

# Detecção e identificação de falhas em um sistema preventor de explosões submarino (BOP)

Vanderson Carvalho<sup>1</sup>, Bruno Nunes<sup>2</sup>, Alessandro Copetti<sup>3</sup> e Ana Paula Sobral<sup>3</sup>

<sup>1</sup> GE Oil & Gas do Brasil, Macaé, Rio de Janeiro, Brasil

<sup>2</sup> GE Global Research, Rio de Janeiro, Rio de Janeiro, Brasil

<sup>3</sup> Universidade Federal Fluminense, Rio das Ostras, Rio de Janeiro, Brasil

**Resumo** O sistema de prevenção de explosões em poços de petróleo submarinos (*BlowOut Preventer*-BOP) possui papel indispensável na segurança das atividades exploratórias de petróleo em alto mar. Como se trata de um equipamento de segurança, a rapidez e a eficiência na descoberta de falhas são vitais para a rentabilidade do trabalho, a segurança da população na plataforma de petróleo e a proteção do meio ambiente. Atualmente, para os sistemas de BOP em geral, a análise dos dados históricos não é comum dada a sua complexidade e número de variáveis envolvidas. Contudo, quando essa análise é realizada, é feita de forma manual e o seu resultado totalmente dependente da experiência do operador encarregado de executá-la. No melhor de nosso entendimento, não existe uma metodologia que oriente o operador para detecção e identificação de falhas em sistemas BOP. Este artigo propõe uma metodologia contemplando preparação dos dados, aplicação da Análise de Componentes Principais e utilização de Cartas de Controle estatístico ( $T^2$  de Hotelling e Estatística  $Q$ ) para detectar a variabilidade e consequentemente falhas no histórico de dados do BOP. Além da detecção e identificação de falhas, a metodologia é capaz de extrair informações relevantes, como a frequência e a relação entre alarmes.

**Keywords:** BOP, preventor de explosão, detecção de falhas, estatística multivariada, análise de componentes principais, cartas de controle.

## 1 Introdução

No período de perfuração de um poço de petróleo em águas profundas existe a possibilidade de haver um *blowout*, ou fluxo indesejado de hidrocarbonetos (gás/óleo) proveniente da formação para a superfície, o que pode levar a situações catastróficas. Assim, todo poço de petróleo durante a fase de perfuração possui um equipamento Preventor de Explosão Submarino ou *BlowOut Preventer*-BOP, responsável por isolar o poço de petróleo caso necessário.

Os operadores do sistema BOP têm como responsabilidade monitorar os alarmes e atuar de acordo. Contudo, a complexidade da operação, a enorme quantidade de dados gerada e a falta de tempo hábil, dificultam a análise de qualquer possível falha que esteja associada ao alarme gerado. Além disso, mesmo havendo redundância de sensores, caso ocorra falha, os sistemas de BOP em geral não selecionam automaticamente o sensor redundante, ficando essa ação a cargo do operador. Por exemplo, um sensor defeituoso que ainda reporte um valor fixo,

dentro do limite operacional, dificilmente será identificado antes que sua falha gere consequências danosas à operacionalidade do sistema[10].

Neste trabalho, adota-se detecção de falha como o ato de determinar se um dado sistema experimenta problemas. Já a identificação da falha determina os agentes que contribuem para a ocorrência da falha, conforme [4]. Métodos estatísticos multivariados, tais como a Análise de Componentes Principais (ACP), em conjunto com testes estatísticos e cartas de controle, são bem estabelecidos como métodos eficazes de detecção e identificação de falhas [6][16]. No entanto, não existem metodologias que apresentem a aplicação de um conjunto de técnicas ou ferramentas para esse propósito em sistemas BOP.

A metodologia proposta neste trabalho compreende a preparação dos dados históricos do BOP, a aplicação da ACP para redução do conjunto de dados e a aplicação de testes estatísticos ( $T^2$  de Hotelling e Estatística  $Q$ ) para a detecção e identificação de falhas. Além disso, a metodologia é capaz de extrair informações relevantes, como a frequência e a relação entre alarmes.

Este artigo está organizado da seguinte forma: a seção 2 expõe o referencial teórico utilizado neste trabalho; a seção 3 apresenta a proposta da metodologia para a detecção e identificação de falhas e na seção 4 a metodologia proposta é aplicada a dados reais e dados simulados.

## 2 Referencial Teórico

Esta seção apresenta uma breve descrição do sistema BOP e as principais técnicas utilizadas na metodologia a ser apresentada na seção 3.

