



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
ENG07053 - TRABALHO DE DIPLOMAÇÃO EM ENGENHARIA
QUÍMICA



Uso da Distância de Mahalanobis para a Detecção de Rompimentos em Linhas de Injeção de Gás

Autor: Eduardo João Damiani

Orientador: Prof. Dr. Marcelo Farenzena

Coorientador: Me. Fabio Cesar Diehl

Porto Alegre, maio de 2021

Autor: Eduardo João Damiani

Uso da Distância de Mahalanobis para a Detecção de Rompimentos em Linhas de Injeção de Gás

Trabalho de Conclusão de Curso apresentado à COMGRAD/ENQ da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Bacharel em Engenharia Química

Orientador: Prof. Dr. Marcelo Farenzena
Coorientador: Me. Fabio Cesar Diehl

Banca Examinadora:

Prof. Dr. Jorge Otávio Trierweiler, Universidade Federal do Rio Grande do Sul

Me. Thiago Koichi Anzai, CENPES/Petrobras

Porto Alegre

2021

AGRADECIMENTOS

Aos meus pais, Andréia e Ivair, o agradecimento especial por todos os anos de convívio, aprendizado, amor e evolução que a Vida nos proporcionou. Agradeço as oportunidades que vocês me permitiram acessar e que poucas pessoas têm o privilégio de receber nesse país. João Pedro, obrigado pela cumplicidade nas conversas que só tenho com um irmão. Obrigado pelas mensagens de incentivo, pelas ligações às noites para saber como estava, as boas-vindas em Passo Fundo e as despedidas quando elas também eram necessárias. Navegamos por mares turbulentos quando sabemos da existência de portos seguros.

Aos meus avós, Darci, Lucila, Gezerino e Ivo, agradeço o presente de ser neto de pessoas tão especiais. Obrigado pelo amor que recebi durante o meu crescimento. Os seus ensinamentos jamais serão esquecidos. Uma pena que o destino não nos permitiu comemorar todos os frutos do “trabalho de avós” que vocês teriam durante as conquistas que estão surgindo em minha vida. Desejo muita luz onde estiverem.

À Universidade Federal do Rio Grande do Sul, em nome dos seus professores, técnicos, servidores e terceirizados, pela recepção de braços abertos do aluno com sede de adquirir conhecimento e promover para o desenvolvimento do Brasil. Vocês são dignos de aplausos por fazerem tanto diante do desinvestimento que Universidades Públicas sofrem nos últimos anos. Ao Laboratório de Materiais Aplicados e Interfaces, ao Diretório Acadêmico dos Estudantes de Engenharia Química, ao Centro do Estudantes de Engenharia e à EPR Consultoria, por permitirem o complemento da minha formação técnica e pessoal. Espero ter retribuído pelo menos uma parte daquilo que recebi das pessoas com quem tive o privilégio de conviver.

À Braskem, pela oportunidade do estágio o qual consolidou a minha formação como Engenheiro Químico. Em especial, Gustavo, Cristhian, Alexandre, Luciana e demais, muito obrigado pelo conhecimento compartilhado nesses últimos dois anos.

Agora sim, aos amigos que estiveram comigo em tantos momentos ao longo dos anos. Nos bastidores de todas as conquistas, vocês são as estrelas da minha graduação e espero que assim se mantenham para a Vida. Débora, Natália, Tainara e Amanda: do início ao fim, juntos. À família Doo, agradeço pela oportunidade que tenho de crescer em todos os nossos encontros. Rodrigo, em especial, obrigado pelo apoio com esse trabalho. Daniele, Pietro, Henrique, Bárbara, Arthur, Bibiana, Tamizi e Luiza, agradeço a amizade e o orgulho daquilo que vivemos juntos.

Jamais serei capaz de retribuir o que tantas pessoas fizeram por mim. Espero ter escolhido com sabedoria os nomes para representar algo tão importante para mim.

RESUMO

O transporte dos mais variados fluidos através de tubulações sempre representou uma alternativa de baixo custo para a distribuição de produtos de extrema importância para a sociedade. Ademais, a tentativa constante de melhoria em processos industriais promove a busca pelo controle e minimização de perdas por vazamentos em redes de distribuição, não sendo diferente no segmento do Óleo e Gás. Devido à implementação de recentes tecnologias em unidades de exploração e produção de petróleo, incidentes relacionados com a ruptura de dutos flexíveis utilizados para a injeção de gases em reservatórios têm despertado interesse para a pesquisa sobre as causas e tentativas de mitigação dessa falha. Na área da Engenharia de Segurança de Processos, pesquisas se concentram na identificação da ocorrência dos vazamentos nas tubulações com o objetivo de minimizar as consequências provocadas pelos eventos. Contudo, abordagens que utilizem métodos estatísticos são pouco exploradas pela literatura. Dessa forma, o presente estudo propôs o cálculo da Distância de Mahalanobis, utilizado para o reconhecimento de padrões e a *clusterização* de bancos de dados, como método para a identificação da ocorrência de rompimentos em tubulações de transporte de gás. O algoritmo implementado para o monitoramento do processo apresentou a capacidade de identificar o rompimento das tubulações em 9 dos 10 casos e apresentou resultados de falsos positivos em 9 dos 33 desvios ou distúrbios de processo estudados. Os resultados da detecção foram comparados com a técnica de classificação por Redes Neurais a fim de avaliar, de forma comparativa, os resultados do método. Os resultados para a detecção por redes neurais apresentaram 100% de acurácia na detecção dos rompimentos, tornando-se uma alternativa promissora para a detecção da falha estudada. Também foi possível determinar que o método de detecção pelo cálculo da Distância de Mahalanobis pode ser utilizado como forma de monitoramento da ocorrência de rompimentos de *risers* em aplicações de injeção de gás em águas profundas desde que avaliados os limites de confiabilidade da sua detecção e os riscos associados à presença de falsos positivos.

Palavras-chave: *Detecção de Vazamentos, Reconhecimento de Padrões, Tubulações de Gás, Distância de Mahalanobis.*

ABSTRACT

Fluid transport through pipelines always represent a low-cost alternative for the distribution of extremely important products. Moreover, the constant pursuit for industrial process improvements promotes the search for controlling and minimizing leakages losses in distribution networks, also in Oil and Gas segment. Due to recent technologies implementation in petroleum extraction and production units, events related to rupture of riser pipelines used for gas injection in oil wells have been drawing attention of this flaw cause and possibilities of mitigating it. Process safety engineering research focus on identifying the leakage, in order to mitigate the negative effects caused by the event. However, statistical methods are rarely proposed by literature. Thus, the current study proposes a new method for leakage events identification in off-shore units flexible pipelines from Mahalanobis Distance, commonly used for pattern recognition and data base classification. The algorithm presented for the process monitoring showed a potential of rupture detection in 9 of 10 case and false positives results in 9 of 33 process' changes. Detection results obtained were compared to Neural Network classification technic to assess the method efficiency. The results for the detection by neural networks showed 100% accuracy, making it a promising alternative for the detection of the studied failure. It was also possible to determine that the detection method by calculating the Mahalanobis Distance can be used as an alternative to monitor the occurrence of riser ruptures in deep water gas injection applications provided that the limits of reliability of its detection and risks associated with the presence of false positives.

Keywords: *Leak detection, Pattern Recognition, Gas Pipelines, Mahalanobis Distance*

LISTA DE FIGURAS

- Figura 1.** Desenho esquemático da tubulação com as respostas da pressão (P) e vazão total (qt) para: (a) Estado estacionário antes do vazamento; e (b) Estado estacionário depois do vazamento.....5
- Figura 2.** Comportamento transiente da pressão em um ponto de medição ($X_T = 0,8L$) ao longo da tubulação para os casos sem vazamento (preto) e com vazamentos de diferentes magnitudes (cor).....6
- Figura 3.** Diagrama simplificado de uma planta de compressão e injeção de gás..... 10
- Figura 4.** Representação hipotética das amostras coletadas para o cálculo da Distância de Mahalanobis nos instantes: a) Início do rompimento; e b) Durante a ocorrência do rompimento.....15
- Figura 5.** Comportamento dinâmico do sistema diante de desvios de processo iguais a 10% e de um caso rompimento da tubulação (1% de abertura e $C_v = 106$ US GPM) para as variáveis: a) FIC-FEED – PV; e b) PIC-SPEED – PV.....18
- Figura 6.** Espaço de Fase para a Variável de Processo e a Saída do controlador PIC-SPEED nos casos de desvio do FIC-FEED, do PIC-SPEED, na fração molar de CO_2 e nos casos de rompimento.....20
- Figura 7.** Comportamento das Distâncias de Mahalanobis calculadas ao longo do tempo para um caso de rompimento da linha B de injeção de gás.....22
- Figura 8.** Matriz de Confusão do método de detecção por Distância de Mahalanobis para os dados das bases: a) treino; e b) teste..... 23
- Figura 9.** Comportamento da variável PIC-SPEED – PV diante de desvios de grande magnitude no processo.....24
- Figura 10.** Matriz de Confusão de treino (a) e teste (b) dos resultados para o melhor resultado do método de detecção para a rede neural de 10 neurônios.....25
- Figura 11.** Matriz de Confusão dos conjuntos de treino (a) e teste (b) dos resultados para o melhor resultado do método de detecção para a rede neural de 5 neurônios.....25

LISTA DE TABELAS

Tabela 1. Variáveis utilizadas no estudo.....	11
Tabela 2. Lista dos casos de desvios de processo simulados no estudo.....	14
Tabela 3. Lista dos casos de rompimento simulados no estudo.....	14
Tabela 4. Resultados dos tempos de detecção dos casos de rompimentos para diferentes tamanhos de amostra (n) de dados coletados continuamente do processo.....	26

SUMÁRIO

1	Introdução	1
2	Revisão Bibliográfica	3
2.1	Métodos de Detecção de Vazamentos em Tubulações	3
2.1.1	Medições Acústicas	3
2.1.2	Modelos de escoamento	4
2.1.3	Análise de transientes	5
2.2	Distância de Mahalanobis	7
3	Metodologia	10
3.1	Descrição do processo de compressão e injeção de gás	10
3.2	Seleção das variáveis de processo	11
3.3	Coleta de dados de simulação	12
3.4	Sensibilidade do sistema	13
3.5	Método de Detecção a partir da Distância de Mahalanobis	14
3.6	Método de detecção por redes neurais	16
3.6.1	Construção da rede neural e treinamento	17
4	Resultados	18
4.1	Análise do comportamento do sistema	18
4.2	Detecção do rompimento a partir da Distância de Mahalanobis	21
4.3	Detecção do rompimento utilizando Redes Neurais	24
4.4	Redução dos tempos de detecção dos rompimentos	26
5	Conclusões e Trabalhos Futuros	28
	REFERÊNCIAS	29
	APÊNDICE A	31
	APÊNDICE B	34
	APÊNDICE C	35

1 Introdução

Devido às distâncias entre os locais de produção e consumo de parte dos fluidos utilizados pela sociedade, o transporte de produtos como água potável, gás natural, petróleo e seus derivados ocorre através do escoamento desses fluidos por tubulações.

Uma das oportunidades de melhoria durante a operação desses sistemas, principalmente em linhas com grandes extensões, é a de redução de perdas provocadas por vazamentos. Estima-se que essas perdas variem entre 6,5% e 24,6% em redes de distribuição de água na Europa (Ghazali et al., 2012). Além da dificuldade imposta pela escala de grandes redes de distribuição, o problema da falta de monitoramento de vazamentos que acontecem sob o solo ou no fundo do mar não é capaz de ser facilmente solucionado, por exemplo, com rotinas de inspeção desses equipamentos. Dessa forma, restringe-se a capacidade de identificar a presença desses modos de falha ao acompanhamento do processo por parte de equipes de Engenharia.

Sob a perspectiva desse problema, operações de produção de petróleo têm apresentado desafios durante a consolidação de novas tecnologias. Nos últimos anos, tem-se implementado uma alternativa para o destino de correntes gasosas de baixo interesse econômico que promove a compressão e o envio dos gases de volta ao reservatório. Tal processo evita que o gás seja queimado em *flare* após a separação das correntes principais e ainda promove a pressurização do reservatório para o aumento da sua capacidade de produção (Zain et al., 2001).

O Relatório de Segurança de Número 1 emitido pela Agência Nacional do Petróleo, responsável pelo segmento de exploração e produção de petróleo, tornou pública a informação a respeito do incidente de rompimento de um *riser* flexível, utilizado para o transporte dos gases até o reservatório em unidades *off-shore*. No relatório, atribui-se o fenômeno de Corrosão sob Tensão por CO₂ (em inglês, *CO₂ Stress Corrosion Cracking – SCC*) como causa para o rompimento completo da tubulação (ANP, 2017). Devido à presença de CO₂ na corrente gasosa, aliado a condições de alta pressão impostas pela operação em águas profundas, os *risers* têm rompido inesperadamente em operações com somente 10% do tempo de vida útil projetado para o duto.

