os resultados dos períodos “brancos” para fins de comparação.

- C. Todos os sensores envolvidos possuem uma relação física que os conecta.

Todos os sensores submarinos que possuam uma interação física conhecida com o sensor que falhou foram utilizados. Os únicos sensores excluídos dessas análises foram aqueles que geram dados “ruins” para o método, como sensores de posição de válvula ou qualquer sensor com sinal em degraus.

- D. No momento da falha o processo se encontrava em estado estacionário. Quando uma estabilidade perfeita não fosse possível garantia-se que, pelo menos dentro do período analisado, não houvesse nenhuma grande variação.

Para complementar, foram escolhidos dois casos em um período transiente. O objetivo disto foi avaliar o comportamento do método nessas condições.

## 5.5 Casos Estudados

Existem alguns tipos mais comuns de falhas observadas nos sensores estudados. Na maioria das vezes o valor medido pelo sensor no momento da falha passa a ser zero e não muda mais. Algumas vezes o zero é substituído pelo valor máximo ou mínimo daquele sensor (topo ou fundo da escala).

Outras formas de falha observadas foram oscilações muito grandes e rápidas nos valores mostrados e/ou a estagnação da leitura em um valor fixo qualquer. Não houve nenhum caso em que um sensor passasse a mostrar valores ligeiramente acima ou abaixo do valor real, sem outra mudança no comportamento do sinal enviado.

Buscou-se avaliar esses diferentes tipos de falha nos casos estudados.

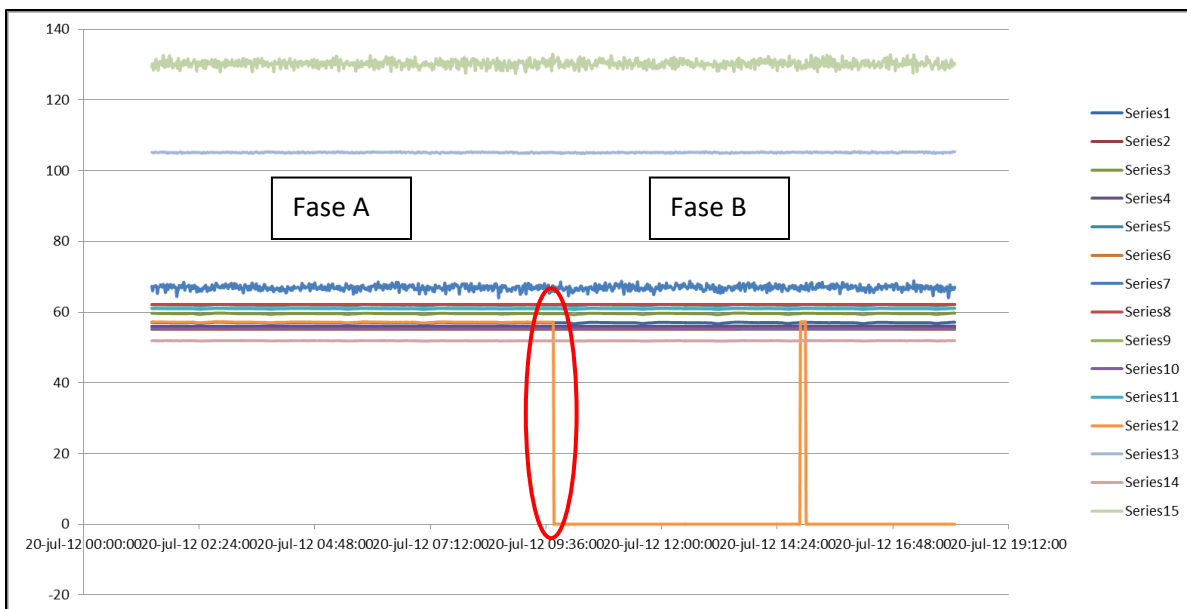
São apresentadas a seguir as séries temporais nos intervalos analisados com duração de 1000 minutos, sendo 500 minutos da Fase A e 500 minutos da Fase B. A falha ocorre sempre no minuto 501. As análises com 600 minutos totais de duração foram feitas sobre os mesmos dados: foi feito um corte na parte final da série temporal.

### 5.5.1 Caso 1

Nesse primeiro caso o sensor de pressão falha e passa a informar 0.0 (zero). Trata-se de uma falha instantânea e total (ocorre de um instante para o outro e não volta a funcionar mais) ocorrida durante um período estável de produção. Esse sensor especificamente não é essencial para a produção, mas muito importante para o monitoramento da integridade do revestimento da *caisson*. Ele é utilizado para garantir que não há vazamento de dentro do *caisson* para o exterior. Trata-se de um sensor que é observado periodicamente para se executar testes de integridade.

Neste caso foram analisados 15 sensores concomitantemente. Esses sensores estão simbolicamente representados na Figura 5-5. São sensores de vazão, pressão, temperatura, frequência e corrente. O instante inicial da falha encontra-se marcado na figura.

Note na Figura 5-5 que o sensor ruim (laranja) volta a mostrar o valor correto por um instante alguns minutos após a falha, e em seguida retorna ao zero indefinidamente.



*Figura 5-5: Sensores analisados no Caso 1.*

### 5.5.2 Caso 2

O caso 2 apresenta uma falha quase simultânea em dois sensores que se localizam na árvore de natal de um poço de injeção de água. Ambos os sensores, um de pressão e outro de temperatura, pertencem a um mesmo dispositivo, estando localizados no mesmo ponto da tubulação. Ambos passam a oscilar de forma repentina entre zero e o valor original que estavam lendo antes da falha.

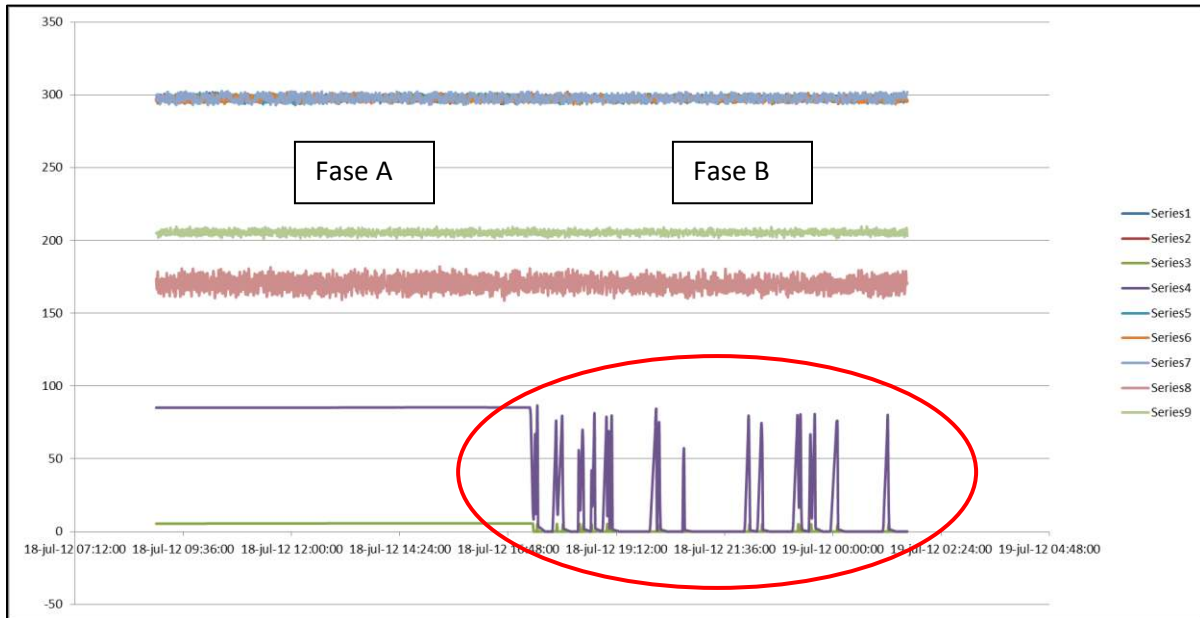
A falha dupla ocorre em etapas separadas por 2 semanas uma da outra. Num primeiro momento os sensores falham, porém voltam a funcionar dias depois. Em seguida, eles falham novamente e em definitivo. Ambos os eventos foram estudados.

Foram feitas análises com 9 sensores – onde 2 falham – e com 8 sensores, com apenas um falhando (Figura 5-6). Foi decidido fazê-lo dessa forma para garantir um melhor parâmetro de comparação entre este caso e os demais. Assim, aproveitou-se para avaliar como o método se comporta diante de uma falha dupla, além de avaliar a aderência do modelo as falhas oscilatórias.

Vale ressaltar que o número total reduzido de sensores neste caso em relação ao anterior se deve a dois motivos:



- 1- Por se tratar de um poço de injeção de água, há menos sensores. Esses poços têm menos parâmetros a serem medidos devido à maior simplicidade do seu fluxo monofásico. Poços produtores são mais complexos, por isso têm mais sensores.
- 2- Sensores de posição de válvulas ou sensores que variam em degraus não foram incluídos nesta análise.



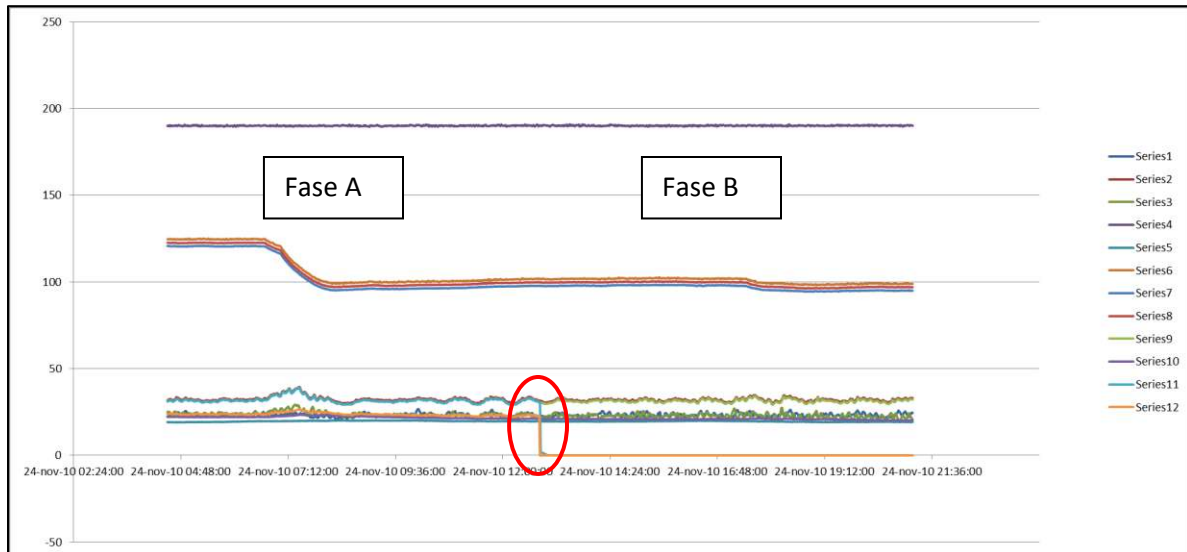
*Figura 5-6: Sensores analisados no Caso 2. A Figura mostra o instante da primeira falha.*

### 5.5.3 Caso 3

Dois sensores, um de temperatura e um de pressão, ambos localizados na árvore de natal de um poço produtor, falham e passam a apresentar o valores fixos de 0.