### 2.1 Sistema BOP

O sistema BOP é formado basicamente por um conjunto de válvulas, responsáveis por controlar e monitorar os fluxos de hidrocarbonetos nos poços de petróleo e gás. O BOP é projetado de forma a selar completamente o poço de petróleo e permitir que os influxos no poço sejam controlados [15]. Para isto, são utilizados diversos tipos de válvulas de grande porte e dispositivos mecânicos, hidráulicos, pneumáticos e elétricos, mantendo ligação direta do poço com a plataforma de perfuração através de um sistema de controle computacional.

O sistema de controle submarino do BOP opera através de dois módulos de controle eletro-hidráulicos localizados na parte superior do BOP, denominados *PODs* de controle azul e amarelo. Os *PODs* são independentes para prover redundância total ao sistema e portanto, uma maior confiabilidade. Os *PODs* têm como objetivo principal receber as informações do sistema de controle computacional de superfície, processar e executar as funções definidas pelo operador e ativar as válvulas de controle. Cada *POD* de controle possui também um conjunto de sensores utilizados para monitorar importantes variáveis do sistema BOP, tais como: pressão de suprimento do sistema, pressão de atuação das funções, temperatura dos fluidos, parâmetros elétricos, etc. Os valores históricos destas variáveis são armazenados pelo sistema de controle na superfície.

### 2.2 Análise de Componentes Principais

A Análise de Componentes Principais (ACP) é uma abordagem estatística que pode ser utilizada para analisar inter-relações entre um grande número de variáveis

em termos de suas dimensões inerentes comuns (fatores), além de fornecerem uma estimativa empírica da estrutura das variáveis consideradas. Isso significa que a ACP identifica um número menor de novas variáveis alternativas, não correlacionadas e que, de algum modo, sumarizem as informações principais das variáveis originais encontrando os fatores ou variáveis latentes [12].

A ACP é uma técnica que consiste em se reduzir o espaço amostral, transformando os sinais (variáveis) correlacionados em um novo conjunto de dados onde se mantém a maior parte da variância original.

A técnica desacopla o espaço amostral do vetor  $X$  em um novo conjunto de variáveis não correlacionadas que correspondem aos elementos do vetor  $Y$ . O  $j$ -ésimo componente principal pode ser obtido pela Equação 1.

$$Y_j = e_j' X = e_{j1} X_1 + e_{j2} X_2 + \dots + e_{jp} X_p \quad (1)$$

Onde,  $Y_j$  é denominado *score*.  $e_j$  é a  $j$ -ésima componente principal da matriz de amostras  $X$  normalizada pela média e desvio padrão.

Os elementos do autovetor  $e_j$  funcionam como pesos de importância de  $X_1, \dots, X_p$  na composição do  $j$ -ésimo componente principal  $Y_j$ , e são denominadas cargas (*loadings*), e são obtidos através da decomposição em valores.

O número de componentes principais é selecionado de forma que represente ao menos 95% da variância total do sistema[12].

### 2.3 Cartas de controle estatístico

O Controle Estatístico de Processos é um conjunto de técnicas estatísticas utilizadas para avaliação de um processo, com o objetivo de controle e melhoria de sua qualidade [14].

Neste trabalho serão utilizadas Cartas de Controle multivariada baseadas em componentes principais [11]. Essas cartas utilizam os resultados da ACP aplicados às Cartas de Controle multivariada  $T^2$  de Hotelling e da estatística  $Q$ , descritas respectivamente nas Subseções 2.4 e 2.5.

### 2.4 Estatística $T^2$ de Hotelling

A estatística  $T^2$  de Hotelling foi um dos primeiros métodos de controle multivariado. A detecção de falha com o uso da ACP e da estatística  $T^2$  de Hotelling, consiste no cálculo de  $T^2$  de menor dimensão calculado para cada nova amostra  $x$  através da Equação 2 [4]:

$$T^2 = x' P \Lambda_p^{-1} P' x \quad (2)$$

Onde,

$p$  é o número de componentes principais selecionados

$x$  é o vetor com uma amostra

$P$  é a matriz de componentes principais

$\Lambda_p$  é a matriz de autovalores da decomposição em valores singulares da matriz de covariância  $S$  reduzida a dimensão  $p$

A carta de controle de Hotelling tem como métrica a distância estatística entre uma amostra multivariada e a parte central do modelo estatístico dos

dados. A estatística  $T^2$  é definida pela distância de Mahalanobis. Em outras palavras, a estatística  $T^2$  é a medida da variação no modelo ACP [7].