Tais eventos provocam perdas de produção em grandes proporções, além de consequências ambientais pelo vazamento de gases como metano e CO₂, que podem provocar alterações em propriedades químicas como o pH da água e afetar o ambiente aquático nas regiões próximas ao local do acidente (Mohammadian et al., 2019).

O cálculo da Distância de Mahalanobis possui como vantagem a utilização das variáveis do próprio processo para o monitoramento do sistema, sem a necessidade da instalação de instrumentos específicos desenvolvidos por métodos de detecção acústicos, ou do uso de uma matemática complexa, em estudos de construção de modelos de balanço de massa e energia. Além disso, a possibilidade da escolha de quaisquer variáveis do processo para o cálculo do método, independentes das suas ordens de grandeza, permite a seleção dos dados que melhor diferenciem as situações de rompimento dos casos de distúrbios no processo.

Tendo em vista a complexidade da constatação de vazamentos em sistemas submarinos, aliado à necessidade da rápida identificação do evento durante o monitoramento do processo, o presente trabalho tem por objetivo propor um método de fácil implementação para a detecção dos rompimentos em tubulações de transporte de gases, que classifique as situações de rompimento dos distúrbios de processo em um curto espaço de tempo entre a ocorrência do evento e sua identificação.

Para atingir o objetivo geral do trabalho, faz-se necessário:

- Gerar um banco de dados de referência a partir de uma simulação que represente o processo estudado para a avaliação dos métodos de detecção. Dessa forma, é possível mapear os efeitos dos desvios de processo estudados e que não estão disponíveis em bases de dados do processo real.
- Utilizar a Distância de Mahalanobis como método para a detecção do rompimento de tubulações em unidades de compressão e injeção de gás.
- Comparar os resultados obtidos pela Distância de Mahalanobis com um algoritmo de reconhecimento de padrões utilizando Redes Neurais Artificiais.
- Analisar o efeito do tamanho de amostra utilizada para o monitoramento contínuo do processo.

2 Revisão Bibliográfica

Este capítulo traz uma revisão das diferentes abordagens disponíveis na literatura sobre o problema de detecção de vazamentos em tubulações de transporte de fluidos. Em diferentes escalas e aplicações, mantém-se o interesse de evitar perdas durante o escoamento dos mais variados componentes.

2.1 Métodos de Detecção de Vazamentos em Tubulações

A imensa variedade de trabalhos que buscam evitar a ocorrência prolongada de vazamentos pode ser reunida em dois grandes grupos quanto à necessidade de implementação de infraestrutura própria para a detecção da falha ou não: métodos de mudanças em *hardware* e *software*. Ainda, dentro desses dois grupos, os métodos podem ser classificados a partir dos diferentes fenômenos provocados por um vazamento. Exemplos desses subgrupos são os métodos de detecção acústicos, de modelagem por balanços de massa e volume ou por transientes de tempo (RTTM), gradientes de pressão, de processamento de sinais, entre outros (Zhang et al., 2015). Esta seção traz uma revisão dos métodos comumente utilizados na literatura para a identificação de vazamentos em processos industriais.

2.1.1 Medições Acústicas

No instante da ocorrência de um vazamento, efeitos no escoamento do fluido ou em regiões próximas da tubulação podem ser percebidos através do monitoramento de consequências, como ondas de pressão e bolhas de gás provocadas pelo evento. A detecção de vazamentos em tubulações usando métodos acústicos tem sido motivo para importantes estudos devido à alta acurácia da sua detecção (Leighton e White, 2012).

A utilização de equipamentos capazes de detectar a propagação de ondas sonoras formadas por perturbações no escoamento, aliado ao uso de redes neurais que diferenciam os sinais em casos de vazamento e “não-vazamento”, foi objeto de estudo por Santos et al., (2014). No estudo, os autores instalaram um microfone dentro do vaso localizado à montante da tubulação onde a abertura de válvulas “*on/off*” simula a ruptura que dá início a um vazamento e utilizaram o sinal elétrico de três diferentes frequências para monitorar a ocorrência do evento. A partir do acompanhamento em tempo real do sinal, foi possível detectar o início e o tamanho do orifício do vazamento com base no aumento da amplitude do sinal coletado.

Já a vibração causada por um vazamento na tubulação pode ser uma das maneiras de detectar o fenômeno. No estudo sobre a identificação de vazamentos em linhas de transporte de água, Yang et al., (2008) propuseram um método que promovia a instalação de um

acelerômetro piezoelétrico na parede da tubulação para a captura de sinais de vibração causadas pelo problema no escoamento. As vibrações, por sua vez, são originadas do colapso das bolhas decorrentes da cavitação da água no vazamento, por consequência da mudança da velocidade e do aumento da turbulência no escoamento da fase líquida. Os autores ainda utilizaram a função de autocorrelação para descrever as características dos sinais do vazamento e um algoritmo de entropia para a extração dessas características. Por sua vez, os valores desses sinais, correspondentes aos casos de vazamento e não vazamento, foram usados como entradas da Rede Neural de Elman, escolhida para detecção do problema.

2.1.2 Modelos de escoamento

Um grupo expressivo de trabalhos relacionados à identificação de falhas em sistemas hidráulicos se baseia na construção de modelos de escoamento. Diante da alteração das variáveis medidas, esses modelos, capazes de representar os processos com similaridade, conseguem evidenciar a presença de um vazamento ao longo de uma tubulação e, algumas vezes, até mesmo identificar o local ele acontece (Bai et al., 2017).

Uma grande variedade de modelos é implementada para sistemas com água, contudo os mesmos modelos não podem ser aplicados para escoamento de gases devido à compressibilidade dos componentes. O modelo de escoamento implementado por Reddy et al., (2011) buscou compreender a complexidade de sistemas gasosos ao implementar um modelo capaz de determinar gradientes de vazão e pressão aproximando os dados estimados pelo modelo dos dados coletados do processo em alguns pontos ao longo da linha. No trabalho, os autores utilizaram equações de conservação de massa e quantidade de momento para o escoamento subsônico de gases em uma tubulação rígida de diâmetro constante. Além disso, a determinação de parâmetros como a inclinação da tubulação e o cálculo da compressibilidade do gás tiveram de ser realizados devido à escolha das equações para a construção do modelo.

A transformada de Laplace foi utilizada na conversão das equações diferenciais do modelo com o objetivo de resolvê-las analiticamente e reduzir o custo computacional exigido por métodos numéricos durante seu uso em tempo real para o monitoramento do *status* do processo. A detecção do vazamento utiliza a abordagem de janela de tempo móvel para a coleta de dados e estimação do estado do sistema durante sua operação, além do procedimento *Generalized Likelihood Ratio* (GLR) para a comparação entre os resultados. Dessa forma, quando a função objetivo, que define a proximidade entre os dados coletados do processo e as variáveis calculadas pelo modelo, aumenta e atinge um valor acima do limite pré-estabelecido, identifica-se a presença de um vazamento.

Outra abordagem para a identificação das falhas em tubulações, aplicada aos casos de operação em águas profundas, foi proposta por Kam, (2010) no escoamento de misturas envolvendo fases líquida e gasosa. Inicialmente, a tubulação foi dividida em trechos de linha de comprimento arbitrário. Entre dois trechos, “Nós” foram utilizados como referência para

os locais de cálculo das propriedades do sistema. Assim, para o caso da pressão, conhecidos os parâmetros e o valor da variável em um dado nó da tubulação, foi possível determinar a pressão dos nós adjacentes a partir do seu valor no ponto anterior (ou posterior) mais a perda de carga calculada para o trecho de linha entre esses dois pontos.

Após o cálculo para a construção dos perfis de pressão e vazão, o autor manteve constantes a vazão de entrada e a pressão na saída da tubulação, além da compressibilidade dos componentes. Dessa forma, como mostrado na Figura 1, o vazamento pôde ser detectado ao se identificar a: 1) Redução da vazão total na saída da tubulação; e 2) Redução da pressão na entrada da tubulação. Na última, a redução da pressão é consequência da menor perda de carga nos trechos de linha à jusante do vazamento pela redução da vazão causada pelo evento. É possível ainda estimar o tamanho dos orifícios ou os locais do vazamento a partir da comparação entre os resultados do modelo (pressão na entrada da tubulação) para os parâmetros definidos e os valores obtidos com dados do processo diante de um caso real de perda de contingência.

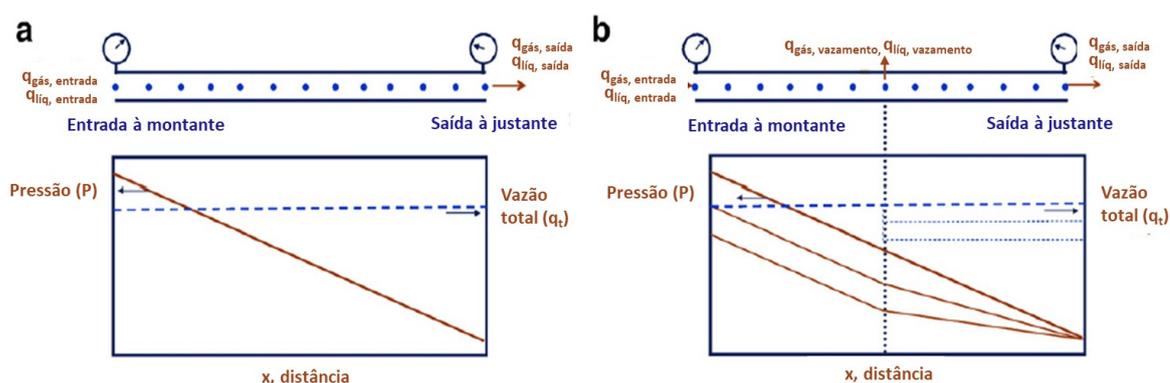


Figura 1. Desenho esquemático da tubulação com as respostas da pressão (P) e vazão total (q_t) para: (a) Estado estacionário antes do vazamento; e (b) Estado estacionário depois do vazamento.

Fonte: Adaptado de Kam, (2010)

2.1.3 Análise de transientes

Os métodos de detecção de vazamentos que fazem uso de instrumentos, alguns inclusive já instalados no processo, para a medição das condições de escoamento e que monitoram as alterações nas variáveis de processo durante sua operação são classificados dentro do grupo de análise de transientes (Colombo et al., 2009).

O acompanhamento da resposta de variáveis do escoamento diante de alterações planejadas em variáveis manipuladas é uma das alternativas para a identificação de perdas ao longo de uma tubulação. Nesse sentido, Warda et al., (2004) propuseram uma abordagem na

qual, a partir de um arranjo de válvulas e medidores instalados ao longo de uma tubulação utilizada para os experimentos, o fechamento controlado das válvulas localizadas à montante ou à jusante das tubulações, com conseqüente redução da vazão de água (20 – 80 %), são capazes de determinar a ocorrência e a localização de vazamentos por meio da análise dos diferentes comportamentos dos transientes de pressão.

Na Figura 2, é possível perceber a diferença do comportamento do transiente de pressão, causado pelo fechamento da válvula à jusante da tubulação, a depender da magnitude do vazamento (q/Q), onde q é a vazão de água que sai através do orifício e Q é a vazão de água que escoava no trecho de tubulação à montante do vazamento. Por fim, os autores ainda construíram um modelo capaz de simular o comportamento do sistema na presença de um vazamento a fim de comparar seus resultados com dados reais do processo.

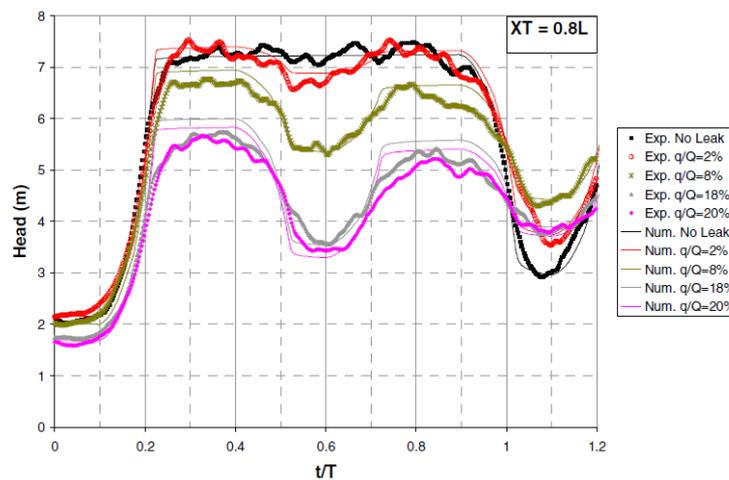


Figura 2. Comportamento transiente da pressão em um ponto de medição ($XT = 0,8L$) ao longo da tubulação para os casos sem vazamento (preto) e com vazamentos de diferentes magnitudes (cor).