Análogo ao caso anterior, este apresenta uma falha simultânea de dois sensores pertencentes a um mesmo dispositivo. Porém neste caso a falha é total e não mostra recuperação ou oscilação após o primeiro evento (Figura 5-7).

Novamente foram analisadas tanto a falha dupla quanto a falha de apenas um dos sensores separadamente. Foi utilizado um conjunto de 12 sensores – ou 11 na análise com apenas uma falha – para a análise.

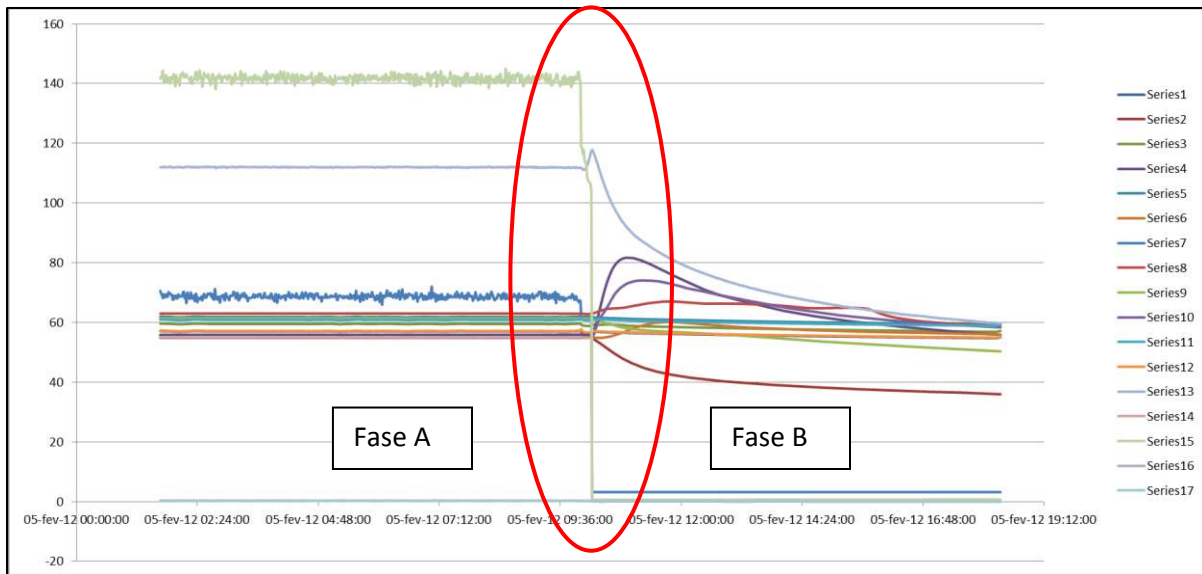


*Figura 5-7: Caso 3: dois sensores perdem suas leituras e passam a indicar zero simultaneamente.*

#### 5.5.4 Caso 4

No caso 4 utilizou-se um período transiente do processo. Foi selecionado um momento em que uma BCS parou de funcionar de forma inesperada, interrompendo repentinamente o fluxo dos poços. Nesse instante os sensores associados a essa bomba sofrem alterações rápidas e bruscas. O objetivo da análise de uma situação como esta é verificar como o PCA se comporta frente a uma mudança na operação.

Foram avaliados 17 sensores, análogos aos utilizados no Caso 1. A ilustração do momento da parada pode ser vista na Figura 5-8.

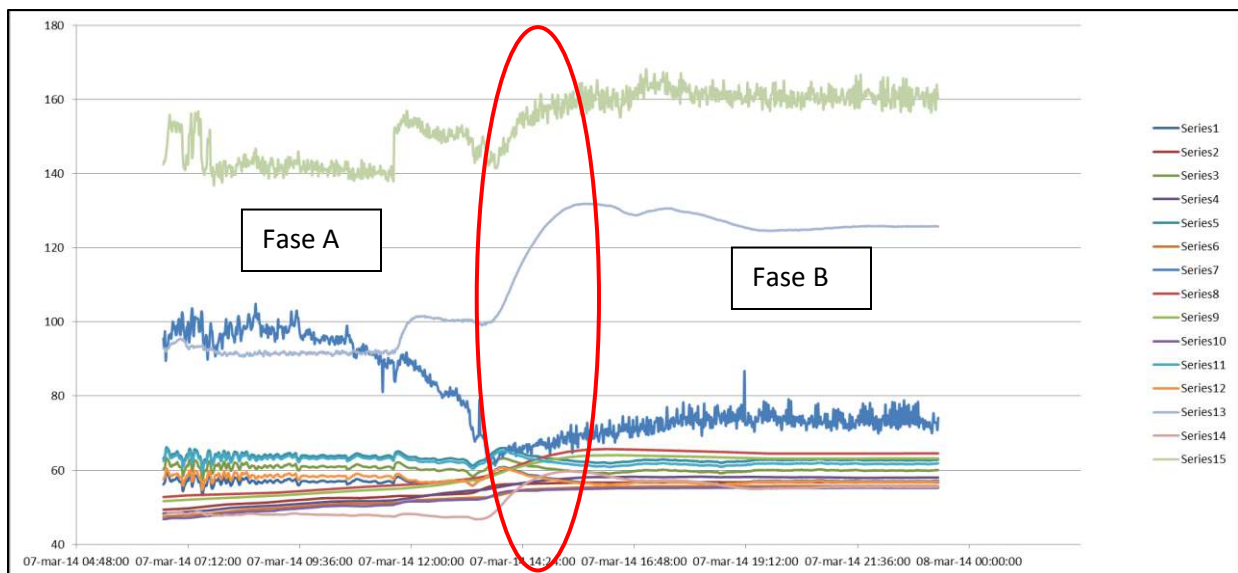


*Figura 5-8: Caso 4: parada repentina de uma BCS devido a problemas na operação.*

### 5.5.5 Caso 5

Neste último caso foi decidido analisar um aumento gradual da vazão de uma das bombas BCS da planta. Escolheu-se um aumento gradual e lento. Esses dados, representados na Figura 5-9, foram comparados com um período estável de produção, alguns dias antes. Quinze sensores equivalentes aos utilizados no Caso 1 foram observados (sensores com mesma função, porém em uma unidade diferente).

Durante o aumento de vazão parâmetros como temperatura, pressão de descarga, corrente e frequência do motor são elevados. Já outros parâmetros apresentam quedas, como a pressão dentro do *caisson* e o nível de líquido no mesmo. Note também que alguns sensores mostram diferença na frequência dos ruídos antes e após a mudança, possivelmente devido à mudança de regime de escoamento ou alteração das condições de equilíbrio dentro do *caisson*.



*Figura 5-9: Aumento gradual da produção da BCS avaliada no Caso 5.*

A Tabela 5-1 traz o conjunto de casos analisados, assim como suas características principais e as diferentes análises que foram realizadas.

*Tabela 5-1: Resumo dos casos analisados.*

Caso	Tipo de falha	Sensores analisados	Sensores ruins	Descrição
<b>Caso 1</b>	Instantânea e total	17	1	Sensor de pressão passa a marcar 0 bar
<b>A</b>	Oscilatória	8	1	Sensor de temperatura passa a oscilar (primeiro evento de falha seguido de recuperação alguns dias depois)
<b>Caso 2</b>	<b>B</b> Oscilatória	8	1	Sensor de temperatura passa a oscilar (segundo evento de falha, definitivo)
<b>C</b>	Oscilatória	9	2	Dois sensores, de pressão e temperatura, passam a oscilar (primeiro evento de falha seguido de recuperação alguns dias depois)
<b>Caso 3</b>	<b>A</b> Instantânea e total	11	1	Sensor de pressão passa a marcar 0 bar
<b>B</b>	Instantânea e total	12	2	Dois sensores, de pressão e temperatura, passam a marcar 0
<b>Caso 4</b>	Não houve falha	17	0	Parada repentina na operação de uma BCS
<b>Caso 5</b>	Não houve falha	15	0	Aumento gradual da vazão de uma BCS

## 6 Resultados

Os resultados de todas as análises executadas são apresentados neste capítulo. Conforme detalhado no capítulo anterior, foram 5 casos estudados gerando um total de 8 análises diferentes (Tabela 5-1). O caso 1 teve uma análise apenas. O caso 2 gerou 3 análises: 2A, 2B e 2C. O caso 3 gerou duas análises, 3A e 3B, e os casos 4 e 5 geraram apenas uma análise cada um. Os casos 1, 2 e 3 apresentaram falha de sensor. Os quantificadores foram então avaliados segundo sua capacidade em detectar a falha ocorrida. Já os casos 4 e 5 não apresentaram falha de sensor, mas sim uma mudança no processo, no caso 4, uma mudança brusca, no caso 5, uma mudança controlada. Os quantificadores foram então avaliados de acordo com a quantidade de falsos positivos.

Note que foram feitas duas avaliações distintas:

- Capacidade de detectar a falha ocorrida nos casos 1, 2 e 3 (total de 6 análises);
- Capacidade de lidar com mudança do processo sem gerar falso positivo nos casos 4 e 5 (total de 2 análises).

Visando a análise de sensibilidade mencionada na seção 5.3, para cada análise descrita acima foram avaliados 42 resultados. Estes incluem a avaliação dos seguintes parâmetros:

- Quatro indicadores que se desdobraram em sete:
  - SPE com três valores de  $\lambda$  diferentes (0.5, 0.75 e 1.0);
  - Spca;
  - Sdist;
  - SF com dois valores de  $\alpha$  diferentes ( $1/2$  e  $2/3$ ).
- Dois intervalos de diferentes durações das fases (600 minutos e 1000 minutos);
- Três limites mínimos de representatividade por parte dos PCs (90%, 95% e 99%);

Obteve-se assim um total de 336 resultados (42 x 8). Destes, 252 se referem a análises de períodos com falha real de sensor e 84 se referem a análises de mudanças de processo.