A fim de se obter limite superior de controle para um determinado intervalo de confiança, cujo o valor é geralmente 95% [9], utiliza-se a distribuição  $F$  de Snedecor, conforme mostra a Equação 3:

$$T_{\alpha,p,n}^2 = \frac{p(n-1)(n+1)}{n(n-p)} F_{\alpha,p-n,p} \quad (3)$$

onde,  $p$  é o número de componentes principais,  $n$  o número de amostras de dados de treinamento, e  $\alpha$  é o nível de significância. Com isso, temos as seguintes condições para  $T^2$ : Se  $T^2 \geq T_{\alpha,p,n}^2 \implies$  Falha ou Se  $T^2 < T_{\alpha,p,n}^2 \implies$  Não-falha / condição normal

## 2.5 Estatística Q

De forma a complementar a estatística  $T^2$ , a porção do espaço amostral correspondente aos menores autovalores pode ser monitorada através da estatística  $Q$  ou SPE (*Squared Prediction Error*) [8]. A estatística  $Q$  é uma distância entre a amostra original e sua projeção no modelo ACP obtido, cuja dimensão é reduzida [5]. Essa distância é uma forma de se avaliar a qualidade do conjunto reduzido obtido, uma vez que durante o processo de redução dimensional há uma perda inerente de informações. Basicamente a estatística  $Q$  é a soma do quadrado de cada linha na matriz de resíduos, conforme Equação 4 [4]

$$Q = \|r\|^2 = \|x - \hat{x}\|^2 = \|(I - PP^T)x\|^2 \quad (4)$$

onde,  $r$  é o vetor de resíduos, ou seja, uma projeção do vetor  $x$  no espaço residual e  $P$  a matriz de carga (*loading matrix*).

A fórmula para o limite superior de controle ( $Q_\alpha$ ) para um determinado nível de significância  $\alpha$ , é mostrada na Equação 5 [9].

$$Q_\alpha = \theta_1 \left[ \frac{c_\alpha \sqrt{\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (5)$$

onde, seus parâmetros  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  e  $h_0$  são mostrados abaixo:

$$\theta_1 = \sum_{i=k+1}^m l_i, \theta_2 = \sum_{i=k+1}^m l_i^2, \theta_3 = \sum_{i=k+1}^m l_i^3, h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$$

onde,  $l$  são os autovalores da matriz de covariância dispostos em ordem decrescente.  $k$  é o número de componentes principais,  $m$  o número de variáveis.  $c_\alpha$  representa uma distribuição normal com um nível de significância  $\alpha$ . Normalmente o intervalo de confiança  $(1 - \alpha)$  para esta aplicação se encontra entre 95% e 98% [9].

Quando a estatística  $Q$  é calculada para um dado vetor de dados, seu valor é comparado com o limite superior de controle ( $Q_\alpha$ ), e caso este limite de controle seja violado, uma falha é detectada [9]. Logo, para os valores de  $Q_\alpha$  temos as seguintes condições: Se  $Q \geq Q_\alpha \implies$  Falha e Se  $Q < Q_\alpha \implies$  Não-falha / condição normal.

### 3 Metodologia Proposta

A metodologia proposta é constituída de quatro etapas e busca responder as seguintes questões:

- Q1: Qual a frequência individual de cada alarme e quais ocorrem com mais frequência?
- Q2: Quais alarmes ocorrem em conjunto?
- Q3: Em que ponto do histórico de dados uma falha é detectada?
- Q4: Quais as variáveis (sensores) que mais contribuem para a detecção de uma determinada falha?

O fluxograma das etapas da metodologia, onde estão destacadas as respostas para as questões (Q1...Q4), pode ser visualizado na Figura 1.

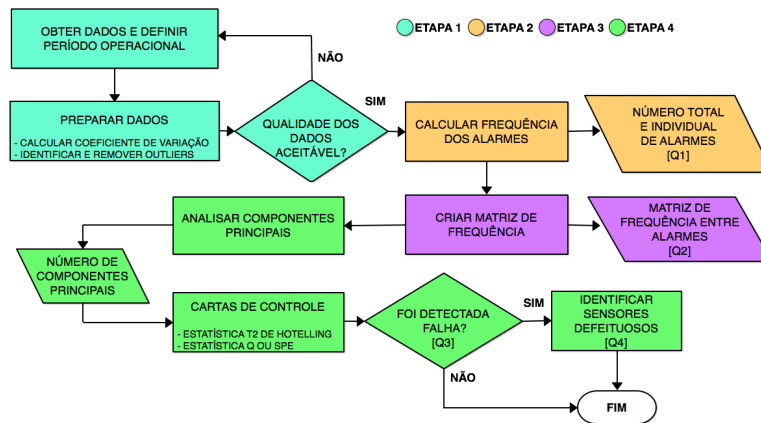


Figura 1: Fluxograma da metodologia proposta neste trabalho

As subseções 3.1, 3.2 e 3.3 apresentam as etapas 1, 2 e 3, respectivamente. A etapa 4 é apresentada na seção 4 e aplicada a uma base de dados que contém três anos de operação de um BOP a bordo de uma plataforma de perfuração, localizada na Bacia de Campos-RJ.