Fonte: (Warda et al., 2004).

A ruptura repentina de uma tubulação provoca o surgimento de uma onda de pressão negativa que se propaga nos dois sentidos do escoamento. A identificação dessas ondas através da medição dos sinais de pressão por instrumentos na linha foi o objeto do estudo desenvolvido por Misiunas et al., (2005). Utilizando o algoritmo de teste de Soma Cumulativa (CUSUM), foi possível identificar o início do vazamento em linhas de distribuição de água subterrâneas. A partir da filtragem dos dados pelo algoritmo de Mínimos Quadrados Recursivos (RLS), evitando que ruídos dificultem o trabalho de detecção, o método CUSUM é utilizado para identificar quaisquer alterações que surgirem nos dados coletados através da soma acumulada dos desvios caso eles excedam um valor limite determinado para a operação normal do sistema entre dois instantes de tempo conseqüentes.

Além da identificação do vazamento, os autores ainda exploraram outra dificuldade durante a operação de linhas de água subterrâneas: a descoberta do local onde se encontram as fraturas na tubulação para que manutenções possam acontecer. Através da determinação dos tempos em que ocorrem três mudanças no valor da pressão em um mesmo ponto da tubulação, causadas por uma onda originada no início do evento e duas ondas refletidas em cada um dos limites do sistema, e da velocidade de propagação das ondas de pressão no fluido em questão, foi possível estimar a proximidade dos locais dos rompimentos estudados.

A detecção de vazamentos em linhas de produção de petróleo também foi motivo de estudo dentro do grupo de métodos de análise de transientes. Utilizando-se da infraestrutura de veículos submarinos não-tripulados necessária para a aquisição de sinais de pressão da linha subaquática, além da compressão e tratamento dos sinais, (Zhao et al., 2019) sugeriram um método a partir da análise de mudanças abruptas e picos de pressão na sucção da bomba de alimentação de óleo e na entrada e na saída da tubulação. Esses sinais, filtrados dos distúrbios causados por condições marítimas adversas, permitiram a detecção dos vazamentos de óleo através do acompanhamento dos picos dos sinais em *trends* das variáveis analisadas.

Em problemas relacionados a redes de distribuição de gás, Zuo et al., (2015) estudaram uma série de linhas de gás natural implantadas em solo Chinês com o objetivo de definir valores-limite para o fechamento completo de válvulas de controle do sistema diante da ruptura de um trecho da tubulação. Os autores escolheram a taxa de despressurização da tubulação como variável para a detecção da falha. Dessa maneira, o cálculo da taxa comparou as pressões em um local da linha entre dois instantes de tempo iguais a 60 segundos. Para a detecção do rompimento e fechamento da válvula de estanqueidade do sistema, o estudo considerou a permanência de taxas de despressurização acima de um valor recomendado para cada válvula do sistema durante, pelo menos, 120 segundos.

A recomendação dos valores limítrofes para o fechamento de cada válvula em casos de despressurização ocorreu dependendo da pressão de operação de cada trecho de linha estudado e da velocidade de escoamento do gás na tubulação. Em ambos os casos, quanto maior são seus valores, maiores são as taxas de despressurização.

2.2 Distância de Mahalanobis

A Distância de Mahalanobis, formulada pelo matemático indiano Prasanta Chandra Mahalanobis, é um método estatístico de medição da proximidade entre duas amostras distintas (Mahalanobis, 1936). Em outras palavras, a Distância atribui um valor à diferença entre os dados de uma amostra em relação à outra. A técnica é utilizada para o reconhecimento de padrões e a *clusterização* de bancos de dados.

Por definição, para duas amostras genéricas $A = \{x_{i1}, y_{i1}; x_{i2}, y_{i2}; \dots\}$ e $B = \{x_{j1}, y_{j1}; x_{j2}, y_{j2}; \dots\}$ com duas variáveis x e y , a Distância de Mahalanobis (D) pode ser calculada através da equação 1:

$$D = \sqrt{(\mu_x - \mu_y)^T V^{-1} (\mu_x - \mu_y)} \quad (1)$$

onde V^{-1} é a inversa da matriz de covariância conjunta das amostras, μ_x é a diferença das médias aritméticas da variável x das duas matrizes (Equação 2), enquanto μ_y representa a mesma diferença aplicada para a variável y . O sufixo T representa a matriz transposta (McLachlan, 1999).

$$\mu_x = \text{média}(x_{i1}, x_{i2}, \dots) - \text{média}(x_{j1}, x_{j2}, \dots) \quad (2)$$

Por sua vez, a matriz de covariância conjunta (V) é calculada através da média ponderada, pelo número de linhas das matrizes A e B respectivamente, de cada um dos elementos das matrizes de covariância individuais.

A matriz de covariância para a amostra A é calculada a partir das covariâncias possíveis entre variáveis:

$$V_A = \begin{bmatrix} \text{cov}(x_i, x_i) & \text{cov}(x_i, y_i) \\ \text{cov}(y_i, x_i) & \text{cov}(y_i, y_i) \end{bmatrix} \quad (3)$$

O mesmo cálculo é feito para a matriz de covariância aplicado à amostra B .

A matriz de covariância representa as relações que as variáveis possuem com elas mesmas e as demais dentro de uma mesma amostra. Dessa maneira, a diagonal principal é determinada pela variação de x e de y em relação a elas mesmas, enquanto os demais elementos são as covariâncias cruzadas entre as variáveis. As matrizes de covariância são simétricas, pois o valor da covariância de uma variável x com relação à y é o mesmo da covariância de y em relação à x .

Se não forem consideradas as covariâncias cruzadas das variáveis da amostra, o cálculo se resume à Matriz Euclidiana Normalizada, que é representada por uma matriz de valores,

diferentes de zero apenas na diagonal principal, das variâncias de cada uma das variáveis das amostras.

A metodologia descrita para amostras com apenas duas variáveis pode ser expandida para casos multivariáveis. Assim é possível calcular as covariâncias entre todas as F variáveis presentes desde que ambas as amostras estudadas possuam o mesmo número de variáveis (colunas da matriz). No entanto, não há a necessidade de que as amostras possuam o mesmo número de pontos estudados (linhas da matriz). Na prática, isso significa ser possível calcular a distância de Mahalanobis entre duas amostras com tamanhos distintos ou ainda utilizar um banco de dados, com dados da operação normal do processo, para determinar sua proximidade com uma amostra desconhecida de dados da operação.

A Distância de Mahalanobis captura de maneira eficiente a estrutura de correlação em dados de alta dimensão, o que a torna particularmente importante em conjuntos de dados com muitas dimensões. Esta abordagem simples apresenta vantagens surpreendentes sobre métodos mais complexos baseados em distância em termos de precisão e complexidade computacional (Aggarwal, 2017). Caso fossem analisadas em gráficos, as correlações entre as variáveis seriam limitadas a um espaço F -dimensional igual a 3. O método ainda é capaz de normalizar os dados, permitindo que se utilizem diferentes ordens de grandeza e conferindo-lhe versatilidade para escolha do grupo de variáveis que melhor representem as particularidades dos diferentes grupos de cada uma das amostras (Ferreira, 2020).

Um estudo relacionado à identificação de falhas de equipamentos em processos industriais usando métodos estatísticos. Zanolli e Astolfi, (2012) desenvolveram um estudo utilizando distâncias geométricas, aplicadas para o reconhecimento de padrões e *clusterização* de grupos de amostras, para a detecção de quatro diferentes modos de falha em compressores centrífugos de uma planta industrial. No trabalho os autores utilizaram a Distância de Mahalanobis para aperfeiçoar o classificador de falhas já utilizado em um estudo anterior. Em relação ao antigo método, baseado no cálculo da Distância Euclidiana, foi possível classificar as diferentes falhas com uma maior probabilidade de acerto.

Além disso, um menor número de dados foi necessário para o reconhecimento do problema, reduzindo de 3 h para 40 min o tempo entre a identificação da falha e o início dos sintomas apresentados pela máquina em relação ao resultado obtido pela Distância Euclidiana. A conclusão dos autores é de que, com a utilização da distância de Mahalanobis, são consideradas as correlações (covariâncias) entre os dados dentro de um mesmo grupo que representa um modo de falha, tornando mais robusta a capacidade de classificação das falhas pela abordagem proposta.

3 Metodologia

Este capítulo tem por objetivo apresentar o processo de compressão e injeção de gás, além de descrever os procedimentos necessários para o desenvolvimento do trabalho. Tais pontos serão apresentados na seguinte ordem: Descrição do processo de compressão e injeção de gás, Seleção das variáveis de processo, Coleta de dados de simulação, Sensibilidade do Sistema e os métodos de detecção a partir da Distância de Mahalanobis e por Redes Neurais.

3.1 Descrição do processo de compressão e injeção de gás

Instalada na própria unidade *off-shore*, a planta de compressão e injeção de gás em reservatórios tem por objetivo comprimir uma corrente gasosa, composta majoritariamente por metano, dióxido de carbono e etano, e retorná-la ao reservatório promovendo um destino alternativo à queima dessa corrente em *flare* ou emissão de CO₂. A Figura 3 representa o diagrama esquemático de uma planta de compressão e injeção de gás como a utilizada no estudo.

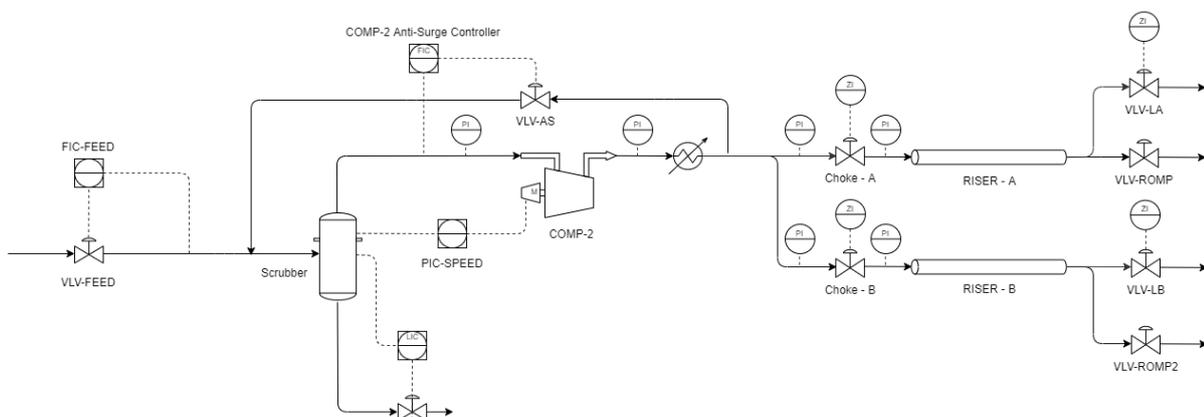


Figura 3. Diagrama simplificado de uma planta de compressão e injeção de gás.

A corrente gasosa oriunda das unidades de separação chega até o sistema através da válvula de alimentação de gás VLV-FEED. A válvula é responsável por controlar a vazão volumétrica de gás, em STD m³/dia, que entra na planta de compressão através do ajuste do *setpoint* do controlador FIC-FEED.

O vaso pulmão “Scrubber”, também chamado de Demister ou Depurador, tem a finalidade de remover gotículas de condensado potencialmente arrastadas nos processos à montante

do compressor. Nele, um medidor de pressão está instalado e o sinal desse instrumento representa a variável PIC-SPEED.PV, que pode ser acompanhada durante a operação da planta. Ainda, essa variável de processo é utilizada para o controle da velocidade de rotação do compressor COMP-2, que realiza a compressão do gás destinado à injeção, através do ajuste do *setpoint* do seu controlador (PIC-SPEED.SP).

A planta também possui um reciclo de gás para o compressor com controle de vazão realizado pelo controlador “COMP-2 Anti-Surge Controller”, que opera em malha fechada e controla a abertura da válvula do reciclo (VLV-AS) a partir da vazão de gás na sucção do compressor de maneira a evitar a ocorrência de Surge, com consequentes vibrações excessivas, em compressores.

Após o sistema de compressão, o gás pode ser distribuído para duas linhas de injeção denominadas linhas A e B. Cada linha possui uma válvula *choke*, a qual o operador controla a abertura para o ajuste de vazão, instalada à montante de *risers* flexíveis, que são comumente utilizados em operações de produção de petróleo, e tubulações rígidas no fundo do mar. Ao fim da tubulação, no local da perfuração do poço de petróleo, as válvulas VLV-LA e VLV-LB representam o fim do processo para cada uma das respectivas linhas ao permitir que todo o fluxo de gás seja “reinjetado” no reservatório.