Os quantificadores foram testados também nos intervalos de tempo onde não havia falha, os períodos “brancos” (Fase A), anteriores a Fase B. Os resultados das análises das falhas foram comparados diretamente com os resultados desses “brancos”. Definiu-se então que a falha teria sido detectada pelo indicador quando houvesse uma variação superior a 10% entre os dois valores. No caso do Spca, Sdist e SF, uma *diminuição* de 10% em relação ao valor do período “branco”, sem falha, indica detecção de falha. No caso do SPE, um *aumento* de 10% indicaria o mesmo.

Finalmente, os resultados de todas as análises foram compilados e as estatísticas de cada quantificador e de cada caso foram avaliadas.

## 6.1 Resultados dos quantificadores

### 6.1.1 SPE

O índice de erro quadrático de predição foi o que apresentou os melhores resultados. Até 92% das falhas reais foram corretamente identificadas (Figura 6-1). O valor de  $\lambda = 1.0$ , ou seja, ausência do efeito memória, foi aquele que gerou melhor desempenho.

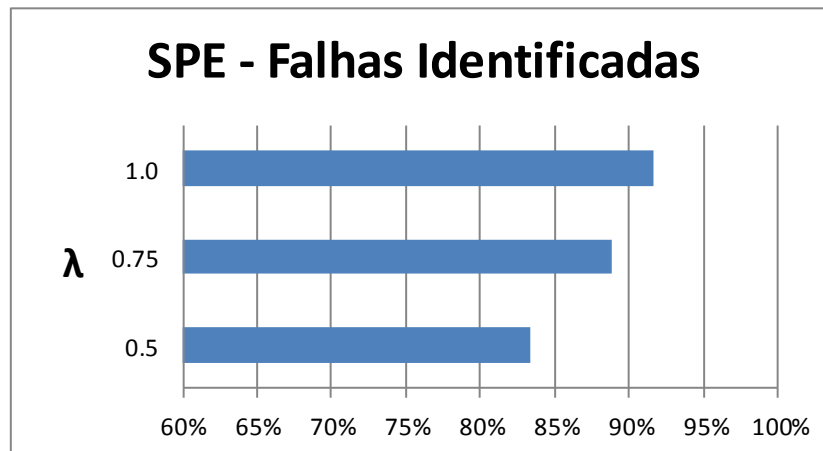
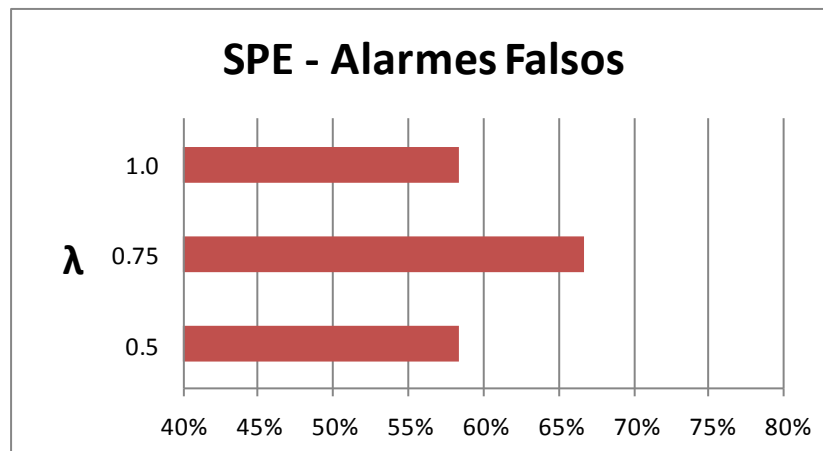


Figura 6-1: Porcentagem de falhas corretamente identificadas nos casos 1, 2 e 3 pelo SPE.

O aumento do efeito memória aumenta a contribuição dos dados sadios (Fase A) sobre o indicador, causando uma espécie de efeito “tampão” que dificulta a identificação de anomalias

(presentes na Fase B). Esperar-se-ia desta forma que um maior efeito memória funcionasse em prol de um menor índice de detecções, que foi de fato observado.

Entretanto, de forma análoga, seria esperado que um maior efeito memória melhorasse o desempenho do indicador em relação aos alarmes falsos nos casos 4 e 5. Não foi o que ocorreu, como se pode ver na Figura 6-2.

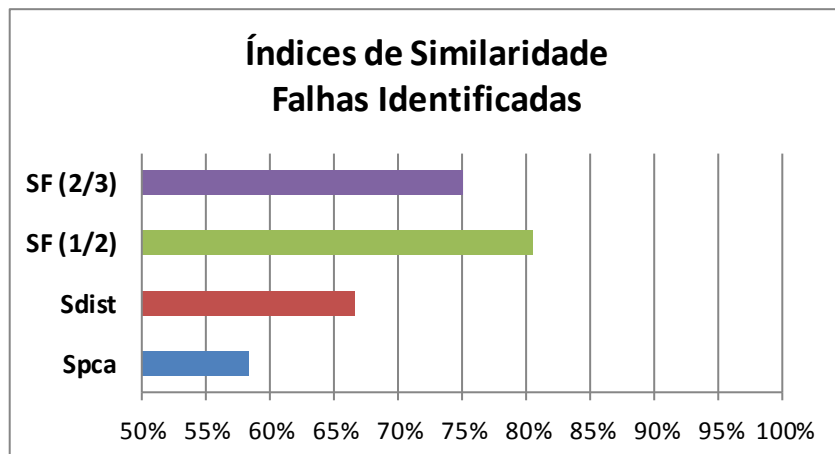


*Figura 6-2: Porcentagem de alarmes falsos gerados pelo SPE nos casos 4 e 5*

Nota-se que o SPE com valor de  $\lambda$  igual a 1.0 gerou tantos alarmes falsos quanto aquele com  $\lambda$  igual a 0.5, sendo o resultado não-conclusivo.

### 6.1.2 Índices de Similaridade

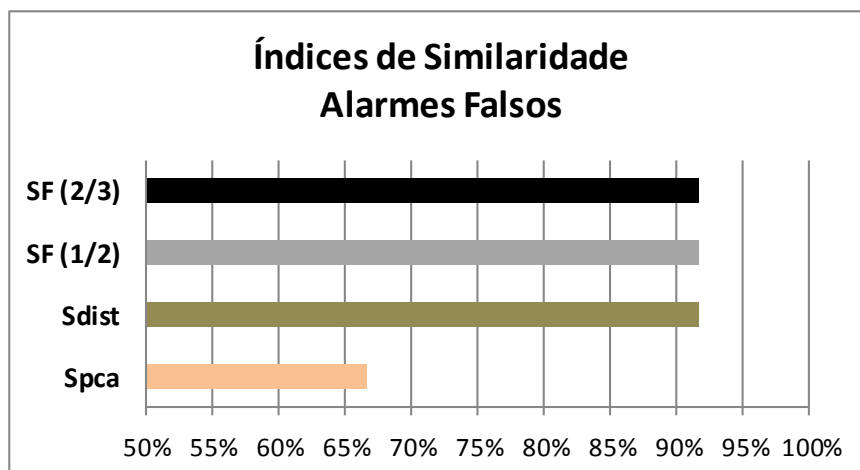
Os quantificadores  $S_{pca}$ ,  $S_{dist}$  e  $SF$  são apresentados neste item. Os resultados obtidos são mostrados na Figura 6-3.



*Figura 6-3: Porcentagem acertos na detecção de falhas por parte dos índices de similaridade estudados. Valores de  $\alpha$  entre parêntesis.*

O Fator de Similaridade – SF – foi o que apresentou melhores resultados, principalmente quando empregada a média aritmética ( $\alpha = 1/2$ ) entre Spca e Sdist. O Spca falhou em identificar grande parte das falhas estudadas. Porém, a combinação de ambos os índices (SF) supera o desempenho deles separadamente.

Os resultados destes mesmos indicadores nas análises dos casos 4 e 5 são mostrados na Figura 6-4.



*Figura 6-4: Porcentagem de alarmes falsos gerados pelos índices de similaridade estudados. Valores de  $\alpha$  entre parêntesis.*



Os casos 4 e 5 geraram grande quantidade de alarmes falsos nesses indicadores. O Spca, que apresentou o pior desempenho na detecção de falhas reais, foi o melhor neste quesito. Porém, o sucesso do Spca aqui pode não ser função de sua capacidade de evitar falsos alarmes, mas sim de sua incapacidade de detectar quaisquer variações nos dados, sejam falhas ou não.

## 6.2 Comparação entre os casos

Em todos os casos o número de componentes principais empregados foi em torno de 40% da quantidade total de sensores utilizados. Houve casos em que apenas 2 componentes principais foram suficientes e casos em que foram necessários 10 PCs (dentre 17 sensores) para se garantir 99% de representação.

A Tabela 6-1 resume o desempenho geral dos quantificadores em cada um dos casos e análises onde houve falha real de sensores. Ao todo foram 6 análises diferentes dos 3 casos de falhas reais e 7 indicadores foram avaliados (SPE x 3, Spca, Sdist, SF x 2), um total de 45 resultados para cada análise.

*Tabela 6-1: Comparação dos resultados caso a caso.*

	<b>Ranking</b>	<b>Falhas Detectadas</b>	<b>% de Detecção</b>
<b>Caso 2C</b>	1	38	90%
<b>Caso 3B</b>	2	37	88%
<b>Caso 3A</b>	3	36	86%
<b>Caso 2A</b>	4	34	81%
<b>Caso 1</b>	5	32	76%
<b>Caso 2B</b>	6	19	45%

Vê-se primeiramente que os casos com maior índice de sucesso são os casos onde observam-se duas falhas simultâneas: 2C e 3B. Aparentemente uma segunda falha intensifica a perda de correlação entre os sensores, aumentando a disparidade entre as Fases A e B e conseqüentemente melhorando a detecção via PCA.

As análises 2B e 2C, diferentes entre si apenas por *um* sensor apresentaram resultados muito díspares. Enquanto que 2C apresentou a melhor performance das 6 análises, 2B apresentou a pior taxa de detecção.

A partir dos resultados detalhados pode-se notar que o caso 2B não foi bem resolvido por nenhum dos fatores de similaridade. O Sdist, por exemplo, falhou em identificar anomalias nos dados dessa análise independente dos parâmetros empregados. Os outros indicadores também tiveram desempenho muito ruim nesse caso específico.