#### 3.1 Preparação dos dados - Etapa 1

Com o objetivo de garantir a qualidade dos dados sob análise, removendo informações irrelevantes, se faz necessário identificar os períodos em que o BOP estava em operação. Para tal apresenta-se a seguinte hipótese: o período de operação do BOP pode ser verificado através da análise dos dados provenientes de sensores de pressão embarcados no BOP. Após uma análise inicial para definir os sensores relevantes para inferir o início e o fim do período operacional, foram selecionados os sensores que medem a pressão hidrostática (cada POD de controle possui um par destes sensores) e o sensor de pressão que mede a pressão de fechamento do conector de travamento do BOP à cabeça do poço. Com isso, parte-se das seguintes premissas:

- a) Como trata-se de um equipamento utilizado em águas profundas, definiu-se que o valor da pressão hidrostática deve ser maior que 1500 psi, o que corresponde aproximadamente a 1000 m de lamina d'água, ou seja:  $P_{hidrostatica} \geq 1500 \text{ psi}$ . No entanto, não basta apenas verificar que o BOP encontra-se a uma profundidade  $\geq 1000\text{m}$ . É necessário verificar se ele encontra-se acoplado a cabeça do poço. É preciso então verificar a  $P_{wellheadconnector}$ .
- b) Uma vez que a pressão nominal de trabalho é da ordem de 1500 psi, a pressão de travamento dos conectores de travamento da cabeça de poço deve ser maior que 1500 psi, ou seja:  $P_{wellheadconnector} \geq 1500 \text{ psi}$ .

O próximo passo após a definição do período operacional consiste em definir o número de sensores a ser utilizado, tendo como critério a relevância em termos de contribuição para falhas no sistema BOP. O estudo realizado por [15] obteve a relação dos componentes do BOP que, naquele contexto, mais contribuíram para falha e conseqüentemente no tempo de parada de operação (*downtime*) do sistema. Através de questionário enviado a 10 especialistas [3] e com o auxílio dos resultados obtidos em [15], foram selecionados os dados provenientes de 37 sensores, os quais refletem o universo dos componentes com maior incidência de falhas no sistema do BOP.

Todos os dados provenientes dos sensores, alarmes e demais parâmetros operacionais do equipamento são gravados no sistema de controle na superfície, em uma tabela diária de dados. Estas tabelas diárias foram agregadas em uma nova tabela anual exportada para o ambiente estatístico R. Foi realizada a remoção de dados espúrios, remoção de *outliers* e análise do Coeficiente de Variação, onde o conjunto de dados apresentou um valor de dispersão de 11%.

### 3.2 Frequência dos alarmes - Etapa 2

O objetivo da segunda etapa é computar a frequência individual e total dos alarmes gerados pelo sistema. A análise de frequência para cada alarme é realizada através da contagem do número de ocorrências para cada código de alarme. Os alarmes são agrupados pelo seu identificador independentemente do POD de controle ao qual pertencem, uma vez que a troca do POD de controle durante o período de operação pode ser realizada a qualquer momento pelo operador. No estudo de caso, destaca-se que um alarme (ID 32), o qual se refere a entrada de água no compartimento de solenoide B, ocorreu 80656 vezes, o que equivale a 90,85% do número total de alarmes.

### 3.3 Matriz de frequência - Etapa 3

Nesta etapa 3 da metodologia, busca-se extrair do conjunto de dados a informação de quais alarmes ocorrem em conjunto, mesmo que não possuam uma relação direta um com o outro. Neste ponto, deseja-se investigar a relação numérica entre alarmes, ou seja, quando um determinado alarme ocorre, quais outros alarmes ocorrem em conjunto e com qual frequência.

Com esse objetivo, foi desenvolvida uma matriz de frequência, que mostra o cruzamento de frequência entre alarmes. Esta matriz é uma matriz quadrada  $n \times n$ , onde  $n$  é o número de variáveis investigadas, e indica a relação entre os alarmes contidos no histórico de dados, em um determinado intervalo de

tempo ( $t$ ). Para este estudo de caso, o valor considerado para  $n$  foi 37, que corresponde ao número de sensores monitorados. Já o intervalo de tempo  $t$  foi definido inicialmente em 10 segundos, pois o sistema de controle é configurado de forma que caso o valor da variável se encontre fora do intervalo pré definido, existe um *delay* da ordem de 10 segundos para que o alarme relacionado a essa variável seja efetivamente ativado. Isto ocorre para evitar que variações espúrias ativem alarmes inadvertidamente. A Figura 2 ilustra o gráfico da matriz de frequência. O par de variáveis onde existe um maior cruzamento de frequência possui uma maior intensidade de cor.