3.2 Seleção das variáveis de processo

Para o melhor entendimento do processo e seu comportamento diante das principais alterações durante a operação da planta, um conjunto de variáveis foi escolhido para a análise dinâmica do processo e o desenvolvimento dos estudos de detecção de rompimentos dos *risers* flexíveis da unidade. A escolha buscou incluir os principais equipamentos do processo, instrumentos de medição e controladores. A Tabela 1 lista as variáveis utilizadas para o estudo.

Tabela 1. Variáveis utilizadas no estudo.

Variável	Unidade	Descrição
FIC-FEED - SP	STD_m ³ /dia	Setpoint do controlador da vazão de alimentação de gás
FIC-FEED - PV	STD_m ³ /dia	Vazão da alimentação de gás
FIC-FEED - OP	%	Output do controlador e posição de abertura da válvula VLV-FEED
PIC-SPEED - SP	barg	Setpoint do controlador da velocidade de rotação do compressor centrífugo
PIC-SPEED - PV	barg	Pressão do vaso Scrubber
PIC-SPEED - OP	%	Output do controlador da velocidade de rotação do compressor centrífugo

11 - P	barg	Pressão na sucção do compressor centrífugo COMP-2
12 - P	barg	Pressão na descarga do compressor centrífugo COMP-2
COMP-2 Anti-Surge Controller - SP	m ³ /h	Setpoint do controlador da vazão do reciclo de gás do compressor
COMP-2 Anti-Surge Controller - PV	m ³ /h	Vazão de gás do compressor centrífugo INJ-2
COMP-2 Anti-Surge Controller - OP	%	Output do controlador e posição de abertura da válvula VLV-AS
Choke-A-PD	barg	Perda de carga da válvula choke da linha A
Choke-A-AP	%	Posição de abertura da válvula choke da linha A
Choke-B-PD	barg	Perda de carga da válvula choke da linha B
Choke-B-AP	%	Posição de abertura da válvula choke da linha B
20 - Pressure	barg	Pressão à jusante da válvula choke da linha A
21 - Pressure	barg	Pressão à jusante da válvula choke da linha B
VLV-LA - ZI	%	Posição de abertura da válvula de injeção de gás da linha A
VLV-LB - ZI	%	Posição de abertura da válvula de injeção de gás da linha B

3.3 Coleta de dados de simulação

Para o desenvolvimento do estudo, foi necessário um mapeamento extenso das condições operacionais da unidade, contudo, a falta de um banco de dados impossibilitou a coleta de informações do sistema real para os distúrbios de processo avaliados no estudo. Além disso, eventos de rompimento como os estudados no trabalho são extremamente raros industrialmente para que fosse possível construir uma base de dados com suas ocorrências. Assim, a obtenção dos dados foi realizada utilizando uma simulação da planta de compressão e injeção de gás implementada e validada pela equipe de uma empresa do setor a partir de dados reais do processo utilizando o software Aspen HYSYS.

Para as simulações dos estudos de caso, utilizou-se a Versão 11 do software Aspen HYSYS. O modelo GERG-2008 foi escolhido como pacote termodinâmico. Parâmetros necessários para a simulação dinâmica como o tempo total da simulação e o passo da Integração foram selecionados na aba *Integrator*. Nela, definiu-se os tempos do início e do fim da simulação sendo iguais a 0 e 200 min respectivamente. O Passo da Integração foi definido em 0,5 s. A coleta de dados utilizou a ferramenta *Stripchart* para a obtenção dos resultados a cada 1 s de tempo real do processo para as variáveis selecionadas anteriormente. O tempo de execução computacional para cada simulação foi igual a 25 min.

Durante os primeiros 10 min de cada estudo de caso, manteve-se inalterada quaisquer variáveis para garantir a condição de estado estacionário antes que fossem realizadas as alterações no processo. Dessa forma, foi possível isolar os efeitos dos desvios estudados neste trabalho.

Após os 10 minutos de simulação, realizaram-se alterações no processo para a coleta e análise das variáveis selecionadas para o estudo do sistema. Os resultados das simulações foram salvos em arquivos Excel, onde foi possível manipular os dados obtidos e convertê-los em gráficos para entendimento da dinâmica do sistema em cada um dos casos analisados. Além disso, os resultados dos cenários de rompimento e de distúrbios no processo também foram utilizados para a verificação da eficácia dos algoritmos estudados.

Para o estudo do caso do rompimento da tubulação, a simulação possui as válvulas VLV-ROMP e VLV-ROMP2, com Coeficientes de Vazão (Cv) iguais ou superiores a 10^5 US GPM, localizadas no fim de cada linha de injeção de gás. A abertura dessas válvulas com um Cv tão alto é capaz de representar um vazamento de grandes proporções como no caso do rompimento completo da linha de injeção de gás. A escolha do local da válvula que simula o rompimento foi conservativa já que, quão maior for a profundidade do vazamento, mais difícil é sua detecção em razão da contrapressão externa no sistema (Kam, 2010).

3.4 Sensibilidade do sistema

Fez-se necessário realizar testes de sensibilidade e obter as curvas de reação das variáveis estudadas para o entendimento do comportamento do sistema diante dos distúrbios ou alterações em variáveis manipuladas. Além disso, foram simulados rompimentos de diferentes proporções e os efeitos de cada um dos casos pôde ser estudado.

A determinação das possíveis alterações do processo ocorreu a partir da identificação dos principais controladores do sistema e dos distúrbios de processo que frequentemente afastam as variáveis dos seus respectivos *setpoints*.

Os desvios do processo estudados são:

1. *Setpoint* do controlador da válvula de alimentação de gás (FIC-FEED -SP);
2. *Setpoint* do controlador da velocidade do compressor de gás (PIC-SPEED - SP);
3. Abertura da válvula choke da linha B de injeção de gás (Choke-A - AP);
4. Fração molar de CO₂ na corrente de alimentação de gás do sistema;

Foram realizadas simulações com desvios positivos e negativos de degraus iguais a 1%, 5%, 10%, 20% e 50% em comparação aos dados de operação padrão. A lista completa com os casos simulados considerando a operação da planta com injeção de gás apenas pela linha B é apresentada na Tabela 2.

Tabela 2. Lista dos casos de desvios de processo simulados no estudo.

Desvio de processo	Unidade	Valor E.E.	Step 1 (1%)	Step 2 (5%)	Step 3 (10%)	Step 4 (20%)	Step 5 (50%)
FIC-FEED - SP (+)	STD m ³ /dia	3.300.000	3.333.000	3.465.000	3.630.000	3.960.000	4.950.000
FIC-FEED - SP (-)	STD m ³ /dia	3.300.000	3.267.000	3.135.000	2.970.000	2.640.000	1.650.000
PIC-SPEED - SP (+)	barg	200	202	210	220	240	-
PIC-SPEED - SP (-)	barg	200	198	190	180	160	-
Choke-B-PD (-)	%	100	99	95	90	80	50
Fração molar de CO ₂ (+)	-	0,370	0,374	0,389	0,407	0,444	0,556
Fração molar de CO ₂ (-)	-	0,370	0,367	0,352	0,333	0,296	0,185

Não foi possível simular os casos de desvios positivo e negativo de 50% do *setpoint* do controlador PIC-SPEED por limites de cálculo impostos pela simulação. Assim, o projeto de experimentos compreende um total de 33 casos de desvios do processo.

Os casos de rompimento foram simulados a partir de alterações na abertura da válvula VLV-ROMP2 para diferentes valores de Cv da válvula. Ambas as alterações buscaram reproduzir rompimentos de menor ou maior magnitudes. Também foram simulados casos de rompimento para estados de diferentes vazões de alimentação de gás na planta. A lista dos 10 casos simulados para o rompimento da linha B de injeção de gás é mostrada na Tabela 3:

Tabela 3. Lista dos casos de rompimento simulados no estudo.

Caso de Rompimento	Coef. De Vazão	Valor E.E.(%)	Unidade	Step 1	Step 2	Step 3	Step 4
Abertura VLV-ROMP2	10 ⁵	0	%	0,01	1	-	-
Abertura VLV-ROMP2	10 ⁶	0	%	0,01	1	100	-
Abertura 1% VLV-ROMP2 + Maior FIC-FEED	10 ⁶	3.300.000	m ³ /dia	3.333.000	3.465.000	3.630.000	3.960.000

Os dados obtidos com as simulações foram filtrados de maneira que a frequência dos resultados de cada variável fosse igual a um resultado a cada 10s. Os resultados das simulações para cada um dos diferentes desvios foram analisados, utilizando os recursos do Excel, com gráficos do comportamento das variáveis ao longo do tempo, além de gráficos de Espaço de Fase entre a variável de processo e o *output* dos controladores.

3.5 Método de Detecção a partir da Distância de Mahalanobis

Sob o olhar da detecção dinâmica de falhas, pode-se definir a Distância de Mahalanobis como a medida da proximidade entre os dados de uma primeira amostra conhecida, que pode representar um estado normal ou de falha de um processo, e uma segunda amostra

desconhecida de dados coletados durante a operação. Assim, caso ocorram desvios no processo inesperados, um grupo de variáveis pode ser estudado para a avaliação do grau de similaridade dos dados analisados em relação à uma base de dados de referência, por exemplo, da operação normal do processo.

A implementação da metodologia ao estudo da detecção dinâmica da falha do processo, requer o cálculo de duas Distâncias de Mahalanobis: 1) a Distância em relação à condição de operação normal do processo; e 2) a Distância em relação à condição de rompimento.

Para o cálculo de cada uma das Distâncias de Mahalanobis são necessárias duas amostras coletadas em diferentes instantes do processo. A primeira, utilizando dados de referência do processo (para casos de operação normal ou casos de rompimento) e a segunda com dados coletados a partir de uma janela móvel de operação. O algoritmo tem por objetivo determinar a similaridade da amostra continuamente coletada pelo processo com a amostra de referência utilizada.

Como resultados, obtêm-se os valores de duas distâncias que, quando comparados entre si, indicam a ocorrência de um rompimento caso o valor de (2) seja inferior do que (1). Intuitivamente, isso significa que os dados coletados do processo pela janela móvel possuem maior similaridade a uma condição de falha (rompimento) de que a uma condição de operação normal do processo.

A Figura 4 representa a sistemática utilizada para simular uma janela móvel de operação e permitir a abordagem dinâmica para o cálculo da Distância de Mahalanobis.

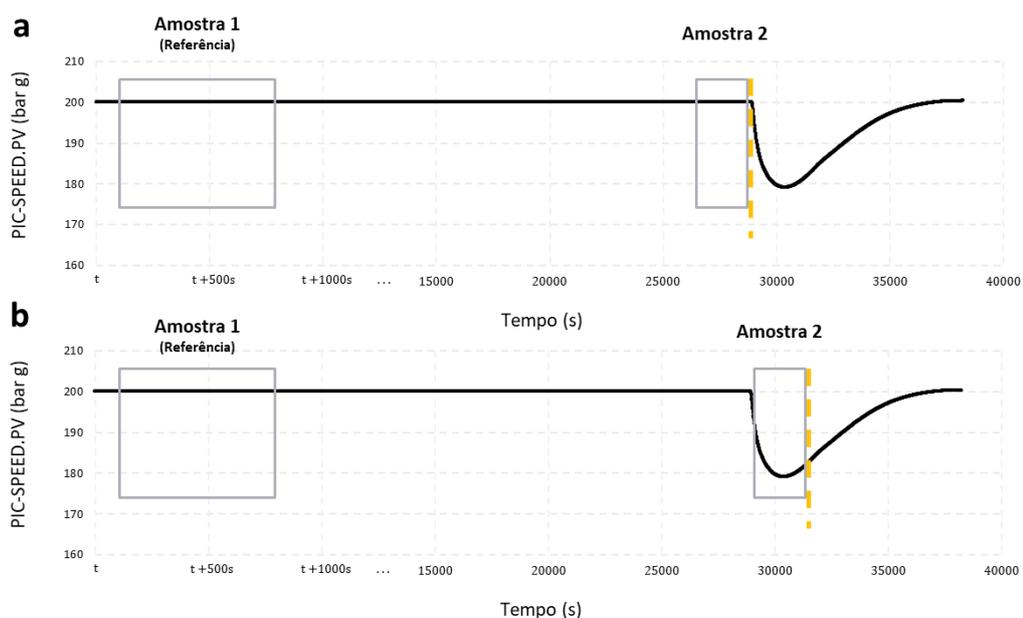


Figura 4. Representação hipotética das amostras coletadas para o cálculo da Distância de Mahalanobis nos instantes: a) Início do rompimento; e b) Durante a ocorrência do rompimento.