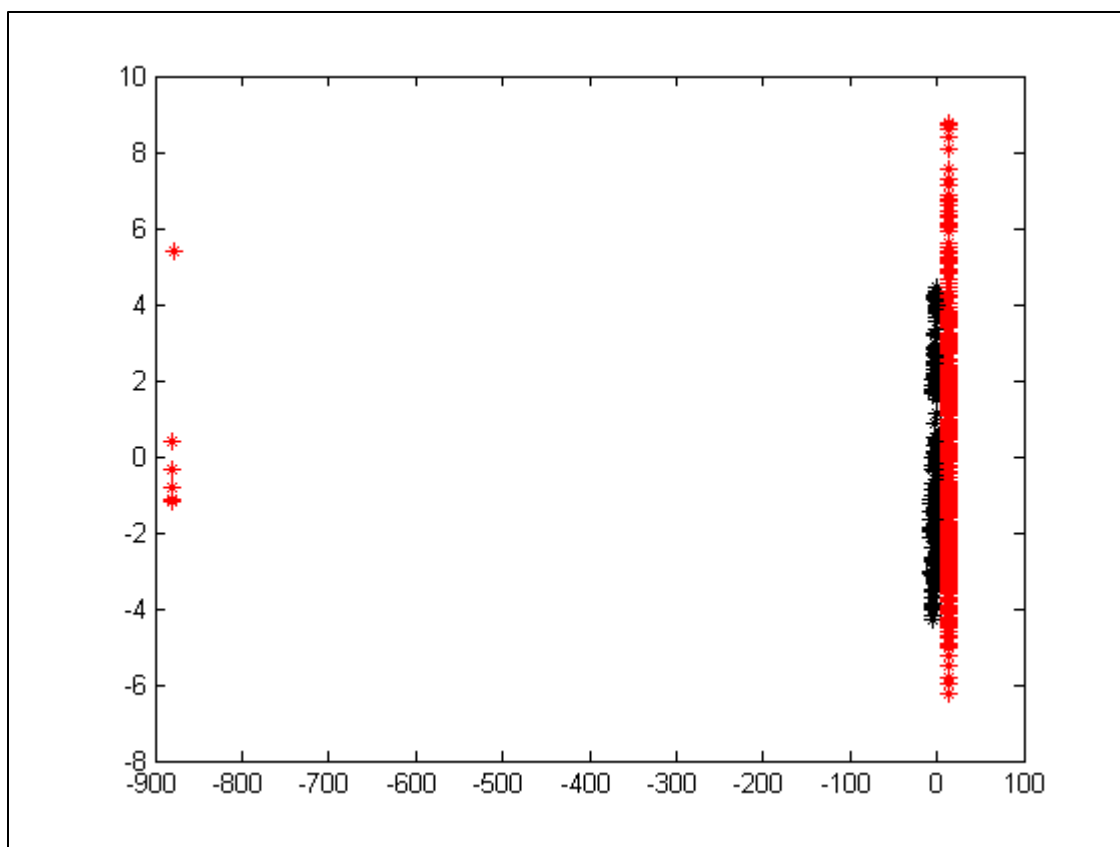
### **6.3 Resultados caso a caso**

Em cada um dos casos foram gerados gráficos na tentativa de se visualizar em 2D a aderência ou desvio do modelo. Essa análise apresentou resultados diversos. São apresentados apenas os gráficos que ilustram melhor o funcionamento do método. Os gráficos que apresentaram clusters aleatórios e que por tanto não permitiram uma análise objetiva, podem ser encontrados no Anexo B. Os gráficos 3D não apresentaram bons resultados, não acrescentando informações novas, logo foram desconsiderados.

#### **6.3.1 Caso 1**

O caso 1, apesar de apresentar uma falha total e instantânea, não apresentou os melhores resultados em detecção de falha, conforme Tabela 6.1.

A Figura 6-5 mostra a comparação entre os dados modelados da Fase A (preto) e os dados modelados da Fase B (vermelho) projetados sobre os dois PCs mais relevantes. Note que o pequeno *cluster* de pontos vermelhos à esquerda representa os pontos em falha. Porém a maior parte dos pontos da Fase B são erroneamente considerados sadios, alocados próximos aos dados da Fase A.



*Figura 6-5: Dados modelados do Caso 1. Em preto é representada a Fase A e em vermelho, a Fase B.*

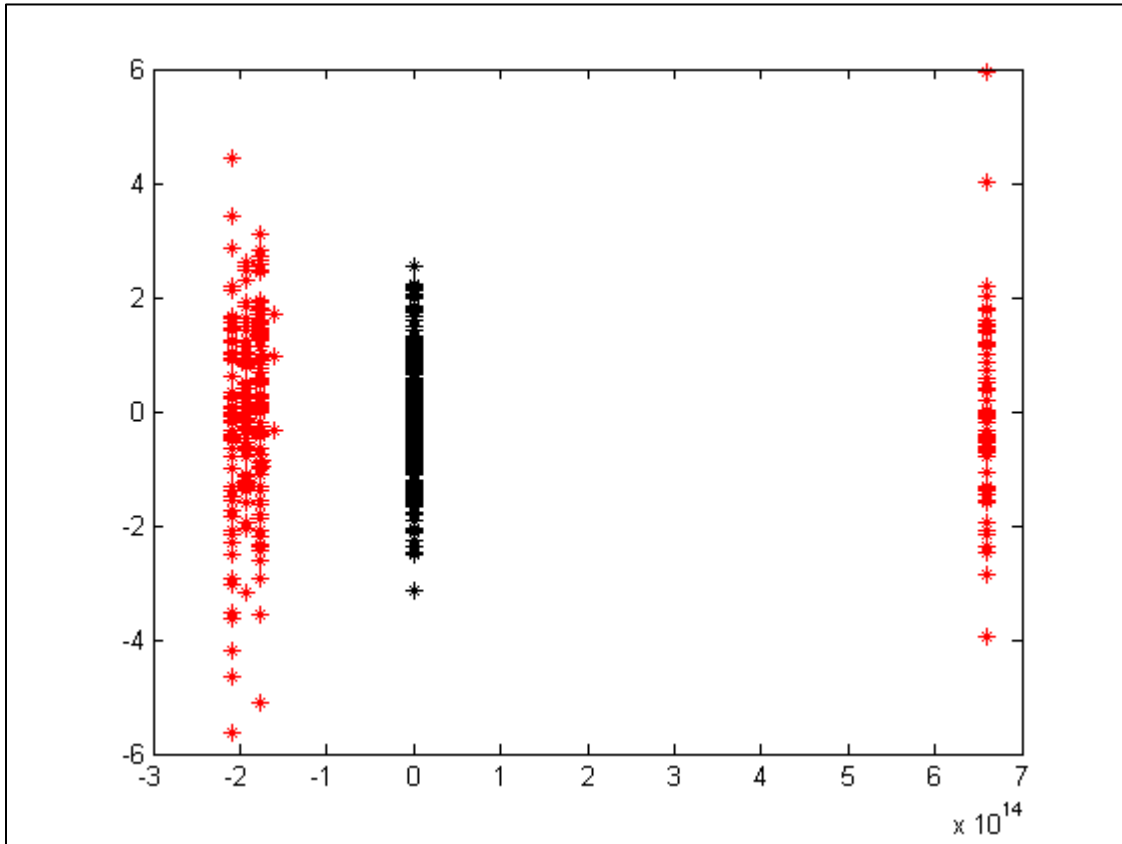
### 6.3.2 Caso 2

O Caso 2 e suas variações apresentaram resultados variados. O caso 2C, com dois sensores falhando concomitantemente, foi o que apresentou resultados mais promissores (Tabela 6.1). É compreensível que o caso 2A apresente um resultado pior devido à menor quantidade de falhas, o que também é apresentado na Tabela 6.1.

O caso 2B, contudo não teve sua falha detectada pelos índices de similaridade na série temporal de 1000 minutos. Spca e Sdist apresentaram valores elevados para todas as análises desse caso, com apenas 3 exceções, todas elas nas análises de menor intervalo (600 minutos).

Quanto à análise gráfica, pode-se ver na comparação entre os dados modelados de cada Fase um detalhe interessante: a Fase B, que apresenta falha oscilatória, representada em vermelho, aparentemente gerou dois clusters (Figura 6.6). Porém o cluster da esquerda, maior, parece indicar uma

leve separação entre seus eixos. Isso pode ser visto como fruto da oscilação dos valores do sensor, deslocando ligeiramente os dados modelados e alterando seu centro.



*Figura 6-6: Dados modelados do Caso 2. Em preto é representada a Fase A e em vermelho, a Fase B.*

### 6.3.3 Caso 3

Em termos gerais, esse caso apresentou bons resultados, conforme Tabela 6.1. Tanto no caso 3A quanto no caso 3B o SPE foi capaz de identificar a falha em todas as análises feitas, independente do valor de  $\lambda$  empregado.

O caso 3 não apresentou gráficos 2D uteis. Os pontos gerados mostraram-se dispersos, não sendo possível concluir nada a partir deles.

#### **6.3.4 Casos 4 e 5**

Os casos 4 e 5 foram usados para testar a hipótese que o PCA não seria capaz de, sozinho, distinguir uma mudança no processo de uma falha de sensor. Além disso, eles foram necessários para testar a premissa usada neste trabalho de que somente dados em estado estacionário poderiam gerar bons resultados.

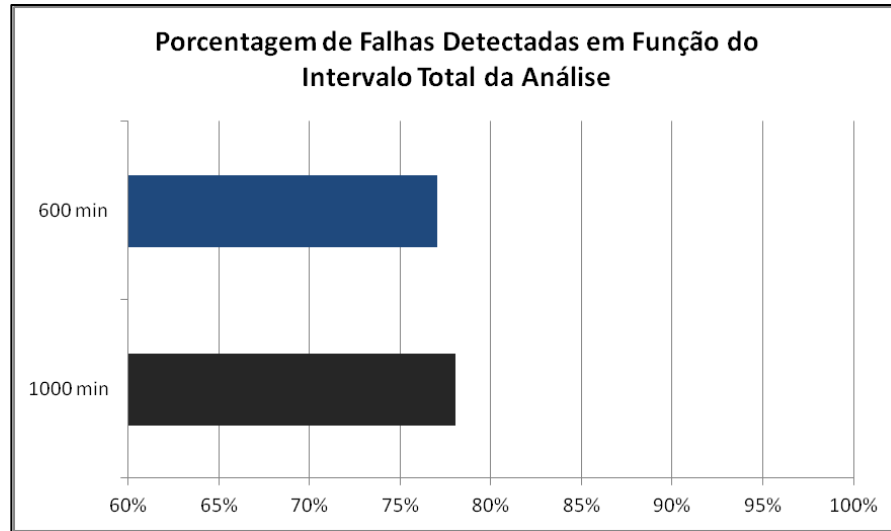
De fato alguns indicadores foram levados a detectar uma falha inexistente na análise dos dados da parada da BCS e nos dados de aumento de vazão gradual da mesma. As oscilações dos valores transmitidos pelos sensores, de forma pouco ordenada, causa um falso positivo.