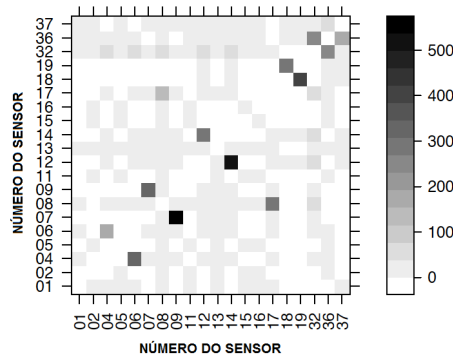


Figura 2: Matriz de cruzamento de frequência dos alarmes

A matriz pode ser utilizada pelos operadores com o objetivo de verificar a influência entre sensores de acordo com os alarmes ativados. Pode-se verificar, por exemplo, a relação não usual existente entre os sensores 32 (Entrada de água - compartimento de solenoides B) e o 36 (Temperatura - compartimento solenoides A). Essa relação pode indicar algum tipo de mal funcionamento do sistema, logo deve ser investigado pela equipe da plataforma.

#### 4 Detecção e identificação de falhas - Etapa 4

Nesta seção é aplicada a Análise de Componentes Principais (ACP) e as cartas de controle estatístico a dois conjuntos de dados; um de treinamento e um de teste. Um passo fundamental antes da aplicação da ACP é a normalização dos dados pelo desvio padrão e média. Isso é importante para garantir que tendências e diferentes escalas de medição, não interfiram no cálculo dos componentes principais.

A correta escolha dos dados de treinamento, responsáveis por treinar o modelo, é fundamental para o processo, uma vez que devem reproduzir a variabilidade natural dos dados [1]. O conjunto de dados de treinamento é constituído por um espaço amostral, onde se considera que as variáveis possuem valores, dentro de sua faixa de operação normal, ou seja, sem apresentar falha. Já os dados de teste correspondem aos dados de alarmes e simulados, caracterizando dados anormais, ou seja, fora do limite usual de operação.

A metodologia é aplicada em dados reais e simulados. Para a simulação são inseridas falhas que podem ocorrer de forma mais usual no dia-a-dia operacional, tais como *drift* nos valores ou problemas de repetibilidade, como descrito em [13]. Tais simulações se tornam necessárias pela falta de dados operacionais confiáveis sobre a ocorrência de falhas reais, uma vez que não é prática comum dos operadores o controle das falhas ocorridas.

#### 4.1 Dados reais de alarme

Nesta seção a etapa 4 da metodologia proposta (detecção e identificação de falhas) é aplicada aos dados dos sensores e alarmes. Através dos dados dos sensores em condição normal de funcionamento, treina-se o modelo para que o mesmo seja capaz de detectar a variabilidade (falhas) no conjunto de alarmes (teste). Nesta etapa foram utilizados apenas dados provenientes dos sensores de pressão (números de 1-20), que correspondem aos componentes do sistema BOP com maior índice de falha.

Foram selecionadas 200 amostras, pois para o conjunto de dados de treinamento mais de 50% dos sensores não possuíam dados a partir deste ponto, e caso a ACP fosse aplicada com essa grande quantidade de dados faltantes, seus resultados não seriam confiáveis [12].

O primeiro passo consiste em aplicar a ACP para ambos os conjuntos de dados (treinamento e teste), conforme explicado na Seção 2.2. Após o cálculo da ACP foram selecionados 7 componentes principais para ambos os conjuntos, que explicam 95% da variância dos dados. Tal resultado, que inclui a matriz de covariância, autovetores e autovalores para o conjuntos de dados de teste, foi utilizado para a recomposição da matriz de dados original utilizando o número de componentes principais selecionados.

Essas informações provenientes da aplicação da ACP são indispensáveis para o cálculo da estatística  $T^2$  de Hotelling e da estatística  $Q$ . O limite superior de controle calculado para o conjunto de dados de treinamento será utilizado como padrão para detecção da variabilidade no conjunto de dados de teste.

Na Figura 3a é mostrado o resultado da estatística  $T^2$  de Hotelling para os dados de teste e o valor do limite superior de 19,84, obtido a partir do conjunto de dados de treinamento. O resultado demonstra que a grande totalidade dos dados de teste está abaixo do limite superior de controle, e por esta razão são considerados dentro dos padrões de normalidade.