Na Figura 4a, instante no qual é representado o início do rompimento, a amostra 2 ainda não contém dados referentes à condição de falha. Já na Figura 4b, a janela móvel obtém, pela primeira vez, 100% dos dados da variável PIC-SPEEP – PV na condição de rompimento para o cálculo da distância de Mahalanobis. Entretanto, pode-se perceber que para a Amostra 1 são utilizados sempre os mesmos dados de referência para qualquer um dos instantes de tempo.

No estudo, foram utilizados os dados obtidos através das simulações para a construção das amostras de referência (Amostra 1). Essas amostras contém dados de 80% dos casos de desvios de processo, listados na Tabela 2, para a determinação da Distância de Mahalanobis em relação às condições normais de processo e dados de 80% dos casos de rompimento, listados na Tabela 3, para o cálculo da Distância de Mahalanobis em relação à condição de Rompimento da tubulação.

A definição dos 80% dos casos de desvio ou dos 80% dos casos de rompimento que compõem a base de dados utilizada como primeira amostra (base de dados de Treino) ocorreu de maneira aleatória. Os casos não selecionados, demais 20%, foram reservados para o teste do método quanto a capacidade de identificar verdadeiros positivos e negativos sem que seus dados estejam presentes na base de dados de Treino.

A Amostra 2 contém os dados de apenas um dos casos estudados de rompimento ou não-rompimento por vez, que são listados nas Tabelas 2 e 3. Para cada um dos casos, são obtidos vetores com os resultados das Distâncias de Mahalanobis em relação à base de operação normal e de rompimento ao longo do tempo.

Ainda, o algoritmo permite a escolha do tamanho da janela móvel de coleta de dados dos casos estudados para o cálculo das Distâncias. É possível escolher os “n” instantes de tempo usados a cada passo da iteração para o cálculo de apenas um valor de Distância de Mahalanobis dentre os demais valores pertencentes ao vetor dos resultados. Como, para o desenvolvimento do estudo, foram considerados dados coletados a cada 10 s, o tempo de amostragem para a amostra 2 representa dados de 10n segundos de operação. Dessa forma, o primeiro resultado obtido pelo algoritmo representa o cálculo da Distância para os dados entre o início da simulação do caso (0s) e o instante “t” para um n qualquer. O segundo ponto compreende o resultado para o cálculo dos dados entre o instante 10 s e o instante “t” para n + 1 e assim sucessivamente até o tempo final da simulação (200min).

O algoritmo foi implementado em *Python* e seu código se encontra no Apêndice A.

3.6 Método de detecção por redes neurais

Recentemente, métodos de detecção de falhas tem utilizado a inteligência artificial como importante ferramenta para o aumento de confiabilidade de processos industriais (Yang et al., 2008). Com o objetivo de reproduzir estruturas neurais de organismos inteligentes, as Redes

Neurais Artificiais (RNA) podem ser descritas como um conjunto de modelos matemáticos baseados em uma estrutura biológica que, através do seu treinamento a partir de uma base de dados, é capaz de adquirir conhecimento para a resolução de problemas complexos.

Os principais elementos de uma rede neural são os próprios neurônios, divididos na maioria dos casos em Neurônios de Entrada, Intermediários e de Saída. Esses elementos são responsáveis, respectivamente, por receber as informações de entrada para o treinamento da rede, interligar os neurônios de entrada e de saída dentro da chamada Camada Oculta e inferir as respostas para cada uma das possíveis classes de resposta permitidas pelo usuário na construção da rede neural (Fausett, 1994).

A maneira como são realizados os treinamentos de redes neurais segue uma rotina de atribuição de pontos aos neurônios envolvidos no processamento dos dados de entrada e classificação dos resultados na sua saída a cada erro ou acerto. O objetivo do treinamento é obter o caminho com o maior número de pontos atribuídos aos neurônios pelos acertos da rede neural na formação da resposta correta. Já o sucesso da classificação de dados de entrada desconhecidos acontece quando a rede neural é capaz de reconhecer padrões semelhantes entre os dados desconhecidos e os dados de treinamento e determinar a classe correta a qual pertencem esses valores.

3.6.1 Construção da rede neural e treinamento

A rede escolhida para o estudo foi uma rede neural simples do tipo de reconhecimento de padrões. Um algoritmo foi implementado utilizando o software Matlab (versão R2020a) para treinamento de 100 redes neurais constituídas de 10 neurônios intermediários do tipo Sigmóide em cada rede. O objetivo do treinamento de 100 redes neurais a partir de valores de inicialização aleatórios é de evitar que, durante o processo de convergência dos resultados, a solução de mínimos locais fosse obtida, reduzindo a capacidade de detecção dos rompimentos pelo método utilizado.

A convergência da rede neural para uma solução local manteria o resultado longe da melhor solução possível para o problema, dessa forma, os resultados obtidos foram aqueles apresentados pela rede neural que apresentou a melhor capacidade de classificar os dados de entrada dentre um total de 100 redes treinadas. As variáveis selecionadas para a detecção do rompimento utilizando a Distância de Mahalanobis foram utilizadas para a aplicação da solução por Redes Neurais. Também foi necessário selecionar os tamanhos dos conjuntos de treino, igual a 75%, e de validação e teste, iguais a 12,5% cada.

4 Resultados

Neste capítulo são apresentados os resultados obtidos a partir das análises do sistema necessárias para o entendimento da problemática, a proposta de detecção de rompimento para o caso estudado e seus resultados quando comparados com outro método de reconhecimento de padrões. Ainda, são avaliados os tempos necessários para a detecção dos rompimentos para diferentes tamanhos das amostras coletadas.

4.1 Análise do comportamento do sistema

Após as simulações dos casos para os desvios de processo e para os rompimentos de diferentes proporções da linha B de injeção de gás, os resultados obtidos foram plotados em gráficos do tipo “Variável de processo x tempo” para a análise do comportamento dinâmico do processo diante das alterações provocadas em cada desvio.

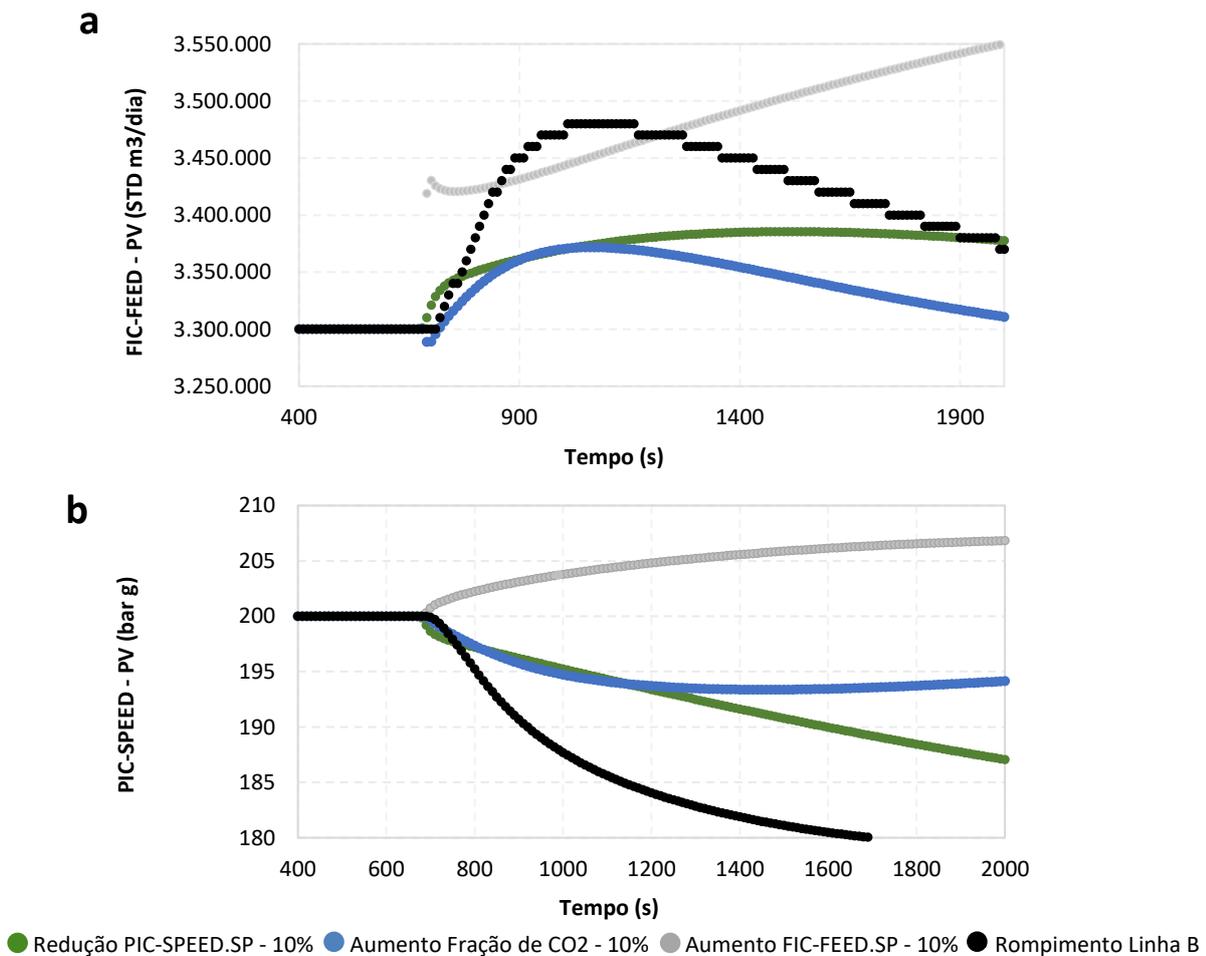


Figura 5. Comportamento dinâmico do sistema diante de desvios de processo iguais a 10% e de um caso rompimento da tubulação (1% de abertura e $C_v = 10^6$ US GPM) para as variáveis: a) FIC-FEED – PV; e b) PIC-SPEED – PV.

Os gráficos presentes na Figura 5 mostram as curvas de reação das variáveis de processo para dois dos principais controladores do sistema: a) Controlador da vazão da unidade através válvula de alimentação de gás do sistema (FIC-FEED); e b) Controlador da pressão de sucção da planta de reinjeção, através da velocidade de rotação do compressor COMP-2 (PIC-SPEED).

Os degraus que provocaram efeitos em magnitudes próximas de um caso de rompimento nas variáveis de processo foram: 5% de aumento no FIC-FEED - SP, 10% de redução no PIC-SPEED – SP e o distúrbio de aumento de 20% na fração molar de CO₂ na corrente de alimentação da planta. Os casos de fechamento da válvula *choke* pouco impactaram no comportamento do sistema, mesmo para o caso de redução de 50% na sua abertura. Acredita-se que seu ajuste isolado, não seja capaz de provocar grandes alterações no processo como os casos das alterações de *setpoint* dos controladores.

A partir da Figura 6a, pode-se perceber a resposta rápida da variável de processo FIC-FEED - PV, numa espécie de “salto” em relação ao valor no estado estacionário, quando os desvios ocorrem no mesmo controlador daquela variável que está sendo analisada. Esse fenômeno provavelmente ocorre por conta da ação proporcional do controlador. Entretanto, para uma situação de rompimento, é possível que os efeitos sejam percebidos com um tempo de atraso (tempo morto) em razão da distância entre a localização do *riser* no fundo do mar e do ponto de medição da variável FIC-FEED - PV na planta, além da quantidade de equipamentos (vasos, válvulas e trocador de calor) entre esses dois locais que amortecem os efeitos provocados pelo rompimento.

É possível perceber as diferentes assinaturas que cada variável apresentada ainda que diante de degraus, de iguais proporções, efetuados no processo. Contudo, em especial para a variável de processo analisada na Figura 5b, pode-se observar o comportamento inverso provocado pelo desvio “Aumento FIC-FEED – 10%” em relação aos demais casos.

Como resultado dessas análises, é possível sugerir que as diferentes assinaturas observadas de forma preliminar expressam a singularidade de cada um dos desvios, permitindo que esses possam ser classificados quanto a ocorrência, ou não, de falhas do processo já que diferem, além das suas magnitudes, de situações típicas de operação da unidade.

A análise de gráficos sem a dependência da variável temporal também pode fornecer informações importantes para a identificação dos casos que apresentam diferenças significativas entre as correlações das suas variáveis. Ademais, analisar duas variáveis que possuem relação direta entre si, como é o caso da variável de processo (PV) e da sua variável manipulada (MV), tornam bastante claras as diferentes reações de um sistema a desvios de processo. Dessa forma, os gráficos de Espaço de Fase podem oferecer a oportunidade de se avaliar a existência das diferentes correlações entre duas variáveis para diversos casos estudados.