Entretanto, o indicador SPE foi mais robusto e manteve-se no mais baixo nível de falsa detecção.

#### **6.4 Intervalo total e influência do instante da falha**

Não foi possível concluir se o método aqui empregado para detecção de falhas fica mais ou menos sensível com o aumento do período de falha dentro do intervalo temporal total de análise. Como se pode observar na Figura 6-7, no total, 77% das falhas foram detectadas quando se utilizou períodos de 1000 minutos e 78%, quando se utilizou períodos de 600 minutos.

Esse resultado pode denotar que intervalos menores podem ser usados. O sistema avaliado é um sistema bastante dinâmico e possivelmente uma escala de tempo mais curta trará os mesmos resultados de forma mais rápida.

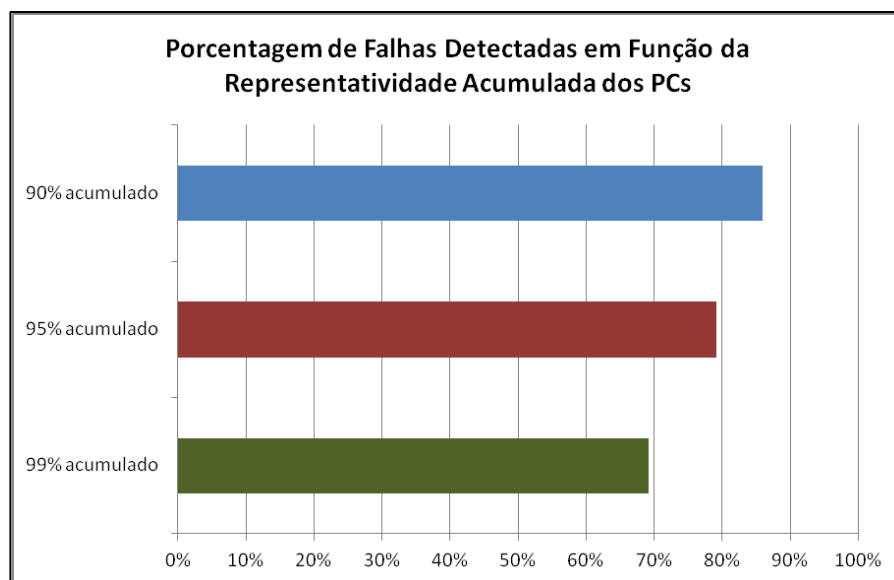


*Figura 6-7: Influência do intervalo total da análise sobre a detecção de falhas.*

O desempenho de cada quantificador, individualmente, também foi bastante homogêneo frente ao intervalo total empregado.

## 6.5 Porcentagem de representação pelos PCs

As diferentes porcentagens de representação por parte dos componentes principais (PCs) apresentaram um resultado inesperado. Análises feitas com menor quantidade de componentes principais, logo menor representação, foram mais bem sucedidas. Dentre as 84 análises feitas para cada valor mínimo de representação – 90, 95 e 99% – a primeira apresentou 72 detecções, a segunda, 66, e a terceira, 58 falhas detectadas. Percentualmente obteve-se 86%, 79% e 69% de acurácia respectivamente (Figura 6-8).



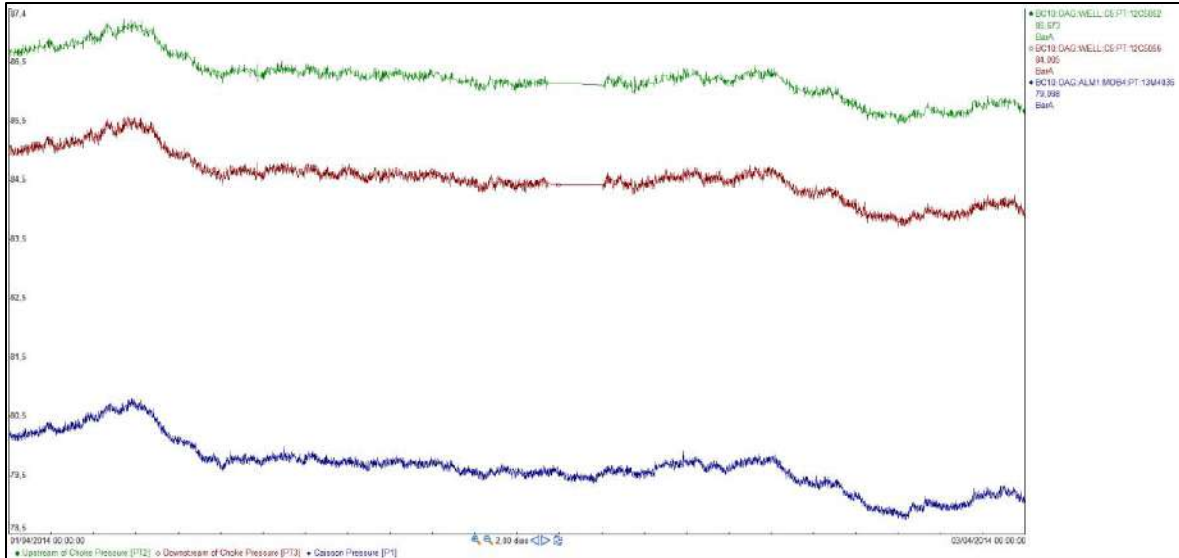
*Figura 6-8: Influência da representação mínima acumulada pelos componentes principais sobre a detecção de falhas.*

## 7 Discussão

### 7.1 Desafios encontrados

Um dos maiores desafios encontrados ao longo deste trabalho foi a obtenção de dados reais válidos para a análise proposta. Muitos dados deixavam de ser utilizáveis por problemas oriundos da natureza dos sensores, ou do sistema de coleta e transmissão de dados ou ainda devido aos processos em si. Ao longo da execução desta pesquisa foi-se obtendo o aprendizado de que tipos de pontos eram aceitáveis e quais deveriam ser evitados.

Por exemplo, é possível verificar no PI pequenos intervalos em que os dados não foram registrados, em geral devido a perdas de conexão com o servidor (Figura 7-1). Este problema costuma se apresentar como uma linha reta e horizontal, em geral curta. Não é algo simples de ser observado quando se está olhando para uma tela com mais de 10 sensores em um longo intervalo de tempo. Além disso, para fins operacionais, esses intervalos com dados interpolados são irrelevantes. Apenas no momento em que se coleta um conjunto de dados para fazer a análise estatística proposta que esse intervalo pode se tornar significativo e desviar o resultado do método.



*Figura 7-1: Momento em que dois sensores apresentam perda de sinal durante algumas horas.*

Complementarmente, alguns sensores geram dados que não podem ser aproveitados pelo método. Por exemplo, sensores que indicam a posição de válvulas costumam apresentar valores constantes, sem qualquer variação, durante todo o período em que a válvula não é atuada (aberta ou fechada ou % de abertura). Uma linha reta e horizontal apresenta desvio padrão igual a zero, logo não pode ser usada na análise proposta. Essa variável ainda apresenta variações em degraus, diferente de todas as outras, lineares, dificultando a correlação numérica. Com isso esses dados não puderam ser usados.

O último exemplo é quando um sensor falha em meio a uma mudança no processo, ou falha antes que uma quantidade suficiente de dados seja gerada para treinar o modelo PCA. Em ambos os casos o método não pôde ser aplicado.

Por todas essas razões, alguns casos, diferentes dos apresentados, tiveram que ser excluídos da análise final. Inclusive, o caso de um sensor com um tipo diferente de falha, na qual este congela em um valor aleatório (no caso, 100 °C), não pôde ser utilizado devido à má qualidade geral dos dados naquele momento.



## 7.2 Discussão dos Resultados

A partir do que foi apresentado as seguintes tendências puderam ser observadas:

- 1- A metodologia apresentada conseguiu detectar as falhas a partir de pelo menos 1 indicador em *todos* os casos.
- 2- O quantificador com maior taxa de sucesso foi o SPE com  $\lambda = 1.0$ . Ele detectou 92% das falhas e gerou a menor quantidade de alarmes falsos - 58%.
- 3- O quantificador com pior desempenho foi o Spca. Esse quantificador apresentou percentualmente mais alarmes falsos do que falhas de fato. Em suma, ele não funcionou propriamente nos dados reais estudados.
- 4- O Sdist apresentou comportamento inadequado em algumas das análises, gerando resultados como 1,00 ou  $10^{-16}$ . Isso leva a crer que alguns dos casos estudados pudessem ser inteiramente idênticos ou completamente díspares na comparação entre Fases A e B, o que não é razoável. Não foi possível verificar uma relação entre esse erro e o tipo de falha em questão. Além disso, a taxa de sucesso desse indicador não foi boa (67%).
- 5- O SF apresentou melhores resultados do que o Spca e o Sdist separadamente. Na maioria das vezes em que apenas um desses dois indicadores detectou a falha, o SF também a detectou. A criação desse quantificador se deu exatamente para unir a capacidade de detecção dos dois parâmetros de similaridade já conhecidos. O que foi observado é que o SF acaba também por minimizar as limitações do Spca e Sdist. O valor de alfa =  $\frac{1}{2}$  foi a melhor escolha.
- 6- A análise gráfica permite ver que o modelo consegue, muitas vezes, seguir os dados experimentais de forma precisa e indicar a falha na forma de espaçamento entre os pontos no espaço 2D. Porém, para detecção e análise de falhas em sistemas multidimensionais, essa análise gráfica não é suficiente e permanece como mero auxílio visual.
- 7- O tamanho do intervalo analisado não apresentou influência direta sobre a detecção de falha. O método provou-se versátil respondendo tanto a falhas no início da Fase B quanto a falhas no final desta mesma Fase. O peso do período *com* falha sobre o período total analisado parece não afetar o desempenho dos indicadores.