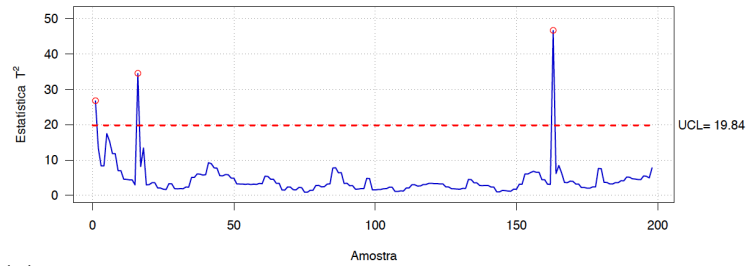
Como os dados de teste estão reconhecidamente fora do limite operacional, pode-se concluir que a estatística  $T^2$  de Hotelling não foi eficaz na detecção de falha neste caso. Isso ocorre, pois a estatística  $T^2$  de Hotelling trata somente da porção do espaço amostral explicado pelos componentes principais, e por essa razão, não é sensível a falhas com baixa magnitude [7], como é o caso da variação entre o conjunto de dados de teste e treinamento. Uma vez que os dados de teste possuem um valor com pouca magnitude acima ou abaixo do valor normal de operação.

O cálculo da estatística  $Q$  e do valor de seu limite de controle superior são realizados utilizando as Equações 4 e 5 respectivamente. A Figura 3b mostra o resultado do cálculo da estatística  $Q$  para o conjunto de dados de teste, e o valor do limite de controle superior de 2,99, calculado a partir dos dados de treinamento. Observa-se que o conjunto de dados de teste apresentam todos os



valores acima do limite superior calculado. Isso significa que todo o conjunto dados de teste possui valores fora da normalidade operacional, e comprova a capacidade do modelo de identificar quais dados se encontram fora do limite operacional, que é o caso dos dados de alarme utilizados como teste do modelo.

(a)



(b)

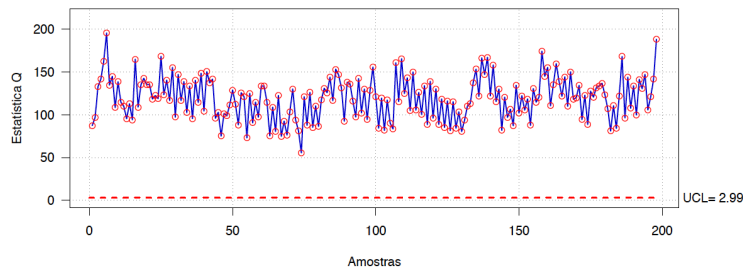


Figura 3: Estatística  $T^2$ (a) e  $Q$ (b) para os dados de teste

A identificação do sensor com falha é realizada através da análise da matriz de resíduos para uma determinada amostra. Aqui um determinado ponto é escolhido para o cálculo e criação do gráfico de contribuição para este ponto. Como demonstram os resultados da estatística  $Q$  para o conjunto de dados de teste, a totalidade dos pontos estão fora da zona de normalidade, por essa razão pode-se selecionar um ponto aleatório para verificação. Neste caso foi selecionado aleatoriamente o ponto amostral número 50, sendo que o sensor 17 exerce a maior influência para a falha detectada.

## 4.2 Simulação - Falha do tipo *sensor drift*

Para a simulação de *sensor drift*, foi selecionado aleatoriamente o sensor número 17, onde seus dados foram simulados a partir da amostra número 100 (cem) do *dataset* de treinamento (i.e., dados originalmente de operação normal, sem falhas). Esta falha simulada é caracterizada pela inserção de informações de pressão que aumentam de valor linearmente em função do tempo (*sensor drift*). Essa é uma falha comum para estes tipos de sensores devido suas características construtivas [13].

Para os dados simulados, após o cálculo da ACP, foram selecionados 6 componentes principais pelos critérios descritos na seção 2.2.

Na Figura 4a é mostrado o resultado para o cálculo da estatística  $T^2$  de Hotelling e seu limite superior de controle. Neste caso, pode ser visto que aproximadamente a partir da amostra número 130 do conjunto de dados, uma falha é detectada.