A Figura 6 mostra os diferentes comportamentos das variáveis do controlador do compressor centrífugo de gás para os desvios estudados. Distanciando-se do ponto de origem em comum, 200barg para a pressão no vaso SG e 55,85% de *output* do controlador PIC-SPEED, as variáveis apresentam trajetórias distintas durante a tentativa de convergirem para novos estados estacionários, expressos pelos vórtices nos extremos de cada um dos resultados.

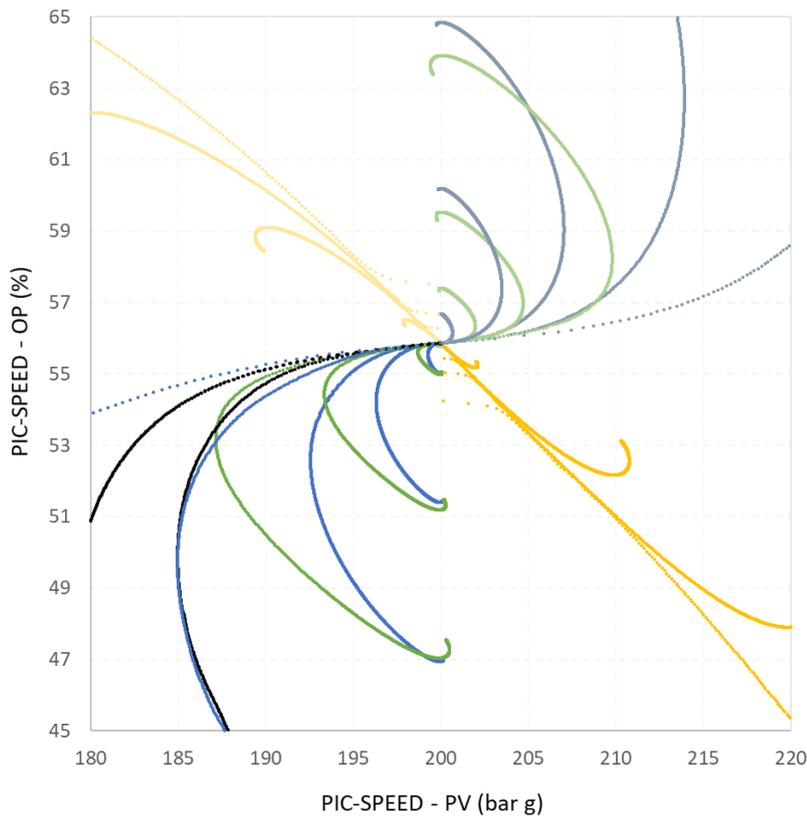


Figura 6. Espaço de Fase para a Variável de Processo e a Saída do controlador PIC-SPEED nos casos de desvio do FIC-FEED, do PIC-SPEED, na fração molar de CO₂ e nos casos de rompimento.

Enquanto para os casos de alterações no *setpoint* do controlador do compressor de gás (amarelo) os desvios são visivelmente distintos do caso de rompimento (preto), para os casos de redução no *setpoint* do controlador FIC-FEED (azul) e aumento da fração molar de CO₂ (verde) os trajetos que as variáveis apresentam durante o estado transiente são bastante próximos entre si, principalmente no início dos desvios.

Essas trajetórias de dados de 160min das variáveis em discrepantes direções e sentidos, configuram as diferentes “formas” dos desvios dentro de um gráfico de Espaço de Fase e reforçam a ideia de que, a partir da seleção adequada de variáveis do sistema, é possível

identificar ou até mesmo categorizar os diferentes casos de desvio e modos de falha de um processo dada a avaliação desses comportamentos.

A representação do Espaço de Fase para variáveis do controlador do compressor de gás na Figura 6 sugere que as demais variáveis também possuem diferentes correlações, de maior ou menor grau, entre si e que se traduzem nas singulares covariâncias das variáveis para as variadas amostras. Assim, faz-se necessária encontrar uma maneira de calcular as covariâncias entre as variáveis para que, através de uma medida, seja possível avaliar o distanciamento entre os casos estudados, além do desenvolvimento de um método capaz de identificar eventos desconhecidos e indesejados na maioria das vezes durante a operação da unidade.

4.2 Detecção do rompimento a partir da Distância de Mahalanobis

Diante das “impressões digitais” que cada um dos grupos de desvios apresenta em seus estados transientes e da percepção de que as diferentes correlações entre as variáveis possam fornecer informações importantes para a detecção de falhas no processo, foi proposto o cálculo da Distância de Mahalanobis na tentativa de detecção dos rompimentos das tubulações de gás subaquáticas.

As variáveis selecionadas para os cálculos da distância de Mahalanobis foram: 1) Vazão da alimentação de gás (FIC-FEED – PV); e 2) Saída do controlador da vazão de alimentação de gás (FIC-FEED – OP) por apresentarem as curvas de reação mais singulares para cada caso avaliado na Seção (4.1) dentre as variáveis analisadas no estudo. Construiu-se uma base com os dados para 80% dos casos de desvios de processo estudados e outra com 80% dos casos de rompimento.

Cada uma das duas bases, formada pelos dados referentes à condição de Operação Normal e à condição de Rompimento da tubulação, representa a Amostra 1 para o cálculo de duas Distância de Mahalanobis, aplicadas para a respectiva base de referência. Quando comparados os valores das duas Distâncias ao longo de uma janela móvel de operação, a expectativa é de que, diante da ocorrência de um rompimento, os valores dos resultados utilizando como referência os dados de Operação Normal apresentem um aumento – correspondendo à dissimilaridade entre a amostra de referência e a amostra analisada. O contrário é esperado para o valor do resultado da Distância para a base de dados de rompimento, indicando o aumento da proximidade entre a amostra de referência e a amostra analisada.

Escolheu-se um valor de n igual a 18 para o número de dados coletados em cada instante de tempo que compõem a segunda amostra necessária para o cálculo da Distância de Mahalanobis. Os dados representam uma janela de operação de 180 s e tal valor foi escolhido em razão da expectativa inicial para o tempo de detecção do rompimento.

O gatilho para a identificação de um rompimento acontece quando, em um certo instante de operação, a “Distância de Mahalanobis – Rompimento” possui um valor inferior ao resultado da “Distância de Mahalanobis – Operação Normal”, refletindo a maior proximidade dos dados de uma situação de falha do que de operação habitual do processo. O resultado para as duas métricas em um caso de detecção de rompimento (abertura de 1% da VLV-ROMP2 para um Cv igual a 10^6 US GPM) é evidenciado na Figura 7. Os valores de tempo negativos indicam instantes de operação anteriores ao rompimento.

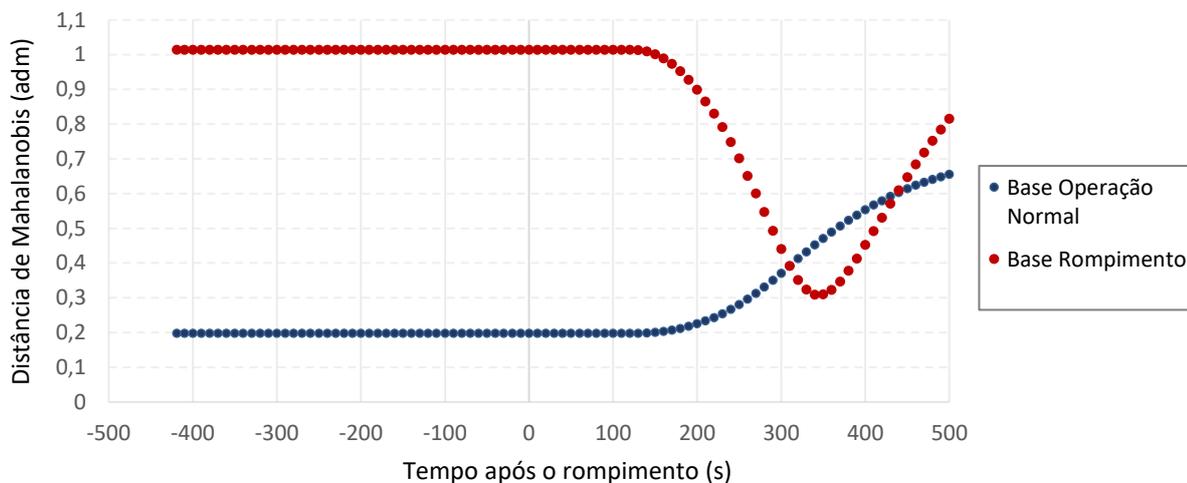


Figura 7. Comportamento das Distâncias de Mahalanobis calculadas ao longo do tempo para um caso de rompimento da linha B de injeção de gás.

Para o caso da Figura 7, foi possível identificar a ocorrência do evento 320 s após o início da falha. Com o método utilizado, a identificação do rompimento ocorre em torno de 5,3 min após o seu início. Isso representa a detecção antecipada da falha no processo, tempo no qual são tomadas medidas de contingência das consequências causadas pelo rompimento.

As matrizes de confusão utilizadas na avaliação do método de detecção proposto para os casos de Treino (80% dos casos utilizados como banco de dados) e os casos de Teste (demais 20%), são mostradas na Figura 8:

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	19 55,9%	7 20,6%
	Rompimento	1 2,9%	7 20,6%

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	5 55,6%	2 22,2%
	Rompimento	0 0%	2 22,2%

Figura 8. Matriz de Confusão do método de detecção por Distância de Mahalanobis para os dados das bases: a) treino; e b) teste.

Os resultados apresentados pelas Matrizes de Confusão foram gerados a partir de dados que representam 15min após o início dos desvios, ou seja, a oportunidade do método de identificar um rompimento dentro desse período para todos os casos estudados. Os resultados evidenciam a ocorrência de falsos negativos e de falsos positivos no método de detecção proposto. A acurácia do método é de 76,7% para os casos avaliados. A lista completa de cada um dos desvios e seus resultados estão presentes no Apêndice B.

O valor de falsos negativos igual a 2,9% representa o caso de um rompimento simulado pela abertura de 0,01% na válvula VLV-ROMP2 para um valor de Cv igual a 10^5 US GPM. Esse caso, que representa a falha de menor magnitude estudada, sugere que o método não é capaz de identificar pequenos vazamento presentes na linha. Por outro lado, não ocorreram casos de falsos negativos para os dados de Teste (b) indicando que o método é capaz de identificar rompimentos de magnitudes diferentes dos casos analisados no estudo, desde que esses ocorram em escala superior ao caso observado na matriz da base Treino.

Os resultados de falsos positivos iguais a 20,6% e 22,2% ocorreram principalmente em casos em que foram realizadas alterações de grande magnitude no processo (ver Apêndice B) e tais desvios, por apresentarem efeitos nas variáveis de processo próximos de um caso de rompimento, talvez não ocorram em uma rotina típica de operação da unidade. A Figura 9. Comportamento da variável PIC-SPEED – PV diante de desvios de grande magnitude no processo. evidencia os fortes efeitos que desvios de grandes magnitudes causam no processo, sendo, em alguns casos, efeitos superiores à ocorrência de um rompimento. Assim, mesmo que para o estudo sejam apontados como falsos positivos, é possível que esses desvios tenham uma baixa probabilidade de acontecer, sem que se tornem de fato uma preocupação durante a realidade das condições de operação.

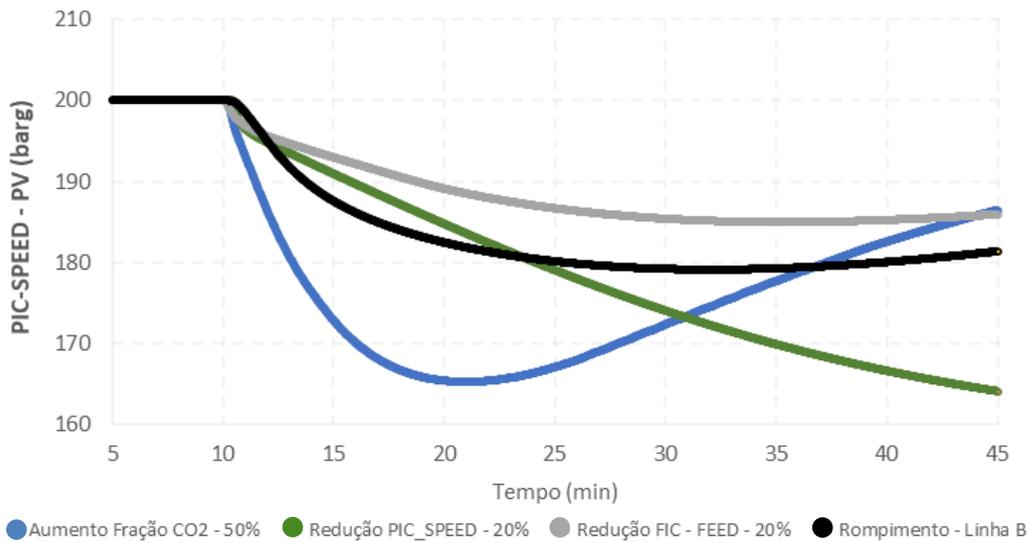


Figura 9. Comportamento da variável PIC-SPEED – PV diante de desvios de grande magnitude no processo.