Finalmente, o Caso 2 – incluindo suas 3 análises – mostrou problemas no desempenho do método. A simples retirada de um sensor com falha fez com que o sucesso na detecção da falha caísse de 92% para 45% (casos 2C e 2B respectivamente). Além disso, comparando-se os casos 2A com 2B, trata-se de falhas análogas, ocorridas no mesmo sensor, usando os mesmos sensores para comparação, porém em momentos distintos. A falha no primeiro momento (2A) foi detectada com relativo sucesso, enquanto que no segundo momento (2B) foi quase invisível ao método. Isso mostra que o grau de confiança que se pode ter nessa análise não é muito elevado. As circunstâncias da falha parecem ter influência elevada sobre o resultado final.

## 8 Conclusão

Neste estudo, dados reais de sensores submarinos foram avaliados visando testar a metodologia conhecida de detecção e análise de falhas via PCA. Os dados sofreram pré-tratamento estatístico e diversos quantificadores foram utilizados para verificar os resultados. Diferentes parâmetros desses indicadores foram alterados e avaliados. Em suma, a metodologia PCA foi implementada de forma ampla, adotando diferentes abordagens encontradas na literatura. Os resultados obtidos em termos de desempenho geral foram razoáveis, porém mostraram que uma metodologia de detecção de falhas puramente baseada na análise dos componentes principais é insuficiente para garantir segurança e confiança nos resultados.

Cinco casos reais diferentes foram estudados, onde três de fato apresentavam falhas de sensores submarinos. Os melhores índices de detecção de falha ocorreram nos casos onde mais de um sensor falhava simultaneamente. A média de sucesso de detecção entre os casos com apenas uma falha foi de 72%, o que não pode ser considerado alto o suficiente para se ter confiança no método. Além disso, uma mesma falha oscilatória apresentou resultados bem diferentes simplesmente por estar dentro de um contexto diferente. Isso revela o quão frágil é o método em relação à seleção dos dados alimentados.

Por outro lado, bons resultados foram obtidos empregando o indicador SPE, tanto na detecção de falhas quanto evitando a geração de falsos alarmes. Esses resultados dão esperanças de que é possível que tal metodologia venha a ser bem sucedida, desde que conte com o acréscimo de técnicas suplementares.

Propõe-se para a continuação do presente estudo a implementação de técnicas complementares ao PCA – como redes neurais e ARX – que aplicados em conjunto aos componentes principais podem reduzir a dimensão dos dados e proporcionar uma análise mais robusta

## Bibliografia

1. *Multi-model based process condition monitoring of offshore oil and gas production process.* **Natarajan, Sathish and Srinivasan, Rajagopalan.** Cingapura : Chemical Engineering Research and Design, 2010, Vol. 88, pp. 572-591.
2. *Notas de aula de pós-graduação: Séries Temporais.* **Vaz Junior, Carlos André.**
3. **Shell, Brasil.** *BC-10 Operating Guidelines. Reference Data - Vol. 02.* s.l. : Documento Interno, Revisão 2013.
4. Shell Brasil. *Shell Intranet.* [Online] [Cited: 04 01, 2014.] <http://sww.shell.com/>.
5. **Shell, Brasil.** *BC-10 Field Development Plan.* s.l. : Documento Interno, 2009.
6. —. *BC-10 Subsea Process Piping & Instrumentation Drawing.* s.l. : Documento Interno, Revisão 2013.
7. —. *Instrument Failure Anomalies.* s.l. : Documento Interno, 2013.
8. *Run to Failure: BP and the Making of the Deepwater Horizon Disaster.* s.l. : W W Norton, 2012.
9. *Detecção e diagnóstico de falhas com base em dados históricos de processo: aplicação em dutovias.* **Vaz Junior, Carlos André.** UFRJ, Rio de Janeiro : Tese de Doutorado em Engenharia Química, 2010.
10. *Fault diagnosis by qualitative trend analysis of the principal components.* **Maurya, M. R., Rengaswamy, R. and Venkatasubramanian, V.** s.l. : Chemical Engineering Research and Design, 2005, Vol. 83, pp. 1122-1132.
11. *PCA-SDG based process monitoring and fault diagnosis.* **Vedam, Hiranmayee and Venkatasubramanian, Venkat.** s.l. : Control Engineering Practice, 1999, Vol. 7, pp. 903-917.

12. *An improved PCA scheme for sensor FDI: application to an air quality monitoring.* **Harkat, Mohamed-Faouzi, Mouro, Gilles and Ragot, José.** Nancy : Journal of Process Control, 2005, Vol. 16, pp. 625-634.
13. *An improved PCA method with application to boiler leak detection.* **Sun, Xi, Marquez, Horacio J. and Chen, Tongwen.** Edmonton : ISA Transactions, 2005, Vol. 44, pp. 379-397.
14. *Use of Principal Component Analysis for sensor fault identification.* **Dunia, Ricardo, et al., et al.** 1996, s.l. : Computers Chem. Engng., 1996, Computers chem. Engng, Vol. 20, pp. 713-718. 6A.
15. *Detection of nonlinearity effects in integrity monitoring methods for offshore jacket-type structures based on principal component analysis.* **Mojtahedi, A., et al., et al.** Tabriz : Marine Structures, 2013, Vol. 33, pp. 100-119.
16. *Data mining based sensor fault diagnosis and validation for building air conditioning system.* **Hou, Zhijian, et al., et al.** s.l. : Energy Conversion and Management, 2006, Vol. 47, pp. 2479-2490.
17. *Monitor and diagnosis of LNG plant fractionation process using k-mean clustering and principal component analysis.* **Pyun, Hahyung, et al., et al.** s.l. : Computer Aided Chemical Engineering, 2011, Vol. 29, pp. 1899-1903.
18. *How to use sustainability indicators for tourism planning: the case of rural tourism in Andalusia (Spain).* **Blancas, F. J., et al., et al.** Sevilla : Science of the Total Environment, 2011, Vols. 412-413, pp. 28-45.
19. *A comparative analysis regarding European tourism competitiveness: emerging versus developed markets.* **Gabor, Manuela Rozalia, Contiu, Lia Codrina and Oltean, Flavia Dana.** s.l. : Procedia Economics and Finance, 2012, Vol. 3, pp. 361-366.
20. *Assessing the need for process re-identification.* **Conner, Jeremy S. and Seborg, Dale E.** 2005, California : Ind. Eng. Chem. Res., 2005, Ind. Eng. Chem., Vol. 44, pp. 2767-2775. 8.
21. *Improved model predictive control using PCA.* **Yuan, Qiaolin and Lennox, Barry.** s.l. : IFAC Fault Detection, Supervision and Safety of Technical Processes, 2006.
22. **Smith, Lindsay I.** *A tutorial on Principal Components Analysis.* 2002.

23. *Model predictive controller monitoring based on pattern classification and PCA.* **Loquasto, Fred and Seborg, Dale E.** California : American Control Conferences, 2003.
24. *Pattern matching in multivariate time series databases using a moving-window approach.* **Singhal, Ashish and Seborg, Dale E.** 2002, California : Ind. Eng. Chem. Res., 2002, Vol. 41. 16.
25. *Multivariate statistical methods for monitoring continuous processes: assessment of discrimination power of disturbance models and diagnosis of multiple disturbances.* **Raich, A.C. and Çinar, A.** s.l. : Chemometrics and Intelligent Laboratory Systems, 1995, Vol. 30, pp. 37-48.
26. **Mathworks.** Matlab Help. [Online] [Cited: 08 25, 2013.] <http://www.mathworks.com/help/matlab/>.
27. **UFMS, Departamento de Engenharia Elétrica.** Curso de Matlab. [Online] [Cited: 08 25, 2013.] <http://www.del.ufms.br/tutoriais/matlab/apresentacao.htm>.
28. *Visualizing PI System Data.* **OSI Soft.** 2012, Vol. 5.4.
29. *Between-groups comparison of principal components.* **Krzanowski, W. J.** s.l. : Journal of American Statistical Association, 1979, Vol. 74, pp. 703-707.
30. *Determining the number of principal components for best reconstruction.* **Qin, S. Joe and Dunia, Ricardo.** Austin : Journal of Process Control, 2000, Vol. 10, pp. 245-250.

# Anexos

## Anexo A

Código do Matlab. As dimensões indicadas nos comentários se aplicam ao Caso 1 com 500 minutos para cada uma das Fases.

```
clear all
close all
clc

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% FELIPE LESBAUPIN - PCA %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% 2013-2014
% UFRJ - Escola de Química
% Monografia de Final de Curso
% Análise de Falha de Sensores Através de Métodos Estatísticos
Baseados em Componentes Principais

%% Passo 1 - Seleções dos dados brutos e determinação dos parâmetros

% ESCOLHER: Carregar os dados dos sensores do Caso desejado:
% Casos "a" são totalmente livres de anomalias, para comparação %

SensoresParaFelipe1a
%SensoresParaFelipe1b
%SensoresParaFelipe2a
%SensoresParaFelipe2b
%SensoresParaFelipe2c
%SensoresParaFelipe2d
%SensoresParaFelipe3a
%SensoresParaFelipe3b
%SensoresParaFelipe3c
%SensoresParaFelipe4a
%SensoresParaFelipe4b
%SensoresParaFelipe5a
%SensoresParaFelipe5b

% Dados originais - Fase A e Fase B, respectivamente:
T1bom;      T1falha;
%500x15    500x15

% Número de sensores:
[n,s] = size(T1bom);

% ESCOLHER: Definir um sensor bom e o sensor que falha:
sb = 1; % Número do Sensor Bom para comparar graficamente
sf = 5; % Número do Sensor que Falha para comparar graficamente
```

```

    % ESCOLHER: "a" define a % de representação dos k autovetores escolhidos
no modelo PCA
    x = [ 0.90 0.95 0.99 ];
        a = 1;
        %a = 2;
        %a = 3;

    % ESCOLHER: "p" define lambda, a ser usado para cálculo de SPE
lambda = [ 0.5 0.75 1.0 ];
        %p = 1;
        %p = 2;
        p = 3;

    % ESCOLHER: "q" define o valor de alfa para cálculo do SF
alfa = [ 1/2 2/3 ];
        q = 1;
        %q = 2;

% Gráfico dos dados brutos, Fase A e Fase B:

u = linspace(0, n-1, n);
u2 = linspace(0, 2*n-1, 2*n);
figure(13)
plot(u2, [T1bom ; T1falha], '-');
figure(15)
plot (u, T1bom', '-');
figure(14)
plot (u, T1falha', '-');

    %% Passo 2 - Tratamento dos dados

% Pré-Tratamento dos dados: Média nula e Desvio unitário:

DPbom = std(T1bom);
%1x15
Mbom = mean(T1bom);
%1x15
T1bomexp = zeros(n,s); % Alocando espaço para o loop
%500x15

for i = 1:s
T1bomexp(:,i) = ( T1bom(:,i) - Mbom(i) ) / DPbom(i);
%500x1          %500x1      %1x1
end

Mfalha = mean(T1falha);
%1x15
T1falhaexp = zeros(n,s); % Alocando espaço para o loop
%500x15

for i = 1:s
T1falhaexp(:,i) = ( T1falha(:,i) - Mfalha(i) ) / DPbom(i);
%500x1          %500x1      %1x1
end