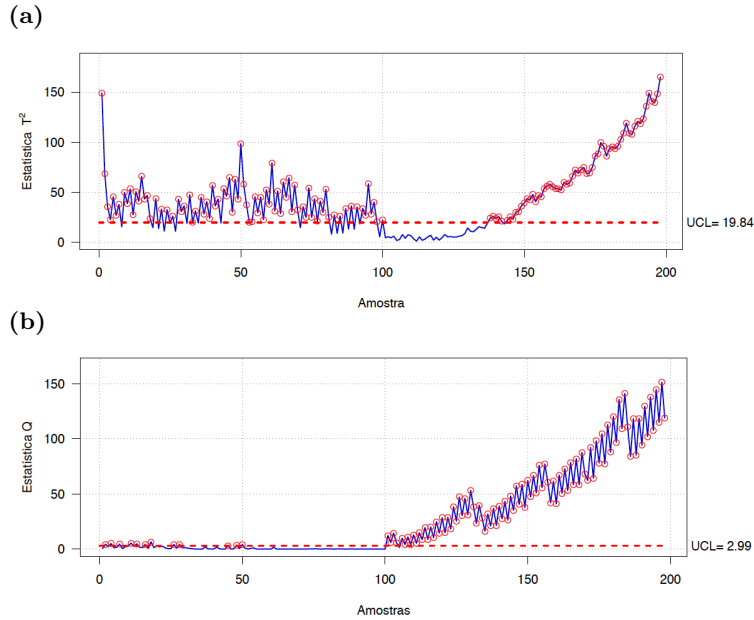


Figura 4: Estatística  $T^2$ (a) e  $Q$ (b) para os dados simulados *sensor drift*

O resultado da estatística  $Q$  para os dados simulados pode ser visto na Figura 4b. Pode-se claramente identificar a ocorrência de falha a partir da amostra número 100. A contribuição para a falha a partir do ponto amostral número 100 pode ser verificada através dos resultados obtidos. Como esperado, o sensor número 17 é o único contribuinte para a falha, uma vez que foi inserida uma "falha fictícia" em seus dados a partir da amostra número 100.

Uma vez obtido o resultado para a estatística  $Q$ , a contribuição para a falha a partir do ponto amostral número 100 pode ser verificada. Como esperado, o sensor número 17 é o único contribuinte para a falha, uma vez que foi inserida uma "falha fictícia" em seus dados a partir da amostra número 100.

### 4.3 Simulação - Falha do tipo repetibilidade

Por fim, foi simulada uma falha de repetibilidade no sensor número 12, onde a partir da amostra número 100 seus dados foram repetidos com o valor fixo de 10000 psi. Valor este, correspondente a máxima pressão que o sensor avaliado é capaz de suportar. Segundo [13], a repetibilidade é uma falha comum de ocorrer, dadas as características operacionais do sistema. Assim como ocorreu para a simulação de *sensor drift*, para os dados simulados, 6 componentes principais foram selecionados.

A partir do resultado do cálculo estatística  $T^2$  de Hotelling (Figura 5a), observa-se que a partir da amostra número 100 a curva pode ser considerada

constante, indicando que para algum sensor do conjunto de dados de teste, seus valores estão de forma anormal. Logo, o valor constante nos dados do sensor número 12 altera o comportamento da estatística  $T^2$  de Hotelling para este conjunto de dados.

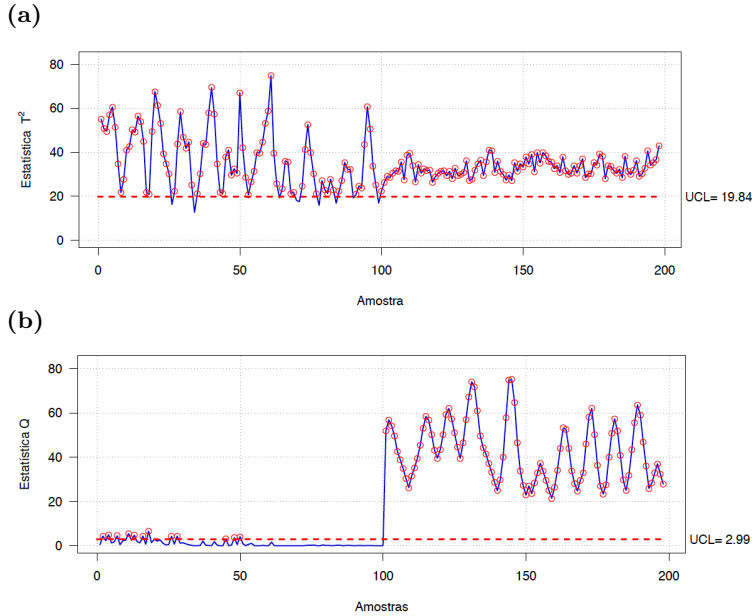


Figura 5: Estatística  $T^2$ (a) e  $Q$ (b) para os dados simulados - Repetibilidade

Pela estatística  $Q$  (Figura 5b), a contribuição para a falha a partir do ponto amostral número 100 pode ser inferida. Como esperado, o sensor número 12 se mostra o único contribuinte para a falha de repetibilidade simulada.