4.3 Detecção do rompimento utilizando Redes Neurais

Com o objetivo de comparar o método de detecção proposto com outros métodos utilizados na literatura para o reconhecimento de padrões em amostras de dados, os resultados da utilização de redes neurais para a detecção dos casos de rompimento são apresentados a seguir.

Utilizaram-se 100 rede neurais, treinadas a partir de valores de inicialização aleatórios, com as variáveis FIC-FEED – PV e FIC-FEED – OP (também utilizadas para o método da Distância de Mahalanobis) para os casos de diferentes desvios e casos de rompimento com o único objetivo de classificar os casos de rompimento como sendo distintos dos casos de alterações ou distúrbios do processo.

Os resultados obtidos para a melhor rede neural treinada para os casos de rompimento e desvios de processo são apresentados na Figura 10. As amostras de dados coletadas para avaliação do método em cada um dos casos estudados representam os mesmos 15 min após o início do rompimento, os quais se espera detectar o rompimento, utilizados pelo método da Distância de Mahalanobis.

A partir dos resultados para a melhor rede neural treinada é possível concluir que o método possui 100% de acurácia na detecção dos rompimentos, sem a presença de falsos positivos que possam afetar a disponibilidade de operação da unidade, para quaisquer um dos desvios estudados.

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	25 80,6%	0 0,0%
	Rompimento	0 0,0%	6 19,4%

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	3 50,0%	0 0,0%
	Rompimento	0 0,0%	3 50,0%

Figura 10. Matriz de Confusão de treino (a) e teste (b) dos resultados para o melhor resultado do método de detecção para a rede neural de 10 neurônios.

A eficiência da detecção do método por redes neurais também foi avaliada para redes com um menor número de neurônios. Foram avaliados os mesmos casos treinando o mesmo grupo de 100 redes neurais com 5 e 2 neurônios. Os resultados obtidos são apresentados Figura 11:

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	25 80,6%	0 0,0%
	Rompimento	0 0,0%	6 19,4%

		PREDIÇÃO	
		Operação normal	Rompimento
REAL	Operação normal	4 66,7%	0 28,6%
	Rompimento	0 0,0%	2 33,3%

Figura 11. Matriz de Confusão dos conjuntos de treino (a) e teste (b) dos resultados para o melhor resultado do método de detecção para a rede neural de 5 neurônios.

Os mesmos resultados para a detecção usando redes neurais com 10 neurônios foram obtidos para a detecção usando redes neurais com apenas 2 neurônios. Esses resultados mostram a facilidade do método em detectar os rompimentos, além de diferenciá-los dos casos de operação normal presentes no estudo sem a ocorrência de falsos positivos. Os

resultados ainda sugerem a utilização do método como uma ferramenta confiável para a identificação do modo de falha estudado.

4.4 Redução dos tempos de detecção dos rompimentos

Além da distância física entre os pontos de medição e controle das variáveis utilizadas para o cálculo da Distância de Mahalanobis e do local do rompimento, outros parâmetros da detecção dinâmica, como a escolha do tamanho de amostragem da coleta contínua no tempo, permitem o atraso no tempo da identificação da falha no processo. Para uma coleta de 18 dados, o tempo entre o início e o instante da detecção do rompimento foi de 320s. Isso quer dizer que o resultado para o instante igual a 320 s contém dados entre os instantes 140 s e 320 s após o início do rompimento.

Na tentativa de reduzir os tempos de detecção para o método da Distância de Mahalanobis, foram calculadas novas Distâncias a partir da coleta de 12, 6 e 3 dados (n) que constituem a Amostra 2, a qual representa os dados coletados no processo através da janela móvel de tempo. Os resultados dos tempos de detecção para cada tamanho de amostragem são mostrados na Tabela 4.

Tabela 4. Resultados dos tempos de detecção dos casos de rompimentos para diferentes tamanhos de amostra (n) de dados coletados continuamente do processo.

Caso de Rompimento	Tempo de detecção (s)			
	$n - 18$	$n - 12$	$n - 6$	$n - 3$
Abertura 0,01% e $Cv = 10^5$	-	-	-	-
Abertura 0,01% e $Cv = 10^6$	320	272	242	222
Abertura 1% $Cv = 10^5$	310	262	232	212
Abertura 1% e $Cv = 10^6$	310	262	232	212
Abertura 100% e $Cv = 10^6$	310	262	232	222

A partir da Tabela 4 é possível observar a redução dos tempos de detecção dos rompimentos devido à redução no tamanho da Amostra 2 para os casos analisados. Os resultados da Distância de Mahalanobis ao longo do tempo para as diferentes amostragens para um mesmo caso de rompimento estão disponíveis no Apêndice C.

A avaliação permitiu a descoberta de uma oportunidade de melhoria no método de identificação do rompimento do ponto de vista da agilidade na sua identificação e redução de danos provocados pelo problema. Acredita-se que os ganhos obtidos com os resultados sejam decorrentes do aumento da sensibilidade do método para a condição de um menor número de instantes de tempo presentes na amostra, tornando os valores das Distâncias calculadas facilmente responsivos na presença dos primeiros desvios das variáveis.

Os valores absolutos das distâncias de Mahalanobis não se alteraram significativamente apesar dos diferentes tamanhos de amostragem. Apesar disso, devem ser mais bem avaliadas as possíveis consequências dessa redução como o aumento no número de falsos negativos que inviabilizariam sua utilização.

5 Conclusões e Trabalhos Futuros

O presente trabalho teve como objetivo propor um método de fácil implementação para a detecção dos rompimentos em tubulações de transporte de gases, que classifique de forma eficaz e rápida a ocorrência do evento. O método estudado identificou 9 dos 10 casos de rompimento estudados em um tempo médio de 4,3min após o seu início, permitindo concluir que ele é capaz de detectar o modo de falha que motivou o desenvolvimento desse trabalho.

O método é possível de ser implementado no sistema de segurança da unidade, servindo como variável monitorada continuamente para a atuação de alarmes à equipe de operação sobre a possível ocorrência de um rompimento. Contudo, a métrica apresentou taxas de falsos positivos entre 20% e 22%, que podem provocar paradas indevidas da unidade de compressão de gás caso a metodologia proposta seja utilizada em um sistema de *trip* da planta. Deve-se estimar a probabilidade em que os desvios de processo que resultaram em falsos positivos ocorrem durante uma rotina de operação e mensurar os riscos da implementação da ferramenta frente aos ganhos esperados pelo monitoramento do processo. Já a detecção por Redes Neurais é capaz de ser utilizada com ótima confiabilidade para a detecção dos rompimentos.

Ainda, os métodos de detecção pela Distância de Mahalanobis e por Redes Neurais são aplicados para casos quando a planta se encontra em plenas condições de operação e seu uso não deve ser englobado para a identificação de um rompimento durante a partida e a parada de uma planta de compressão de gás.

Diante da identificação de novos desvios ou distúrbios do processo não considerados no estudo, as bases de dados construídas como referência para o cálculo da Distância podem ser expandidas para que considerem as covariâncias entre as variáveis dos casos recentes, tornando o método cada vez mais robusto, sendo esse o primeiro trabalho futuro sugerido.

A implementação de outros métodos para a identificação de vazamentos em tubulações, já estudados pela literatura, ou de Redes Neurais, que geraram resultados promissores nesse estudo, podem ser alternativas à detecção usando a Distância de Mahalanobis, contudo devem ser avaliadas a viabilidade da implementação e o custo computacional das técnicas para o monitoramento on-line do processo.

Metodologias estatísticas utilizadas para o reconhecimento de padrões ainda são uma alternativa para a identificação de falhas em processos industriais e possuem grande espaço para o desenvolvimento de futuras soluções. Como abrangência desse trabalho, recomenda-se um estudo para a seleção ideal do grupo de variáveis necessárias para a obtenção de melhores resultados de detecção dos rompimentos utilizando a Distância de Mahalanobis de forma a promover a consolidação do método nessas aplicações, além do estudo de outras técnicas como Análise do Componente Principal (PCA), *Support Vector Machine* (SVM), etc.

REFERÊNCIAS

AGGARWAL, Charu C. *Outlier Analysis*. Springer, 2017. Second Edition. ISBN 978-3-319-47578-3 (eBook), DOI 10.1007/978-3-319-47578-3.

ANP. **Alerta de Segurança 001 - ANP/SSM: Corrosão sob Tensão por CO2 (SCC-CO2)**. Brasília: [s.n.]. Disponível em: <<https://www.gov.br/anp/pt-br/assuntos/exploracao-e-producao-de-oleo-e-gas/seguranca-operacional-e-meio-ambiente/incidentes/alertas-de-seguranca>>.

BAI, Y.; ZHANG, T.; LI, Y.; et al. A new leak detection method for subsea pipelines. **Ships and Offshore Structures**, v. 12, p. 144–152, 2017.

COLOMBO, A. F.; LEE, P.; KARNEY, B. W. A selective literature review of transient-based leak detection methods. **Journal of Hydro-Environment Research**, v. 2, n. 4, p. 212–227, 2009.

FAUSETT, L. *Fundamentals of Neural Networks: Architecture, Algorithms, and Applications*. 1994.

FERREIRA, J. **Deteção de indícios de fraudes no Programa Farmácia Popular do Brasil**. [s.l.] Universidade Federal do Rio Grande do Sul, 2020.

GHAZALI, M. F.; BECK, S. B. M.; SHUCKSMITH, J. D.; et al. Comparative study of instantaneous frequency based methods for leak detection in pipeline networks. **Mechanical Systems and Signal Processing**, v. 29, p. 187–200, 2012.

KAM, S. I. Mechanistic modeling of pipeline leak detection at fixed inlet rate. **Journal of Petroleum Science and Engineering**, v. 70, n. 3–4, p. 145–156, 2010.

LEIGHTON, T. G.; WHITE, P. R. Quantification of undersea gas leaks from carbon capture and storage facilities, from pipelines and from methane seeps, by their acoustic emissions. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 468, n. 2138, p. 485–510, 2012.

MAHALANOBIS, P. On the Generalized Distance in Statistics. **National Institute of Sciences of India**, v. 2, n. 1, p. 49–55, 1936.

MCLACHLAN, G. J. Mahalanobis distance. **Resonance**, v. 4, n. 6, p. 20–26, 1999.

MISIUNAS, D.; VÍTKOVSKÝ, J.; OLSSON, G.; et al. Pipeline Break Detection Using Pressure Transient Monitoring. **Journal of Water Resources Planning and Management**, v. 131, n. 4, p. 316–325, 2005.

MOHAMMADIAN, E.; JAN, B. M.; AZDARPOUR, A.; et al. CO₂-EOR/Sequestration : Current Trends and Future Horizons. In: **Enhanced Oil Recovery Processes - New Technologies**. [s.l.] IntechOpen, 2019.

WARDA, H. A.; ADAM, I. G.; RASHAD, A. B. **A PRACTICAL IMPLEMENTATION OF PRESSURE TRANSIENT ANALYSIS IN LEAK LOCALIZATION IN PIPELINES**. International Pipeline Conference. **Anais**, 2004.

ZAIN, Z.; KECHUT, N. I.; NADESON, G.; et al. **Evaluation of CO₂ Gas Injection for Major Oil Production Fields in Malaysia - Experimental Approach Case Study: Dulang Field**. SPE Asia Pacific Improved Oil Recovery Conference. **Anais**, 2001.

ZANOLI, S. M.; ASTOLFI, G. **Faults diagnosis for a centrifugal machine using the Mahalanobis distance**. 8th IFAC Symposium on Fault Detection. **Anais**.IFAC, 2012.
Disponível em: <<http://dx.doi.org/10.3182/20120829-3-MX-2028.00223>>

ZHANG, T.; TAN, Y.; ZHANG, X.; et al. A novel hybrid technique for leak detection and location in straight pipelines. **Journal of Loss Prevention in the Process Industries**, v. 35, p. 157–168, 2015.

ZUO, L.; JIANG, F.; JIN, B.; et al. Value setting for the rate of pressure drop of automatic line-break control valves in natural gas pipelines. **Journal of Natural Gas Science and Engineering**, v. 26, p. 803–809, 2015.