```



```

% Novas matrizes de dados normalizados:
T1bomexp; T1falhaexp;
%500x15      500x15

% Gráfico dos dados normalizados:

figure(16)
plot(u2, [T1bomexp ; T1falhaexp], '-');
figure(17)
plot (u, T1bomexp', '-');
figure(18)
plot (u, T1falhaexp', '-');

                % Passo 3 - Aplicação do Método PCA

                %%%%%%%%%%%%%%% Aplicando o PCA: %%%%%%%%%%%%%%%

%% FASE A (treino, sem falha):

MatCovbom = cov(T1bomexp);
%15x15      500x15
[Vetbom,Valbom] = eig(MatCovbom);
%15x15  %15x15      15x15

% Rearrmando Autovetores em ordem decrescente de Autovalor:
[Variavelinutil,permutation] = sort(-diag(Valbom));
Valbom = Valbom(permutation,permutation); Vetbom = Vetbom(:,permutation);
%15x15      15x15

% Selecionando quantidade k de componentes principais para x% de
representação:
total = trace(Valbom);  aaa = sum(Valbom);  Racumul = cumsum(aaa,2);
%1x1      1x15      1x15

% Garantir x% de representação pelos Componentes Principais:
k = 1;
while ( Racumul(k) / total ) < x(a)
    k = k+1;
end

% Mostrar resultado ao usuário:
disp('Porcentagem representada pelo modelo:')
    ( Racumul(k) / total ) * 100
disp('Quantos Autovetores foram selecionados:')
    num2str(k)

% Gerando o modelo PCA - Matriz MC e Matriz C:
MCbom = Vetbom (:,1:k);      Cbom = MCbom * MCbom';
%15xk      %15x15  15xk      kx15

% Aplicando modelo PCA (Cbom) nos dados experimentais:
T1bommod = Cbom * T1bomexp';
%15x500    15x15    15x500

```

```

% Checando a aderência do modelo graficamente (Fase A):
% Plot usando apenas 2 das 15 linhas (sensor bom e ruim) de cada
matriz;
% Comparação entre dados experimentais e dados modelados:
figure(1)
plot(T1bommod(sb,:),T1bommod(sf,:), '*r') % Modelo, vermelho
%1x500      1x500
hold all
plot(T1bomexp(:,sb),T1bomexp(:,sf), '*k') % Experimental, preto
%500x1      500x1

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% FASE B (teste, com falha):

MatCovfalha = cov(T1falhaexp); [Vetfalha,Valfalha] = eig(MatCovfalha);
%15x15      %15x15      15x15

% Rearrumando Autovetores em ordem decrescente de autovalor:
[Variavelinutil2,permutation] = sort(-diag(Valfalha));
Valfalha = Valfalha(permutation,permutation); Vetfalha =
Vetfalha(:,permutation);
%15x15      15x15

% Gerando o modelo PCA - Matriz MC e Matriz C.
MCfalha = Vetfalha(:,1:k); Cfalha = MCfalha * MCfalha';
%15xk      %15x15      15xk      kx15

% Obs: k é o mesmo da Fase A:

% Aplicando modelo PCA (Cfalha) nos dados experimentais:
T1falhamod = Cbom * T1falhaexp';
%15x500      15x15      15x500

%Checando a aderência do modelo (Fase B):
% Plot usando apenas 2 das 15 linhas (sensor bom e ruim) de cada
matriz;
% Comparação entre dados experimentais e dados modelados:
figure(2)
plot(T1falhamod(sb,:),T1falhamod(sf,:), '*r') % Modelo, vermelho
%1x500      1x500
hold all
plot(T1falhaexp(:,sb),T1falhaexp(:,sf), '*k') % Experimental, preto
%500x1      500x1

%% Passo 4 - Análises dos Resultados

%% Verificando ocorrência da falha por análise gráfica

T1bommod_2 = MCbom' * T1bomexp';
%kx500      kx15      15x500
T1falhamod_2 = MCfalha' * T1falhaexp';

```

```

%kx500      kx15      15x500

% Análise 2D:

% Usando apenas os 2 primeiros PCs para verificar efeito da redução da
dimensão:
figure(3)
plot(T1bommod_2(1,:),T1bommod_2(2,:), '*k') % Fase A, preto
      %1x500      1x500
hold on
plot(T1falhamod_2(1,:),T1falhamod_2(2,:), '*r') % Fase B, vermelho
      %1x500      1x500

% Análise 3D:

% Usando apenas 3 das 15 linhas de cada matriz para um plot 3D (2 sbom e
1 sfalha):
figure(4)
plot3(T1bommod(sb,:),T1bommod(2,:),T1bommod(sf,:), '*k');
      %1x500      1x500      1x500
hold on
plot3(T1falhamod(sb,:),T1falhamod(2,:),T1falhamod(sf,:), '*r');
      %1x500      1x500      1x500

%%%%%%%%%% QUANTIFICADORES %%%%%%%%%%%

%% Cálculo de SPE: -----

% SPEbom (na Fase A):

T1bommod = T1bommod';
ebom = zeros(n,s);

for t = 1:s
    ebom(:,t) = T1bomexp(:,t) - T1bommod(:,t);
end

ebomfinal = ( (sum(ebom(:,t) .^ 2 , t)) ./ t ) .^ (0.5) ;
SPEbom = abs(ebomfinal) .^ 2;

%SPE (na Fase B):

T1falhamod = T1falhamod';
e = zeros(n,s);

for m = 1:s
    e(:,m) = T1falhaexp(:,m) - T1falhamod(:,m);
end

efinal = ( (sum(e(:,m) .^ 2, m))./ m ) .^ (0.5);
SPEfalha = abs(efinal) .^ 2;

```

```

SPE = [ SPEbom; SPEfalha ];
SPEmedio = [];

for y = 1:2*n
    MEDIA = sum(SPE(1:y)) / y; % Média aritmética dos valores de SPE para
    cada instante n
    NOVO = (1-lambda(p)) * MEDIA + lambda(p) * SPE(y);
           %1x1           1x1           1x1
    SPEmedio = [SPEmedio, NOVO];
end

SPEbom2 = sort(SPEbom,1,'descend');
MAX = SPEbom2(2);
u2 = linspace (0,2*n-1,2*n);

figure(10)
hold on
    plot (SPEmedio, '-k')
    plot(u2, MAX, '-r')
    plot(u2, -MAX, '-r')
    plot(u2, 2*MAX, '-g')
    plot(u2, -2*MAX, '-g')

% Determinação de % de falha segundo SPE:

FALHA = 0;

for p = 1:2*n

    if SPEmedio(p) > MAX
        FALHA = FALHA + 1;
    end
end

FALHA = FALHA / (2*n) * 100;

disp('Porcentagem de Erro por SPE (em %):')
    num2str(FALHA)

%%% Cálculo de Spca: -----

% Quanto mais próximo de 1, menor a probabilidade de falha:

Spca = ( trace (MCbom'*MCfalha * MCfalha'*MCbom) ) / k % k já foi
definido acima

%%% Cálculo de Sdist: -----

% Quanto mais próximo de 1, menor a probabilidade de falha