## 5 Conclusão

Neste trabalho foi apresentada uma metodologia para o aprimoramento da detecção e identificação de falhas em um Sistema Preventor de Explosões Submarino (BOP). Ferramentas estatísticas e computacionais são aplicadas a dados históricos provenientes desse sistema, com o objetivo de detectar e identificar falhas. A metodologia foi aplicada a um conjunto de dados reais fora do limite operacional e a dados simulados com falhas conhecidas para este tipo de equipamento/sensor, tais como: *sensor drift* e repetibilidade, onde em ambos os casos, a metodologia proposta foi capaz de detectar e identificar a origem da falha de forma consistente. Além da detecção e identificação de falhas, a metodologia é capaz de extrair informações relevantes, como a frequência e relação entre alarmes. Estas informações disponibilizadas aos operadores colaboram para a redução do tempo necessário para o diagnóstico e a posterior resolução da falha, contribuindo significativamente para reduzir as perdas operacionais e aumentar a confiabilidade geral do sistema BOP. Pelos resultados obtidos, considera-se

que a metodologia desenvolvida pode ser facilmente adaptada para aplicação em históricos de dados provenientes de outros sistemas BOP.

Por fim, na comparação com outras metodologias de detecção e identificação de falhas não aplicadas ao BOP [6][10][2], este trabalho tem seu diferencial no monitoramento da qualidade dos dados (Etapa 1) e na visualização rápida da relação entre variáveis pela matriz de frequência (Etapas 2 e 3). Já a Etapa 4 é comum nos trabalhos citados.

## Referências

1. Bezergianni, S., Kalogianni, A.: Application of Principal Component Analysis for monitoring and disturbance detection of a hydrotreating process. *Industrial & Engineering Chemistry Research* 47(18), 6972–6982 (2008)
2. Cai, B., Zhao, Y., Liu, H., Xie, M.: A data-driven fault diagnosis methodology in three-phase inverters for pmsm drive systems. *IEEE Transactions on Power Electronics* 32(7), 5590–5600 (2017)
3. Carvalho, V.P.: Aprimoramento da detecção e identificação de falhas em um sistema preventor de explosões submarinas (BOP). Dissertação de mestrado, Universidade Federal Fluminense (2016)
4. Chiang, L.H., Russell, E.L., Braatz, R.D., Russell, E.L.: Fault detection and diagnosis in industrial systems. *Advanced Textbooks in Control and Signal Processing*, Springer London, London (2001)
5. Dunia, R., Qin, S.J., Edgar, T.F., McAvoy, T.J.: Identification of faulty sensors using principal component analysis. *AIChE Journal* 42(10), 2797–2812 (oct 1996)
6. Gupta, S., Saputelli, L., Nikolaou, M.: Applying Big Data Analytics to Detect, Diagnose, and Prevent Impending Failures in Electric Submersible Pumps. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, Dubai (sep 2016)
7. Harrou, F., Kadri, F., Chaabane, S., Tahon, C., Sun, Y.: Improved Principal Component Analysis for anomaly detection: Application to an emergency department. *Computers & Industrial Engineering* 88, 63–67 (2015)
8. He, T., Xie, W.R., Wu, Q.H., Shi, T.L.: Process fault detection and diagnosis based on principal component analysis. *Fifth International Conference on Machine Learning and Cybernetics* 1(August), 3551–3556 (2006)
9. Jackson, J.E., Mudholkar, G.S.: Control procedures for residuals associated with Principal Component Analysis. *Technometrics* 21(3), 341–349 (1979)
10. Ji, H., He, X., Shang, J., Zhou, D.: Incipient fault detection with smoothing techniques in statistical process monitoring. *Control Engineering Practice* 62(February), 11–21 (2017)
11. Martin, E., Morris, A., Zhang, J.: Process performance monitoring using multivariate statistical process control. *IEE Proceedings of Control Theory and Applications* 143(2), 132–144 (2002)
12. Mingoti, S.A.: Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada. Editora UFMG, Belo Horizonte, 1 edn. (2005)
13. Mishra, R., Dubey, S.M., Shrivastava, K.K.: Analysis of system error log using Association Mining. *UACEE International Journal of Computer Science and its Applications* pp. 12–17 (2013)
14. Montgomery, D.: Introduction to statistical quality control. John Wiley & Sons, Inc., 3 edn. (2009)
15. da Silva, S.G.: Confiabilidade de equipamentos submarinos de segurança de sondas flutuantes (BlowOut Preventers). Dissertação de mestrado, UFRJ (2002)
16. Venkatasubramanian, V.: A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering* 27, 293–311 (2003)