APÊNDICE A

```

### Mahalanobis calculator for 2 matrices

import pandas as pd
import numpy as np
import sys

def covarfn(a, b): # calculate covariance between two arrays

    covar = a.cov(b) * (len(a) - 1) / len(a)
    return covar

def generatePandasSeries(a):
    for i in range(len(a)):
        a[i] = pd.Series(a[i])

# calculate mahalanobis distance between two matrices
def calculateDistance(a, b, startIndex, nRows, verbose):
    if verbose:
        print("starting iteration ", startIndex)
    endIndex = startIndex + nRows
    nCols = len(a)
    if len(a) != len(b):
        print("not equal number of columns")
        exit

    populationArray = a
    sampleArray = np.array(b)
    sampleArray = sampleArray[:,startIndex:endIndex]
    if verbose:
        print("sample: ", sampleArray)
    sampleArray = sampleArray.tolist()
    generatePandasSeries(populationArray)
    generatePandasSeries(sampleArray)

    ct = np.zeros((nCols, nCols))
    for i in range(len(ct)):
        for j in range(len(ct[i])):
            if i == j:
                ct[i][j] = (covarfn(populationArray[i],populationArray[j]) *
                    len(populationArray[0]) + covarfn(sampleArray[i],sampleArray[j]) *
                    len(sampleArray[0])) / (len(populationArray[0]) + len(sampleArray[0]))
            else:

```

```

        ct[i][j] = -1 * (covarfn(populationArray[i],populationArray[j]) *
len(populationArray[0]) + covarfn(sampleArray[i],sampleArray[j]) *
len(sampleArray[0])) / (len(populationArray[0]) + len(sampleArray[0]))

det = np.linalg.det(ct)
inverse = np.true_divide(ct, det)
v1 = np.zeros(nCols)

for i in range(len(ct)):
    v1[i] = np.average(populationArray[i]) - np.average(sampleArray[i])
v1 = np.asmatrix(np.flip(v1))
v2 = v1.T

distance = np.matmul(np.matmul(v1, inverse), v2)
distance = np.sqrt(distance)
if verbose:
    print("ct\n", ct, "\ndeterminant ", det, "\ninverse \n", inverse, "\nv1\n", v1,
"\nv2\n", v2, "\ndistance ", distance)
return distance

def getAllDistances(a, b, nRows, verbose):
    numIterations = len(b[0])-nRows+1
    retArr = np.zeros(numIterations)
    for i in range(len(retArr)):
        retArr[i] = calculateDistance(a, b, i, nRows, verbose)
    return retArr

def getArrayFromFile(filename):
    columns = []
    with open(filename) as file_in:
        for line in file_in:
            columns.append(list(map(float, line.rstrip("\n").split(','))))
    return columns

def main():
    np.set_printoptions(threshold=np.inf)
    verbose = False
    if len(sys.argv) < 4:
        print("Wrong number of arguments. Call this executable with the following
arguments:\n"
"- all population file\n"
"- sample file\n"
"- number of rows from sample for each iteration\n"
"- verbose (True or False). If set to true, it will print all the intermediate steps from
the calculations."
"\nExample: python mahanolis.py big-file.txt small.file.txt 180 False")
    return

```

```
a = getArrayFromFile(sys.argv[1])
b = getArrayFromFile(sys.argv[2])
nRows = sys.argv[3]
if len(sys.argv) >= 5:
    if sys.argv[4] == "False":
        verbose = ""
    else:
        verbose = sys.argv[4]

if verbose:
    print("input matrices\n", a, "\n", b)
distances = getAllDistances(a, b, int(nRows), verbose)
print(distances)

if __name__ == "__main__":
    main()
```

APÊNDICE B

Tabela B.1: Lista completa dos resultados do Método da Distância de Mahalanobis em cada um dos casos simulados no trabalho.

Caso	Distância de Mahalanobis - Base Operação Normal	Distância de Mahalanobis - Base Rompimento	Resultado do método
Matriz de Confusão - Base Treino			
Setpoint FIC-Feed 1% (+)	0,197	1,014	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 10% (+)	0,429	1,014	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 20% (+)	0,819	0,989	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 50% (+)	1,946	0,903	FALSO POSITIVO
Setpoint FIC-Feed 1% (-)	0,214	0,941	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 5% (-)	0,342	0,805	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 10% (-)	0,564	0,847	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 50% (-)	2,715	1,013	FALSO POSITIVO
Setpoint PIC-Speed 1% (+)	0,197	1,014	VERDADEIRO NEGATIVO
Setpoint PIC-Speed 5% (+)	0,197	1,014	VERDADEIRO NEGATIVO
Setpoint PIC-Speed 20% (+)	0,732	1,014	VERDADEIRO NEGATIVO
Setpoint PIC-Speed 1% (-)	0,226	0,881	VERDADEIRO NEGATIVO
Setpoint PIC-Speed 10% (-)	0,481	0,365	FALSO POSITIVO
Setpoint PIC-Speed 20% (-)	0,710	0,320	FALSO POSITIVO
Abertura VLV-105 1% (-)	0,197	1,013	VERDADEIRO NEGATIVO
Abertura VLV-105 5% (-)	0,197	1,013	VERDADEIRO NEGATIVO
Abertura VLV-105 10% (-)	0,197	1,013	VERDADEIRO NEGATIVO
Abertura VLV-105 50% (-)	0,197	1,014	VERDADEIRO NEGATIVO
Teor de CO2 1% (+)	0,234	0,846	VERDADEIRO NEGATIVO
Teor de CO2 10% (+)	0,390	0,385	FALSO POSITIVO
Teor de CO2 20% (+)	0,559	0,316	FALSO POSITIVO
Teor de CO2 50% (+)	1,018	0,302	FALSO POSITIVO
Teor de CO2 1% (-)	0,205	0,977	VERDADEIRO NEGATIVO
Teor de CO2 5% (-)	0,198	1,009	VERDADEIRO NEGATIVO
Teor de CO2 10% (-)	0,200	1,002	VERDADEIRO NEGATIVO
Teor de CO2 20% (-)	0,299	0,969	VERDADEIRO NEGATIVO
Abertura 0.01% VLV-ROMP2 (Cv = 10 ⁵)	0,292	0,618	FALSO NEGATIVO
Abertura 0.01% VLV-ROMP2 (Cv = 10 ⁶)	0,588	0,315	VERDADEIRO POSITIVO
Abertura 1% VLV-ROMP2 (Cv = 10 ⁶)	0,725	0,309	VERDADEIRO POSITIVO
Abertura 100% VLV-ROMP2 (Cv = 10 ⁶)	0,725	0,309	VERDADEIRO POSITIVO
Rompimento + 1% FIC-FEED	0,766	0,245	VERDADEIRO POSITIVO
Rompimento + 5% FIC-FEED	0,989	0,030	VERDADEIRO POSITIVO
Rompimento + 10% FIC-FEED	1,315	0,324	VERDADEIRO POSITIVO
Rompimento + 50% FIC-FEED	4,293	3,030	VERDADEIRO POSITIVO
Matriz de confusão - Base Teste			
Setpoint FIC-Feed 5% (+)	0,255	1,014	VERDADEIRO NEGATIVO
Setpoint FIC-Feed 20% (-)	1,052	0,974	FALSO POSITIVO
Setpoint PIC-Speed 10% (+)	0,261	1,014	VERDADEIRO NEGATIVO
Setpoint PIC-Speed 5% (-)	0,457	0,463	VERDADEIRO NEGATIVO
Abertura VLV-105 20% (-)	0,197	1,013	VERDADEIRO NEGATIVO
Teor de CO2 5% (+)	0,303	0,585	VERDADEIRO NEGATIVO
Teor de CO2 50% (-)	1,156	0,915	FALSO POSITIVO
Abertura 1% VLV-ROMP2 (Cv = 10 ⁵)	0,723	0,308	VERDADEIRO POSITIVO
Rompimento + 20% FIC-FEED	2,013	0,958	VERDADEIRO POSITIVO

APÊNDICE C

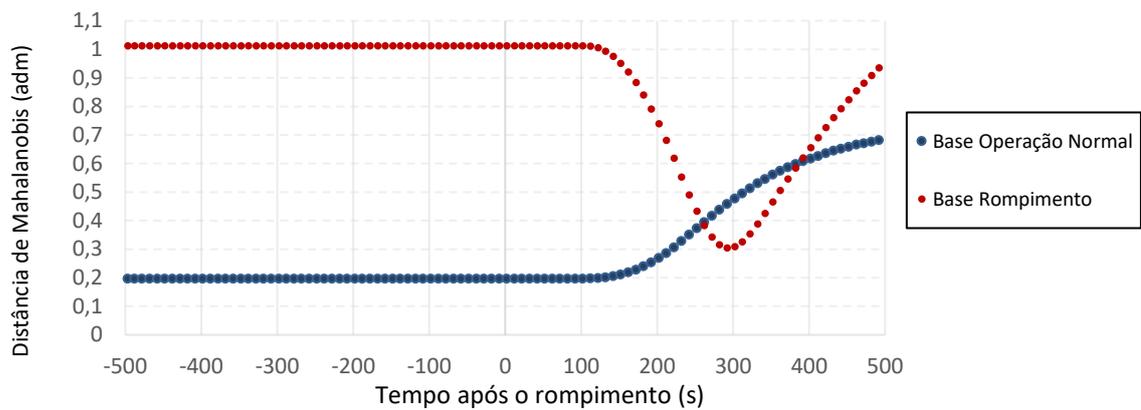


Figura C.1: Comportamento das Distâncias de Mahalanobis calculadas ao longo do tempo para um caso de rompimento (100% de abertura da VLV-ROMP2) e tamanho de amostra (n) igual a 12.

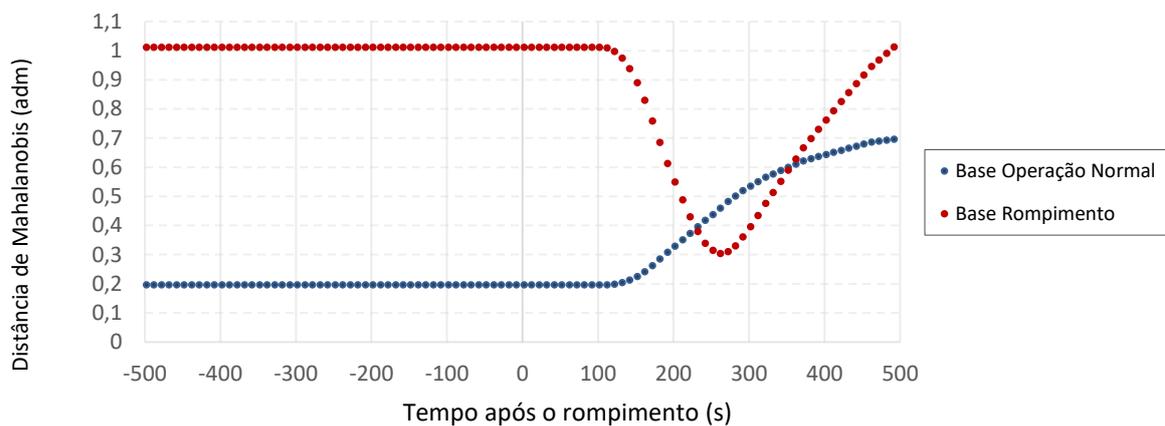


Figura C.2: Comportamento das Distâncias de Mahalanobis calculadas ao longo do tempo para um caso de rompimento (100% de abertura da VLV-ROMP2) e tamanho de amostra (n) igual a 6.

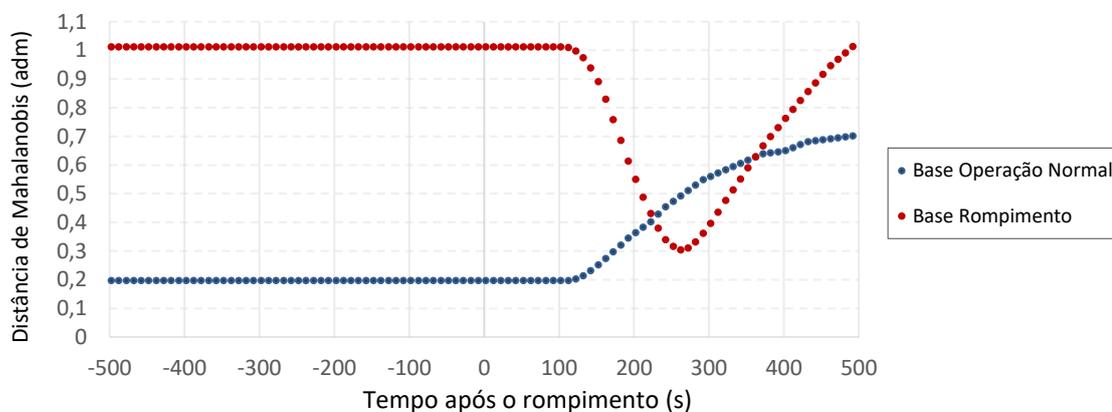


Figura C.3: Comportamento das Distâncias de Mahalanobis calculadas ao longo do tempo para um caso de rompimento (100% de abertura da VLV-ROMP2) e tamanho de amostra (n) igual a 3.