```

```

% Pseudoinverso da matriz covariança:
sigma = pinv(MatCovfalha);
% Distância de Mahalanobis:
teta = sqrt( (Mbom - Mfalha) * sigma * (Mbom - Mfalha)' );
% Calculando a integral:
z = sym('z'); integral = int( exp( (-z^2) / 2 ), z, -99999999, teta);
integral = eval(integral);
% Sdist:
Sdist = 2 * ( 1 - ( 1 / sqrt(2*pi)) * integral )

%%% Cálculo de SF: -----

% Quanto mais próximo de 1, menor a probabilidade de falha

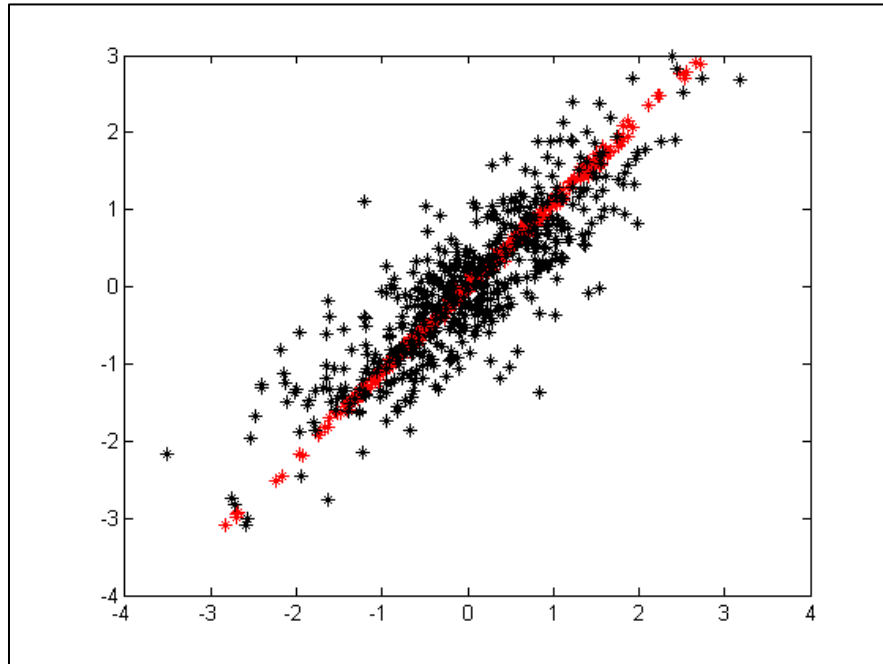
SF = alfa(q) * Spca + (1 - alfa(q)) * Sdist

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% FIM %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

## Anexo B

Gráficos 2D extras:



*Figura B-1: Caso 2B, dados experimentais em preto e dados modelados em vermelho. Dois sensores durante a Fase A são representados nessa figura. O modelo mostra boa aderência.*

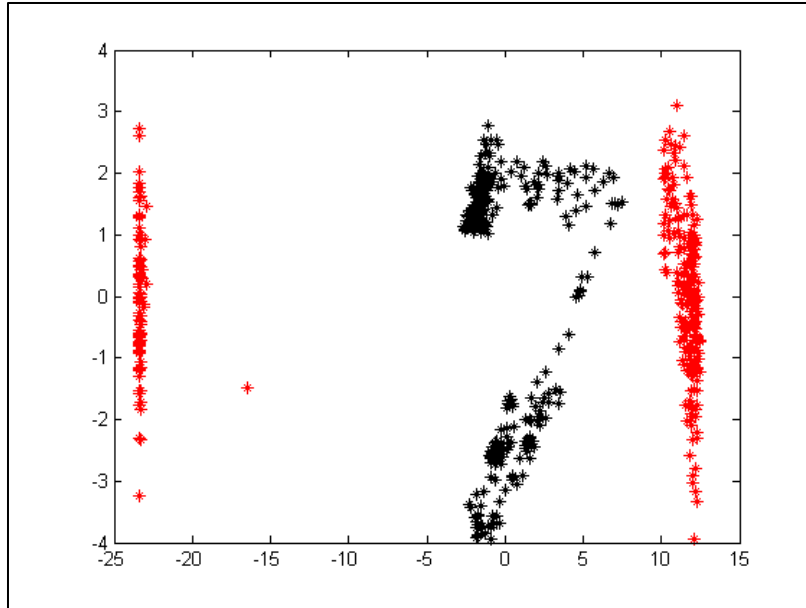


Figura B-2: Caso 3A, dados modelados da Fase A, em preto, e da Fase B, em vermelho. Um sensor sadio e um sensor ruim foram plotados.

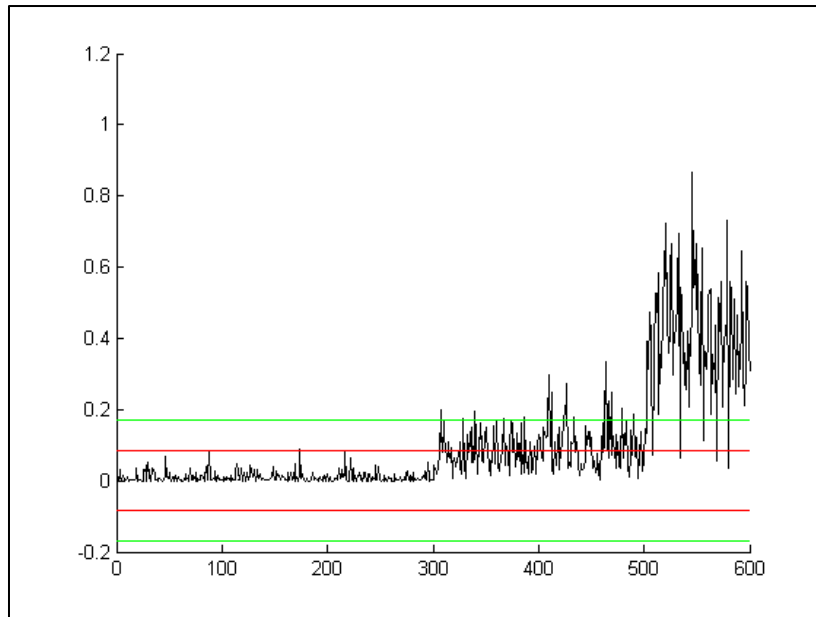


Figura B-3: Análise do SPE no Caso 2A A linha vermelha indica o limite aceitável.

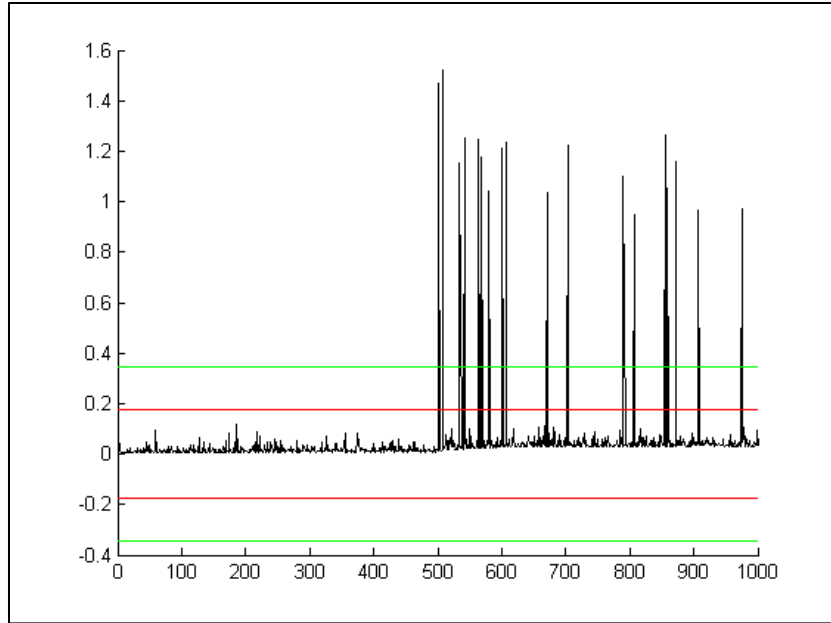


Figura B-4: Análise do SPE no Caso 2B. A linha vermelha indica o limite aceitável.

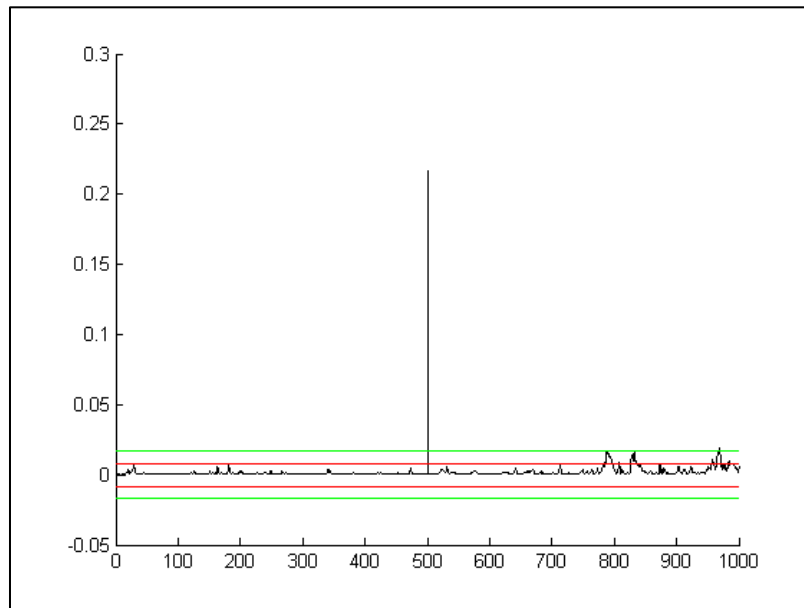


Figura B-5: Análise do SPE no Caso 3A. A linha vermelha indica o limite aceitável.



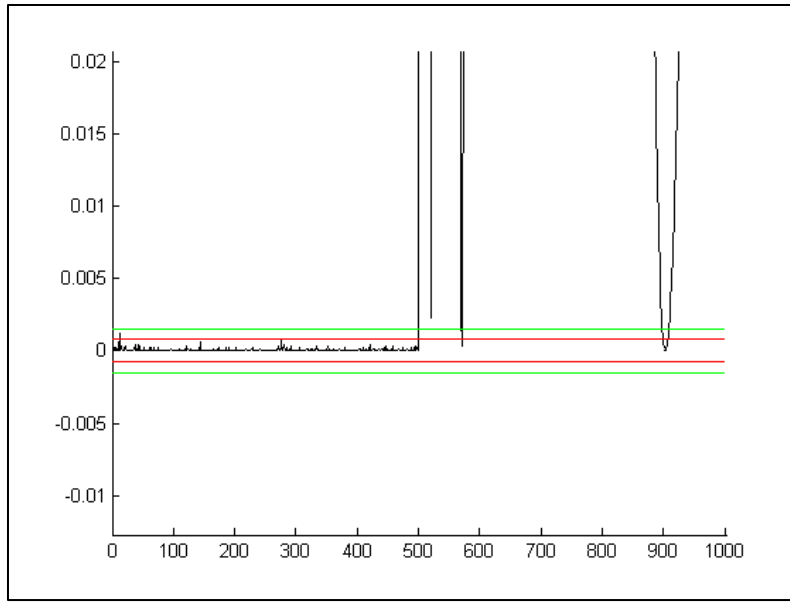


Figura B-6: Análise do SPE no Caso 4. A linha vermelha indica o limite aceitável.

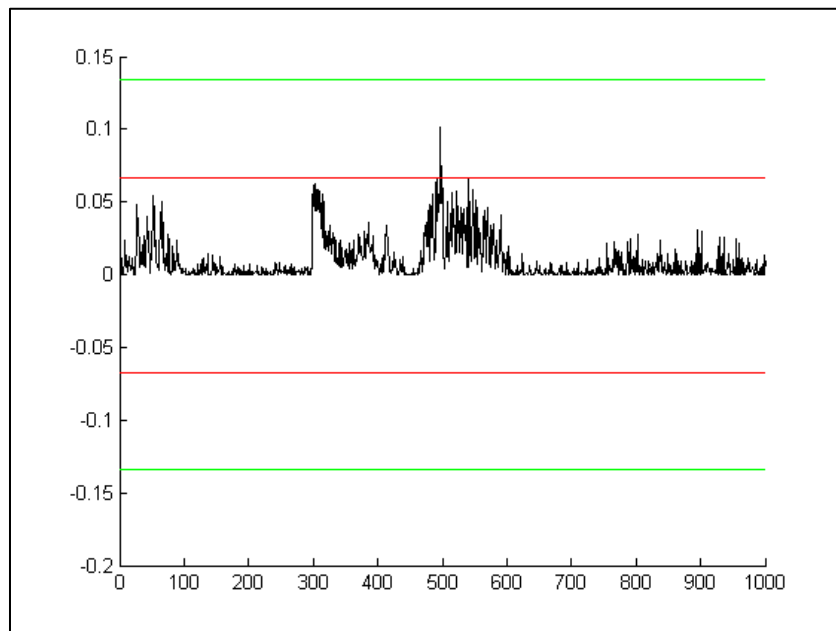


Figura B-7: Análise do SPE no Caso 5. A linha vermelha indica o limite aceitável. Nota-se que existem pequenos desvios já na Fase A devido a pequenas variações nos dados usados.